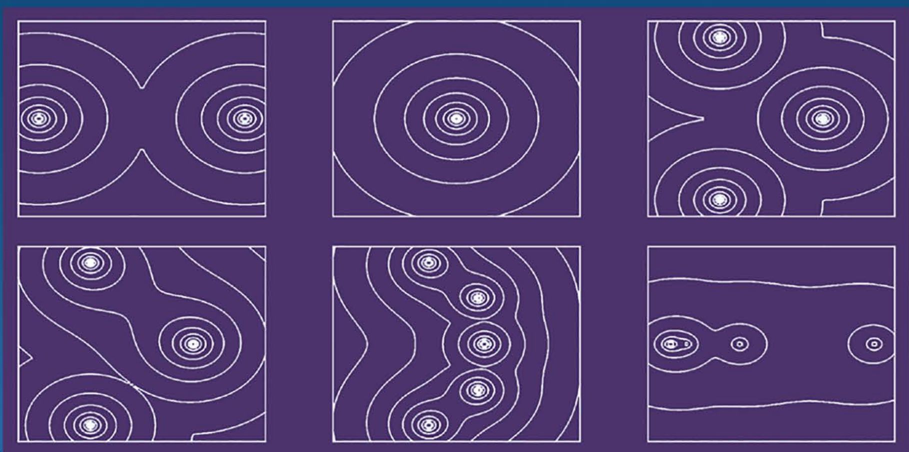
Advances in Applied Mathematics

CRC STANDARD MATHEMATICAL TABLES AND FORMULAS

33RD EDITION



Edited by

Dan Zwillinger

CRC Press
Taylor & Francis Group
A CHAPMAN & HALL BOOK

CRC STANDARD MATHEMATICAL TABLES AND FORMULAS

33RD EDITION

Advances in Applied Mathematics

Series Editor: Daniel Zwillinger

Published Titles

Advanced Engineering Mathematics with MATLAB, Fourth Edition
Dean G. Duffy

CRC Standard Curves and Surfaces with Mathematica®, Third Edition
David H. von Seggern

CRC Standard Mathematical Tables and Formulas, 33rd Edition
Dan Zwillinger

Dynamical Systems for Biological Modeling: An Introduction
Fred Brauer and Christopher Kribs

Fast Solvers for Mesh-Based Computations *Maciej Paszyński*

Green's Functions with Applications, Second Edition *Dean G. Duffy*

Handbook of Peridynamic Modeling *Floriin Bobaru, John T. Foster, Philippe H. Geubelle, and Stewart A. Silling*

Introduction to Financial Mathematics *Kevin J. Hastings*

Linear and Complex Analysis for Applications *John P. D'Angelo*

Linear and Integer Optimization: Theory and Practice, Third Edition
Gerard Sierksma and Yori Zwols

Markov Processes *James R. Kirkwood*

Pocket Book of Integrals and Mathematical Formulas, 5th Edition
Ronald J. Tallarida

Stochastic Partial Differential Equations, Second Edition *Pao-Liu Chow*

Quadratic Programming with Computer Programs *Michael J. Best*

Advances in Applied Mathematics

CRC STANDARD MATHEMATICAL TABLES AND FORMULAS

33RD EDITION

Edited by

Dan Zwillinger

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed on acid-free paper
Version Date: 20171206

International Standard Book Number-13: 978-1-4987-7780-3 (Hardback)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Table of Contents

Preface

It has long been the established policy of CRC Press to publish, in handbook form, the most up-to-date, authoritative, logically arranged, and readily usable reference material available.

Just as pocket calculators have replaced tables of square roots and trig functions; the internet has made printed tabulation of many tables and formulas unnecessary. As the content and capabilities of the internet continue to grow, the content of this book also evolves. For this edition of *Standard Mathematical Tables and Formulae* the content was reconsidered and reviewed. The criteria for inclusion in this edition includes:

- information that is immediately useful as a reference (e.g., interpretation of powers of 10);
- information that is useful and not commonly known (e.g., proof methods);
- information that is more complete or concise than that which can be easily found on the internet (e.g., table of conformal mappings);
- information difficult to find on the internet due to the challenges of entering an appropriate query (e.g., integral tables).

Applying these criteria, practitioners from mathematics, engineering, and the sciences have made changes in several sections and have added new material.

- The "Mathematical Formulas from the Sciences" chapter now includes topics from biology, chemistry, and radar.
- Material has been augmented in many areas, including: acceptance sampling, card games, lattices, and set operations.
- New material has been added on the following topics: continuous wavelet transform, contour integration, coupled analogues, financial options, fractal arithmetic, generating functions, linear temporal logic, matrix pseudospectra, max plus algebra, proof methods, and two dimensional integrals.
- Descriptions of new functions have been added: Lambert, prolate spheroidal, and Weierstrass.

Of course, the same successful format which has characterized earlier editions of the *Handbook* has been retained. Material is presented in a multi-sectional format, with each section containing a valuable collection of fundamental reference material—tabular and expository.

In line with the established policy of CRC Press, the *Handbook* will be updated in as current and timely manner as is possible. Suggestions for the inclusion of new material in subsequent editions and comments regarding the present edition are welcomed. The home page for this book, which will include errata, will be maintained at http://www.mathtable.com/smtf.

This new edition of the *Handbook* will continue to support the needs of practitioners of mathematics in the mathematical and scientific fields, as it has for almost 90 years. Even as the internet becomes more powerful, it is this editor's opinion that the new edition will continue to be a valued reference.

Every book takes time and care. This book would not have been possible without the loving support of my wife, Janet Taylor, and my son, Kent Zwillinger.

May 2017
Daniel Zwillinger
ZwillingerBooks@gmail.com

Editor-in-Chief

Daniel Zwillinger
Rensselaer Polytechnic Institute
Troy, New York

Contributors

George E. Andrews
Evan Pugh University Professor in Mathematics
The Pennsylvania State University
University Park, Pennsylvania

Lawrence Glasser
Professor of Physics Emeritus
Clarkson University
Potsdam, New York

Michael Mascagni
Professor of Computer Science
Professor of Mathematics
Professor of Scientific Computing
Florida State University
Tallahassee, Florida

Ray McLenaghan
Adjunct Professor in Department of Applied Mathematics
University of Waterloo
Waterloo, Ontario, Canada

Roger B. Nelsen
Professor Emeritus of Mathematics
Lewis & Clark College
Portland, Oregon

Dr. Joseph J. Rushanan
MITRE Corporation
Bedford, Massachusetts

Dr. Les Servi
MITRE Corporation
Bedford, Massachusetts

Dr. Michael T. Strauss
President HME
Newburyport, Massachusetts

Dr. Nico M. Temme
Centrum Wiskunde & Informatica
Amsterdam, The Netherlands

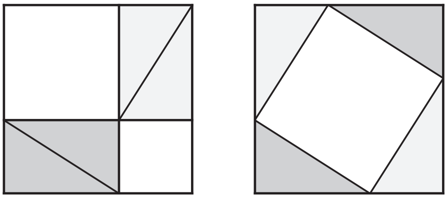
Ahmed I. Zayed
Professor in Department of Mathematical Sciences
DePaul University
Chicago, Illinois

Chapter 1

Numbers and Elementary Mathematics

1.1 PROOFS WITHOUT WORDS

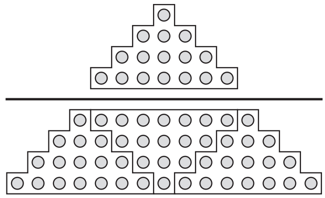
The Pythagorean Theorem



—the *Chou pei suan ching*
(author unknown, circa B.C. 200?)

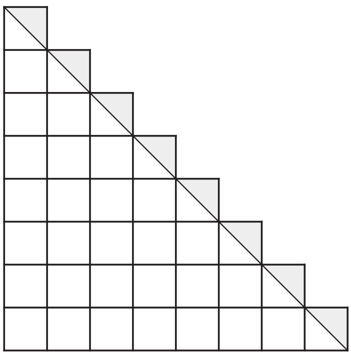
A Property of the Sequence of Odd Integers (Galileo, 1615)

$$\frac{1}{3} = \frac{1+3}{5+7} = \frac{1+3+5}{7+9+11} = \cdots$$



$$\frac{1+3+\cdots+(2n-1)}{(2n+1)+(2n+3)+\cdots+(4n-1)} = \frac{1}{3}$$

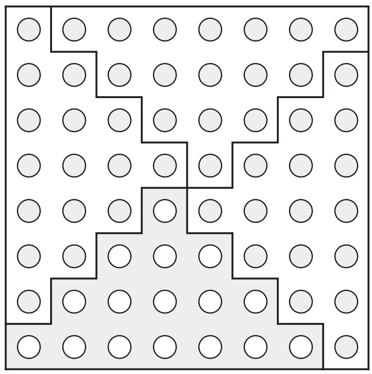
$$1+2+\cdots+n = \frac{n(n+1)}{2}$$



$$1+2+\cdots+n = \frac{1}{2}\cdot n^2 + n\cdot\frac{1}{2} = \frac{n(n+1)}{2}$$

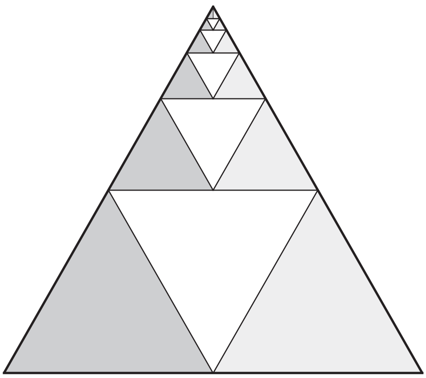
—Ian Richards

$$1+3+5+\cdots+(2n-1) = n^2$$



$$1+3+\cdots+(2n-1) = \frac{1}{4}(2n)^2 = n^2$$

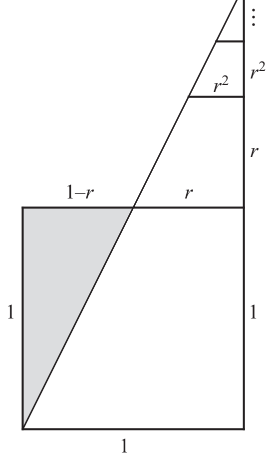
Geometric Series



$$\frac{1}{4} + \left(\frac{1}{4}\right)^2 + \left(\frac{1}{4}\right)^3 + \ldots = \frac{1}{3}$$

—Rick Mabry

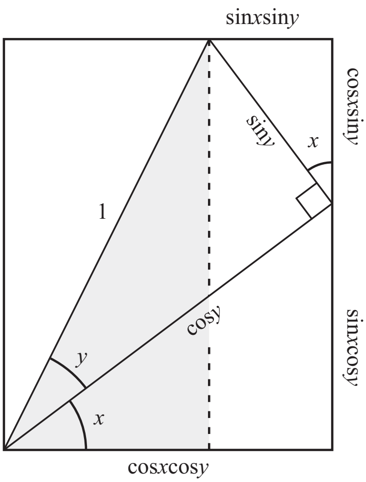
Geometric Series



$$\frac{1 + r + r^2 + \cdots}{1} = \frac{1}{1 - r}$$

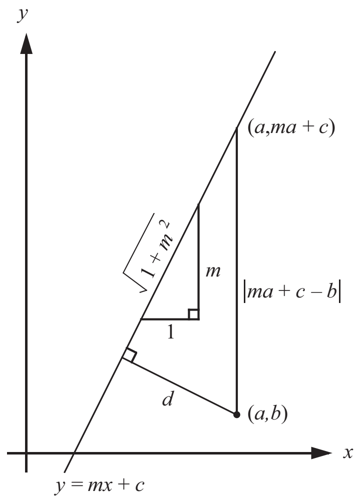
—Benjamin G. Klein
and Irl C. Bivens

Addition Formulae for the Sine and Cosine



$$\sin(x + y) = \sin x \cos y + \cos x \sin y$$
$$\cos(x + y) = \cos x \cos y - \sin x \sin y$$

The Distance Between a Point and a Line

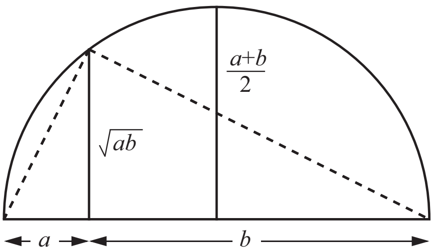


$$\frac{d}{1} = \frac{|ma + c - b|}{\sqrt{1 + m^2}}$$

—R. L. Eisenman

The Arithmetic Mean-Geometric Mean Inequality

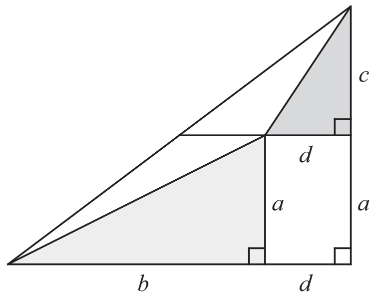
$$a,b > 0 \Rightarrow \frac{a+b}{2} \geq \sqrt{ab}$$



—Charles D. Gallant

The Mediant Property

$$\frac{a}{b} < \frac{c}{d} \quad \Rightarrow \quad \frac{a}{b} < \frac{a+c}{b+d} < \frac{c}{d}$$



—Richard A. Gibbs

1.2 CONSTANTS

1.2.1 DIVISIBILITY TESTS

1. Divisibility by 2: the last digit is divisible by 2
2. Divisibility by 3: the sum of the digits is divisible by 3
3. Divisibility by 4: the number formed from the last 2 digits is divisible by 4
4. Divisibility by 5: the last digit is either 0 or 5
5. Divisibility by 6: is divisible by both 2 and 3
6. Divisibility by 9: the sum of the digits is divisible by 9
7. Divisibility by 10: the last digit is 0
8. Divisibility by 11: the difference between the sum of the odd digits and the sum of the even digits is divisible by 11

EXAMPLE Consider the number $N = 1036728$.

- The last digit is 8, so N is divisible by 2.
- The last two digits are 28 which is divisible by 4, so N is divisible by 4.
- The sum of the digits is $27 = 1 + 0 + 3 + 6 + 7 + 2 + 8$. This is divisible by 3, so N is divisible by 3. This is also divisible by 9, so N is divisible by 9.
- The sum of the odd digits is $19 = 1 + 3 + 7 + 8$ and the sum of the even digits is $8 = 6 + 2$; the difference is $19 - 8 = 11$. This is divisible by 11, so N is divisible by 11.

1.2.2 DECIMAL MULTIPLES AND PREFIXES

The prefix names and symbols below are taken from Conference Générale des Poids et Mesures, 1991. The common names are for the United States.

Factor	Prefix	Symbol	Common name
$10^{(10^{100})}$			googolplex
10^{100}			googol
10^{24}	yotta	Y	heptillion
10^{21}	zetta	Z	hexillion
$1\ 000\ 000\ 000\ 000\ 000\ 000 = 10^{18}$	exa	E	quintillion
$1\ 000\ 000\ 000\ 000\ 000 = 10^{15}$	peta	P	quadrillion
$1\ 000\ 000\ 000\ 000 = 10^{12}$	tera	T	trillion
$1\ 000\ 000\ 000 = 10^{9}$	giga	G	billion
$1\ 000\ 000 = 10^{6}$	mega	M	million
$1\ 000 = 10^{3}$	kilo	k	thousand
$100 = 10^{2}$	hecto	H	hundred
$10 = 10^{1}$	deka	da	ten
$0.1 = 10^{-1}$	deci	d	tenth
$0.01 = 10^{-2}$	centi	c	hundredth
$0.001 = 10^{-3}$	milli	m	thousandth
$0.000\ 001 = 10^{-6}$	micro	μ	millionth
$0.000\ 000\ 001 = 10^{-9}$	nano	n	billionth
$0.000\ 000\ 000\ 001 = 10^{-12}$	pico	p	trillionth
$0.000\ 000\ 000\ 000\ 001 = 10^{-15}$	femto	f	quadrillionth
$0.000\ 000\ 000\ 000\ 000\ 001 = 10^{-18}$	atto	a	quintillionth
10^{-21}	zepto	z	hexillionth
10^{-24}	yocto	y	heptillionth

1.2.3 BINARY PREFIXES

A byte is 8 bits. A kibibyte is $2^{10} = 1024$ bytes. Other prefixes for power of 2 are:

Factor	Prefix	Symbol
2^{10}	kibi	Ki
2^{20}	mebi	Mi
2^{30}	gibi	Gi
2^{40}	tebi	Ti
2^{50}	pebi	Pi
2^{60}	exbi	Ei

1.2.4 INTERPRETATIONS OF POWERS OF 10

10^{-43}	Planck time in seconds
10^{-35}	Planck length in meters
10^{-30}	mass of an electron in kilograms
10^{-27}	mass of a proton in kilograms
10^{-15}	the radius of the hydrogen nucleus (a proton) in meters
10^{-11}	the likelihood of being dealt 13 top honors in bridge
10^{-10}	the (Bohr) radius of a hydrogen atom in meters
10^{-9}	the number of seconds it takes light to travel one foot
10^{-6}	the likelihood of being dealt a royal flush in poker
10^{0}	the density of water is 1 gram per milliliter
10^{1}	the number of fingers that people have
10^{2}	the number of stable elements in the periodic table
10^{4}	the speed of the Earth around the sun in meters/second
10^{5}	the number of hairs on a human scalp
10^{6}	the number of words in the English language
10^{7}	the number of seconds in a year
10^{8}	the speed of light in meters per second
10^{9}	the number of heartbeats in a lifetime for most mammals
10^{10}	the number of people on the earth
10^{11}	the distance from the Earth to the sun in meters
10^{13}	diameter of the solar system in meters
10^{14}	number of cells in the human body
10^{15}	the surface area of the earth in square meters
10^{16}	the number of meters light travels in one year
10^{17}	the age of the universe in seconds
10^{18}	the volume of water in the earth's oceans in cubic meters
10^{19}	the number of possible positions of Rubik's cube
10^{21}	the volume of the earth in cubic meters
10^{24}	the number of grains of sand in the Sahara desert
10^{25}	the mass of the earth in kilograms
10^{30}	the mass of the sun in kilograms
10^{50}	the number of atoms in the earth
10^{52}	the mass of the observable universe in kilograms
10^{54}	the number of elements in the monster group
10^{78}	the volume of the universe in cubic meters

(Note: these numbers have been rounded to the nearest power of ten.)

1.2.5 NUMERALS IN DIFFERENT LANGUAGES

	0	1	2	3	4	5	6	7	8	9
Arabic	٠	١	٢	٣	٤	٥	٦	٧	٨	٩
Bengali	০	১	২	৩	৪	৫	৬	৭	৮	৯
Chinese (simple)	〇	一	二	三	四	五	六	七	八	九
Chinese (complex)	零	壹	貳	參	肆	伍	陸	柒	捌	玖
Devanagari	०	१	२	३	४	५	६	७	८	९
Ge'ez (Ethiopic)		፩	፪	፫	፬	፭	፮	፯	፰	፱
Gujarati	૦	૧	૨	૩	૪	૫	૬	૭	૮	૯
Gurmukhi numbers	੦	੧	੨	੩	੪	੫	੬	੭	੮	੯
Kannada	೦	೧	೨	೩	೪	೫	೬	೭	೮	೯
Khmer	០	១	២	៣	៤	៥	៦	៧	៨	៩
Lao	໐	໑	໒	໓	໔	໕	໖	໗	໘	໙
Limbu	᥆	᥇	᥈	᥉	᥊	᥋	᥌	᥍	᥎	᥏
Malayalam	൦	൧	൨	൩	൪	൫	൬	൭	൮	൯
Mongolian	᠐	᠑	᠒	᠓	᠔	᠕	᠖	᠗	᠘	᠙
Myanmar	၀	၁	၂	၃	၄	၅	၆	၇	၈	၉
Odia	୦	୧	୨	୩	୪	୫	୬	୭	୮	୯
Roman		I	II	III	IV	V	VI	VII	VIII	IX
Tamil	௦	௧	௨	௩	௪	௫	௬	௭	௮	௯
Telugu	౦	౧	౨	౩	౪	౫	౬	౭	౮	౯
Thai	๐	๑	๒	๓	๔	๕	๖	๗	๘	๙
Urdu	۰	۱	۲	۳	۴	۵	۶	۷	۸	۹

1.2.6 ROMAN NUMERALS

The major symbols in Roman numerals are I $=1$, V $=5$, X $=10$, L $=50$, C $=100$, D $=500$, and M $=1{,}000$. The rules for constructing Roman numerals are:

1. A symbol following one of equal or greater value adds its value. (For example, II $=2$, XI $=11$, and DV $=505$.)
2. A symbol following one of lesser value has the lesser value subtracted from the larger value. An I is only allowed to precede a V or an X, an X is only allowed to precede an L or a C, and a C is only allowed to precede a D or an M. (For example IV $=4$, IX $=9$, and XL $=40$.)
3. When a symbol stands between two of greater value, its value is subtracted from the second and the result is added to the first. (For example, XIV$=10+(5-1)=14$, CIX$=100+(10-1)=109$, DXL$=500+(50-10)=540$.)
4. When two ways exist for representing a number, the one in which the symbol of larger value occurs earlier in the string is preferred. (For example, 14 is represented as XIV, not as VIX.)

Decimal number	1	2	3	4	5	6	7	8	9
Roman numeral	I	II	III	IV	V	VI	VII	VIII	IX

10	14	50	200	400	500	600	999	1000
X	XIV	L	CC	CD	D	DC	CMXCIX	M

1995	1999	2000	2001	2017	2018
MCMXCV	MCMXCIX	MM	MMI	MMXVII	MMXVIII

1.2.7 TYPES OF NUMBERS

1. **Natural numbers** The set of *natural numbers*, $\{0, 1, 2, \ldots\}$, is customarily denoted by $\mathbb{N}$. Many authors do not consider 0 to be a natural number.

2. **Integers** The set of *integers*, $\{0, \pm 1, \pm 2, \ldots\}$, is customarily denoted by $\mathbb{Z}$.

3. **Rational numbers** The set of *rational numbers*, $\{\frac{p}{q} \mid p, q \in \mathbb{Z}, q \neq 0\}$, is customarily denoted by $\mathbb{Q}$.

   (a) Two fractions $\frac{p}{q}$ and $\frac{r}{s}$ are equal if and only if $ps = qr$.
   (b) Addition of fractions is defined by $\frac{p}{q} + \frac{r}{s} = \frac{ps+qr}{qs}$.
   (c) Multiplication of fractions is defined by $\frac{p}{q} \cdot \frac{r}{s} = \frac{pr}{qs}$.

4. **Real numbers** Real numbers are defined to be converging sequences of rational numbers or as decimals that might or might not repeat. The set of *real numbers* is customarily denoted by $\mathbb{R}$.

   Real numbers can be divided into two subsets. One subset, the *algebraic numbers*, are real numbers which solve a polynomial equation in one variable with integer coefficients. For example; $\sqrt{2}$ is an algebraic number because it solves the polynomial equation $x^2 - 2 = 0$; and all rational numbers are algebraic. Real numbers that are not algebraic numbers are called *transcendental numbers*. Examples of transcendental numbers include π and e.

5. **Definition of infinity** The real numbers are extended to $\overline{\mathbb{R}}$ by the inclusion of $+\infty$ and $-\infty$ with the following definitions

   (a) for x in $\mathbb{R}$: $-\infty < x < \infty$
   (b) for x in $\mathbb{R}$: $x + \infty = \infty$
   (c) for x in $\mathbb{R}$: $x - \infty = -\infty$
   (d) for x in $\mathbb{R}$: $\dfrac{x}{\infty} = \dfrac{x}{-\infty} = 0$
   (e) if $x > 0$ then $x \cdot \infty = \infty$
   (f) if $x > 0$ then $x \cdot (-\infty) = -\infty$
   (g) $\infty + \infty = \infty$
   (h) $-\infty - \infty = -\infty$
   (i) $\infty \cdot \infty = \infty$
   (j) $-\infty \cdot (-\infty) = \infty$

6. **Complex numbers** The set of *complex numbers* is customarily denoted by $\mathbb{C}$. They are numbers of the form $a + bi$, where $i^2 = -1$, and a and b are real numbers.

Operation	computation	result
addition	$(a + bi) + (c + di)$	$(a + c) + i(b + d)$
multiplication	$(a + bi)(c + di)$	$(ac - bd) + (ad + bc)i$
reciprocal	$\dfrac{1}{a + bi}$	$\left(\dfrac{a}{a^2 + b^2}\right) - \left(\dfrac{b}{a^2 + b^2}\right) i$
complex conjugate	$z = a + bi$	$\overline{z} = a - bi$

   Properties include: $\overline{z + w} = \overline{z} + \overline{w}$ and $\overline{zw} = \overline{z}\,\overline{w}$.

1.2.8 REPRESENTATION OF NUMBERS

Numerals as usually written have radix or base 10, so the numeral $a_n a_{n-1} \dots a_1 a_0$ represents the number $a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_2 10^2 + a_1 10 + a_0$. However, other bases can be used, particularly bases 2, 8, and 16. When a number is written in base 2, the number is said to be in binary notation. The names of other bases are:

2	binary	6	senary	10	decimal	20	vigesimal
3	ternary	7	septenary	11	undenary	60	sexagesimal
4	quaternary	8	octal	12	duodecimal		
5	quinary	9	nonary	16	hexadecimal		

When writing a number in base b, the digits used range from 0 to $b-1$. If $b > 10$, then the digit A stands for 10, B for 11, etc. When a base other than 10 is used, it is indicated by a subscript:

$$\begin{aligned} 10111_2 &= 1 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2 + 1 = 23, \\ \mathrm{A3}_{16} &= 10 \times 16 + 3 = 163, \\ 543_7 &= 5 \times 7^2 + 4 \times 7 + 3 = 276. \end{aligned} \tag{1.2.1}$$

To convert a number from base 10 to base b, divide the number by b, and the remainder will be the last digit. Then divide the quotient by b, using the remainder as the previous digit. Continue this process until a quotient of 0 is obtained.

EXAMPLE To convert 573 to base 12, divide 573 by 12, yielding a quotient of 47 and a remainder of 9; hence, "9" is the last digit. Divide 47 by 12, yielding a quotient of 3 and a remainder of 11 (which we represent with a "B"). Divide 3 by 12 yielding a quotient of 0 and a remainder of 3. Therefore, $573_{10} = \mathrm{3B9}_{12}$.

Converting from base b to base r can be done by converting to and from base 10. However, it is simple to convert from base b to base b^n. For example, to convert 110111101_2 to base 16, group the digits in fours (because 16 is 2^4), yielding $1\ 1011\ 1101_2$, and then convert each group of 4 to base 16 directly, yielding $\mathrm{1BD}_{16}$.

1.2.9 REPRESENTATION OF COMPLEX NUMBERS – DEMOIVRE'S THEOREM

A complex number $a + bi$ can be written in the form $re^{i\theta}$, where $r^2 = a^2 + b^2$ and $\tan\theta = b/a$. Because $e^{i\theta} = \cos\theta + i\sin\theta$,

$$\begin{aligned} (a+bi)^n &= r^n(\cos n\theta + i \sin n\theta), \\ \sqrt[n]{1} &= \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n}, \qquad k = 0, 1, \dots, n-1. \\ \sqrt[n]{-1} &= \cos\frac{(2k+1)\pi}{n} + i\sin\frac{(2k+1)\pi}{n}, \qquad k = 0, 1, \dots, n-1. \end{aligned} \tag{1.2.2}$$

1.2.10 ARROW NOTATION

Arrow notation is used to represent large numbers. Start with $m \uparrow n = \underbrace{m \cdot m \cdots m}_{n}$, then (evaluation proceeds from the right)

$$m \uparrow\uparrow n = \underbrace{m \uparrow m \uparrow \cdots \uparrow m}_{n} \qquad m \uparrow\uparrow\uparrow n = \underbrace{m \uparrow\uparrow m \uparrow\uparrow \cdots \uparrow\uparrow m}_{n}$$

For example, $m \uparrow n = m^n$, $m \uparrow\uparrow 2 = m^m$, and $m \uparrow\uparrow 3 = m^{(m^m)}$.

1.2.11 ONES AND TWOS COMPLEMENT

One's and two's complement are ways to represent numbers in a computer. For positive values the binary representation, the ones' complement representation, and the twos' complement representation are the same.

- *Ones' complement* represents integers from $-\left(2^{N-1}-1\right)$ to $2^{N-1}-1$. For negative values, the binary representation of the absolute value is obtained, and then all of the bits are inverted (i.e., swapping 0's for 1's and vice versa).
- *Twos' complement* represents integers from -2^{N-1} to $2^{N-1}-1$. For negative vales, the two's complement representation is the same as the value one added to the ones' complement representation.

Number	Ones' complement	Twos' complement
7	0111	0111
6	0110	0110
5	0101	0101
4	0100	0100
3	0011	0011
2	0010	0010
1	0001	0001
0	0000	0000
−0	1111	
−1	1110	1111
−2	1101	1110
−3	1100	1101
−4	1011	1100
−5	1010	1011
−6	1001	1010
−7	1000	1001
−8		1000

1.2.12 SYMMETRIC BASE THREE REPRESENTATION

In the symmetric base three representation, powers of 3 are added and subtracted to represent numbers; the symbols $\{\downarrow, 0, \uparrow\}$ represent $\{-1, 0, 1\}$. For example, one writes $\uparrow\downarrow\downarrow$ for 5 since $5 = 9-3-1$. To negate a number in symmetric base three, turn its symbol upside down, e.g., $-5 = \downarrow\uparrow\uparrow$. Basic arithmetic operations are especially simple in this base.

1.2.13 HEXADECIMAL ADDITION AND SUBTRACTION TABLE

A = 10, B = 11, C = 12, D = 13, E = 14, F = 15.
Example: 6 + 2 = 8; hence 8 − 6 = 2 and 8 − 2 = 6.
Example: 4 + E = 12; hence 12 − 4 = E and 12 − E = 4.

	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10
2	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11
3	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12
4	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13
5	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14
6	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15
7	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16
8	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17
9	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	18
A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	18	19
B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	18	19	1A
C	0D	0E	0F	10	11	12	13	14	15	16	17	18	19	1A	1B
D	0E	0F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C
E	0F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D
F	10	11	12	13	14	15	16	17	18	19	1A	1B	1C	1D	1E

1.2.14 HEXADECIMAL MULTIPLICATION TABLE

Example: 2 × 4 = 8.
Example: 2 × F = 1E.

	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
1	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
2	02	04	06	08	0A	0C	0E	10	12	14	16	18	1A	1C	1E
3	03	06	09	0C	0F	12	15	18	1B	1E	21	24	27	2A	2D
4	04	08	0C	10	14	18	1C	20	24	28	2C	30	34	38	3C
5	05	0A	0F	14	19	1E	23	28	2D	32	37	3C	41	46	4B
6	06	0C	12	18	1E	24	2A	30	36	3C	42	48	4E	54	5A
7	07	0E	15	1C	23	2A	31	38	3F	46	4D	54	5B	62	69
8	08	10	18	20	28	30	38	40	48	50	58	60	68	70	78
9	09	12	1B	24	2D	36	3F	48	51	5A	63	6C	75	7E	87
A	0A	14	1E	28	32	3C	46	50	5A	64	6E	78	82	8C	96
B	0B	16	21	2C	37	42	4D	58	63	6E	79	84	8F	9A	A5
C	0C	18	24	30	3C	48	54	60	6C	78	84	90	9C	A8	B4
D	0D	1A	27	34	41	4E	5B	68	75	82	8F	9C	A9	B6	C3
E	0E	1C	2A	38	46	54	62	70	7E	8C	9A	A8	B6	C4	D2
F	0F	1E	2D	3C	4B	5A	69	78	87	96	A5	B4	C3	D2	E1

1.3 SPECIAL NUMBERS

1.3.1 POWERS OF 2

1	2	0.5
2	4	0.25
3	8	0.125
4	16	0.0625
5	32	0.03125
6	64	0.015625
7	128	0.0078125
8	256	0.00390625
9	512	0.001953125
10	1024	0.0009765625
11	2048	0.00048828125
12	4096	0.000244140625
13	8192	0.0001220703125
14	16384	0.00006103515625
15	32768	0.000030517578125
16	65536	0.0000152587890625
17	131072	0.00000762939453125
18	262144	0.000003814697265625
19	524288	0.0000019073486328125
20	1048576	0.00000095367431640625
21	2097152	0.000000476837158203125
22	4194304	0.0000002384185791015625
23	8388608	0.00000011920928955078125
24	16777216	0.000000059604644775390625
25	33554432	0.0000000298023223876953125

1.3.2 POWERS OF 10 IN HEXADECIMAL

n	10^n	10^{-n}
0	1_{16}	1_{16}
1	A_{16}	$0.1999999999999999999\ldots_{16}$
2	64_{16}	$0.028F5C28F5C28F5C28F5\ldots_{16}$
3	$3E8_{16}$	$0.004189374BC6A7EF9DB2\ldots_{16}$
4	2710_{16}	$0.00068DB8BAC710CB295E\ldots_{16}$
5	$186A0_{16}$	$0.0000A7C5AC471B478423\ldots_{16}$
6	$F4240_{16}$	$0.000010C6F7A0B5ED8D36\ldots_{16}$
7	989680_{16}	$0.000001AD7F29ABCAF485\ldots_{16}$
8	$5F5E100_{16}$	$0.0000002AF31DC4611873\ldots_{16}$
9	$3B9ACA00_{16}$	$0.000000044B82FA09B5A5\ldots_{16}$
10	$2540BE400_{16}$	$0.000000006DF37F675EF6\ldots_{16}$

1.3.3 SPECIAL CONSTANTS

1.3.3.1 The constant π

The transcendental number π is defined as the ratio of the circumference of a circle to the diameter. It is also the ratio of the area of a circle to the square of the radius (r) and appears in several formulas in geometry and trigonometry

$$\text{circumference of a circle} = 2\pi r, \qquad \text{volume of a sphere} = \frac{4}{3}\pi r^3,$$
$$\text{area of a circle} = \pi r^2, \qquad \text{surface area of a sphere} = 4\pi r^2.$$

One method of computing π is to use the infinite series for the function $\tan^{-1} x$ and one of the identities

$$\begin{aligned}\pi &= 4\ \tan^{-1} 1 = 6\ \tan^{-1}\frac{1}{\sqrt{3}}\\ &= 2\ \tan^{-1}\frac{1}{2} + 2\ \tan^{-1}\frac{1}{3} + 8\ \tan^{-1}\frac{1}{5} - 2\ \tan^{-1}\frac{1}{239}\\ &= 24\ \tan^{-1}\frac{1}{8} + 8\ \tan^{-1}\frac{1}{57} + 4\ \tan^{-1}\frac{1}{239}\\ &= 48\ \tan^{-1}\frac{1}{18} + 32\ \tan^{-1}\frac{1}{57} - 20\ \tan^{-1}\frac{1}{239}\end{aligned} \tag{1.3.1}$$

There are many identities involving π. For example:

$$\pi = \sum_{i=0}^{\infty} \frac{1}{16^i}\left(\frac{4}{8i+1} - \frac{2}{8i+4} - \frac{1}{8i+5} - \frac{1}{8i+6}\right)$$

$$\pi = \lim_{k\to\infty} 2^k \underbrace{\sqrt{2 - \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \sqrt{2 + \cdots + \sqrt{2 + \sqrt{2}}}}}}}}}_{k \text{ square roots}}$$

$$\frac{2}{\pi} = \prod_{k=1}^{\infty} \frac{\overbrace{\sqrt{2 + \sqrt{2 + \cdots + \sqrt{2 + \sqrt{2}}}}}^{k \text{ square roots}}}{2}$$

$$\frac{\pi^3}{32} = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^3} = 1 - \frac{1}{27} + \frac{1}{125} - \frac{1}{343} + \dots$$

To 100 decimal places:

$\pi \approx 3.$ 14159 26535 89793 23846 26433 83279 50288 41971 69399 37510
58209 74944 59230 78164 06286 20899 86280 34825 34211 70679

In different bases:

$$\pi \approx \texttt{11.0010010000111111011010101000100010000101101000 1...}_2$$
$$\pi \approx \texttt{3.1103755242102643021514230630505600670163211220 1...}_8$$
$$\pi \approx \texttt{3.243F6A8885A308D313198A2E03707344A4093822299F...}_{16}$$

To 50 decimal places:

$$\begin{aligned}
\pi/8 &\approx 0.39269\ 90816\ 98724\ 15480\ 78304\ 22909\ 93786\ 05246\ 46174\ 92189\\
\pi/4 &\approx 0.78539\ 81633\ 97448\ 30961\ 56608\ 45819\ 87572\ 10492\ 92349\ 84378\\
\pi/3 &\approx 1.04719\ 75511\ 96597\ 74615\ 42144\ 61093\ 16762\ 80657\ 23133\ 12504\\
\pi/2 &\approx 1.57079\ 63267\ 94896\ 61923\ 13216\ 91639\ 75144\ 20985\ 84699\ 68755\\
\sqrt{\pi} &\approx 1.77245\ 38509\ 05516\ 02729\ 81674\ 83341\ 14518\ 27975\ 49456\ 12239
\end{aligned}$$

In 2016 π was computed to 12.1 trillion decimal digits. The frequency counts of the digits for $\pi - 3$, using 1 trillion decimal places, are:

digit 0:	99999485134	digit 5:	99999671008
digit 1:	99999945664	digit 6:	99999807503
digit 2:	100000480057	digit 7:	99999818723
digit 3:	99999787805	digit 8:	100000791469
digit 4:	100000357857	digit 9:	99999854780

1.3.3.2 The constant e

The transcendental number e is the base of natural logarithms. It is given by

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = \sum_{n=0}^{\infty}\frac{1}{n!}. \tag{1.3.2}$$

$$\begin{aligned}
e \approx 2.&71828\ 18284\ 59045\ 23536\ 02874\ 71352\ 66249\ 77572\ 47093\ 69995\\
&95749\ 66967\ 62772\ 40766\ 30353\ 54759\ 45713\ 82178\ 52516\ 64274\ \ldots
\end{aligned}$$

In different bases:

$$e \approx \texttt{10.1011011111100001010100010110001010001010111011 0...}_2$$
$$e \approx \texttt{2.5576052130505355124652773425420047172363616613 4...}_8$$
$$e \approx \texttt{2.B7E151628AED2A6ABF7158809CF4F3C762E7160F3...}_{16}$$

The exponential function is $e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!}$. Euler's formula relates e and π: $e^{\pi i} = -1$

1.3.3.3 The constant γ

Euler's constant γ is defined by

$$\gamma = \lim_{n\to\infty}\left(\sum_{k=1}^{n}\frac{1}{k} - \log n\right). \tag{1.3.3}$$

It is not known whether γ is rational or irrational.

$$\begin{aligned}
\gamma \approx 0.&57721\ 56649\ 01532\ 86060\ 65120\ 90082\ 40243\ 10421\ 59335\ 93992\\
&35988\ 05767\ 23488\ 48677\ 26777\ 66467\ 09369\ 47063\ 29174\ 67495\ \ldots
\end{aligned}$$

1.3.3.4 The constant ϕ

The golden ratio, ϕ, is defined as the positive root of the equation $\frac{\phi}{1} = \frac{1+\phi}{\phi}$ or $\phi^2 = \phi+1$; that is $\phi = \frac{1+\sqrt{5}}{2}$. There is the continued faction representation $\phi = [\,\overline{1}\,]$ and the representation in square roots

$$\phi = \sqrt{1 + \sqrt{1 + \sqrt{1 + \sqrt{1 + \dots}}}}$$

$$\phi \approx 1.61803\,39887\,49894\,84820\,45868\,34365\,63811\,77203\,09179\,80576\,\dots$$

1.3.4 FACTORIALS

The factorial of n, denoted $n!$, is the product of all positive integers less than or equal to n; $n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$. By definition, $0! = 1$. If n is a negative integer ($n = -1, -2, \dots$) then $n! = \pm\infty$. The generalization of the factorial function to non-integer arguments is the gamma function (see page 478); when n is an integer, $\Gamma(n) = (n-1)!$.

The double factorial of n is denoted by $n!!$ and is defined as $n!! = n(n-2)(n-4)\cdots$, where the last term in the product is 2 or 1, depending on whether n is even or odd. The *shifted factorial* (also called *Pochhammer's symbol*) is denoted by $(a)_n$ and is defined as

$$(a)_n = \underbrace{a \cdot (a+1) \cdot (a+2) \cdots (a+n-1)}_{n \text{ terms}} = \frac{(a+n-1)!}{(a-1)!} = \frac{\Gamma(a+n)}{\Gamma(a)} \tag{1.3.4}$$

Approximations to $n!$ for large n include Stirling's formula (the first term of the following)

$$n! \approx \sqrt{2\pi e}\left(\frac{n}{e}\right)^{n+\frac{1}{2}}\left(1 + \frac{1}{12n} + \frac{1}{288n^2} + \dots\right) \tag{1.3.5}$$

and Burnside's formula

$$n! \approx \sqrt{2\pi}\left(\frac{n+\frac{1}{2}}{e}\right)^{n+\frac{1}{2}} \tag{1.3.6}$$

n	$n!$	$\log_{10} n!$	$n!!$	$\log_{10} n!!$
0	1	0.00000	1	0.00000
1	1	0.00000	1	0.00000
2	2	0.30103	2	0.30103
3	6	0.77815	3	0.47712
4	24	1.38021	8	0.90309
5	120	2.07918	15	1.17609
6	720	2.85733	48	1.68124
7	5040	3.70243	105	2.02119
8	40320	4.60552	384	2.58433
9	3.6288×10^5	5.55976	945	2.97543
10	3.6288×10^6	6.55976	3840	3.58433

n	$n!$	$\log_{10} n!$	$n!!$	$\log_{10} n!!$
20	2.4329×10^{18}	18.38612	3.7159×10^{9}	9.57006
30	2.6525×10^{32}	32.42366	4.2850×10^{16}	16.63195
40	8.1592×10^{47}	47.91165	2.5511×10^{24}	24.40672
50	3.0414×10^{64}	64.48307	5.2047×10^{32}	32.71640
60	8.3210×10^{81}	81.92017	2.8481×10^{41}	41.45456
70	1.1979×10^{100}	100.07841	3.5504×10^{50}	50.55028
80	7.1569×10^{118}	118.85473	8.9711×10^{59}	59.95284
90	1.4857×10^{138}	138.17194	4.2088×10^{69}	69.62416
100	9.3326×10^{157}	157.97000	3.4243×10^{79}	79.53457
150	5.7134×10^{262}	262.75689	9.3726×10^{131}	131.97186
500	1.2201×10^{1134}	1134.0864	5.8490×10^{567}	567.76709
1000	4.0239×10^{2567}	2567.6046	3.9940×10^{1284}	1284.6014

1.3.5 BERNOULLI POLYNOMIALS AND NUMBERS

The *Bernoulli polynomials* $B_n(x)$ are defined by the generating function

$$\frac{te^{xt}}{e^t - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{t^n}{n!}. \tag{1.3.7}$$

These polynomials can be defined recursively by $B_0(x) = 1$, $B'_n(x) = nB_{n-1}(x)$, and $\int_0^1 B_n(x)\,dx = 0$ for $n \geq 1$. The identity $B_{k+1}(x+1) - B_{k+1}(x) = (k+1)x^k$ means that sums of powers can be computed via Bernoulli polynomials

$$1^k + 2^k + \cdots + n^k = \frac{B_{k+1}(n+1) - B_{k+1}(0)}{k+1}. \tag{1.3.8}$$

The Bernoulli numbers are the Bernoulli polynomials evaluated at 0: $B_n = B_n(0)$. A generating function for the Bernoulli numbers is $\sum_{n=0}^{\infty} B_n \frac{t^n}{n!} = \frac{t}{e^t - 1}$. In the following table each Bernoulli number is written as a fraction of integers: $B_n = N_n/D_n$. Note that $B_{2m+1} = 0$ for $m \geq 1$.

n	$B_n(x)$
0	1
1	$(2x-1)/2$
2	$(6x^2 - 6x + 1)/6$
3	$(2x^3 - 3x^2 + x)/2$
4	$(30x^4 - 60x^3 + 30x^2 - 1)/30$
5	$(6x^5 - 15x^4 + 10x^3 - x)/6$

n	N_n	D_n	B_n
0	1	1	1.00000×10^{0}
1	-1	2	-5.00000×10^{-1}
2	1	6	1.66667×10^{-1}
4	-1	30	-3.33333×10^{-2}
6	1	42	2.38095×10^{-2}
8	-1	30	-3.33333×10^{-2}
10	5	66	7.57576×10^{-2}

1.3.6 EULER POLYNOMIALS AND NUMBERS

The *Euler polynomials* $E_n(x)$ are defined by the generating function

$$\frac{2e^{xt}}{e^t+1} = \sum_{n=0}^{\infty} E_n(x)\frac{t^n}{n!}. \tag{1.3.9}$$

Alternating sums of powers can be computed in terms of Euler polynomials

$$\sum_{i=1}^{n}(-1)^{n-i}i^k = n^k - (n-1)^k + \cdots \mp 2^k \pm 1^k = \frac{E_k(n+1)+(-1)^n E_k(0)}{2}. \tag{1.3.10}$$

The Euler numbers are defined as $E_n = 2^n E_n(\frac{1}{2})$. A generating function is

$$\sum_{n=0}^{\infty} E_n\frac{t^n}{n!} = \frac{2e^t}{e^{2t}+1} \tag{1.3.11}$$

n	$E_n(x)$
0	1
1	$(2x-1)/2$
2	x^2-x
3	$(4x^3-6x^2+1)/4$
4	x^4-2x^3+x
5	$(2x^5-5x^4+5x^2-1)/2$

n	E_n
2	-1
4	5
6	-61
8	1385
10	-50521
12	2702765

1.3.7 FIBONACCI NUMBERS

The Fibonacci numbers $\{F_n\}$ are defined by the recurrence:

$$F_1 = 1, \qquad F_2 = 1, \qquad F_{n+2} = F_n + F_{n+1}. \tag{1.3.12}$$

An exact formula is available (see page 186):

$$F_n = \frac{1}{\sqrt{5}}\left[\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n\right] \tag{1.3.13}$$

Note that F_n is the integer nearest to $\phi^n/\sqrt{5}$ as $n \to \infty$, where ϕ is the golden ratio.

n	F_n
1	1
2	1
3	2
4	3
5	5
6	8
7	13
8	21
9	34
10	55

n	F_n
11	89
12	144
13	233
14	377
15	610
16	987
17	1597
18	2584
19	4181
20	6765

n	F_n
21	10946
22	17711
23	28657
24	46368
25	75025
26	121393
27	196418
28	317811
29	514229
30	832040

n	F_n
31	1346269
32	2178309
33	3524578
34	5702887
35	9227465
36	14930352
37	24157817
38	39088169
39	63245986
40	102334155

1.3.8 SUMS OF POWERS OF INTEGERS

1. Define the sum of the first n k^{th}-powers

$$s_k(n) = 1^k + 2^k + \cdots + n^k = \sum_{m=1}^{n} m^k. \tag{1.3.14}$$

   (a) $s_k(n) = (k+1)^{-1}\left[B_{k+1}(n+1) - B_{k+1}(0)\right]$ (where the B_k are Bernoulli polynomials, see Section 1.3.5).

   (b) If $s_k(n) = \sum_{m=1}^{k+1} a_m n^{k-m+2}$, then

$$\begin{aligned} s_{k+1}(n) &= \left(\frac{k+1}{k+2}\right) a_1 n^{k+2} + \cdots + \left(\frac{k+1}{k}\right) a_3 n^k \\ &\quad + \cdots + \left(\frac{k+1}{2}\right) a_{k+1} n^2 + \left[1 - (k+1)\sum_{m=1}^{k+1} \frac{a_m}{k+3-m}\right] n. \end{aligned}$$

   (c) Note the specific values

$$\begin{aligned} s_1(n) &= 1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1) \\ s_2(n) &= 1^2 + 2^2 + 3^2 + \cdots + n^2 = \frac{1}{6}n(n+1)(2n+1) \\ s_3(n) &= 1^3 + 2^3 + 3^3 + \cdots + n^3 = \frac{1}{4}n^2(n+1)^2 = [s_1(n)]^2 \\ s_4(n) &= 1^4 + 2^4 + 3^4 + \cdots + n^4 = \frac{1}{5}(3n^2 + 3n - 1)s_2(n) \\ s_5(n) &= 1^5 + 2^5 + 3^5 + \cdots + n^5 = \frac{1}{12}n^2(n+1)^2(2n^2 + 2n - 1) \end{aligned}$$

2. $\displaystyle\sum_{k=1}^{n}(km-1) = \frac{1}{2}mn(n+1) - n$

3. $\displaystyle\sum_{k=1}^{n}(km-1)^2 = \frac{n}{6}\left[m^2(n+1)(2n+1) - 6m(n+1) + 6\right]$

4. $\displaystyle\sum_{k=1}^{n}(-1)^{k+1}(km-1) = \frac{(-1)^n}{4}\left[2 - (2n+1)m\right] + \frac{m-2}{4}$

5. $\displaystyle\sum_{k=1}^{n}(-1)^{k+1}(km-1)^2 = \tfrac{(-1)^{n+1}}{2}\left[n(n+1)m^2 - (2n+1)m + 1\right] + \tfrac{1-m}{2}$

n	$\sum_{k=1}^{n} k$	$\sum_{k=1}^{n} k^2$	$\sum_{k=1}^{n} k^3$	$\sum_{k=1}^{n} k^4$	$\sum_{k=1}^{n} k^5$
1	1	1	1	1	1
2	3	5	9	17	33
3	6	14	36	98	276
4	10	30	100	354	1300
5	15	55	225	979	4425

1.3.9 NEGATIVE INTEGER POWERS

Riemann's zeta function is defined to be $\zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n}$ (it is defined for Re $n > 1$ and extended to $\mathbb{C}$). Related functions are

$$\alpha(n) = \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^n}, \qquad \beta(n) = \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^n}, \qquad \gamma(n) = \sum_{k=0}^{\infty} \frac{1}{(2k+1)^n}.$$

Properties include:

1. $\alpha(n) = (1 - 2^{1-n})\zeta(n)$
2. $\zeta(2k) = \dfrac{(2\pi)^{2k}}{2(2k)!} |B_{2k}|$
3. $\gamma(n) = (1 - 2^{-n})\zeta(n)$
4. $\beta(2k+1) = \dfrac{(\pi/2)^{2k+1}}{2(2k)!} |E_{2k}|$
5. The series $\beta(1) = 1 - \frac{1}{3} + \frac{1}{5} - \cdots = \pi/4$ is known as Gregory's series.
6. Catalan's constant is $\mathbf{G} = \beta(2) \approx 0.915966$.
7. *Riemann hypothesis*: The non-trivial zeros of the Riemann zeta function (i.e., the $\{z_i\}$ that satisfy $\zeta(z_i) = 0$) lie on the *critical line* given by Re $z_i = \frac{1}{2}$. (The trivial zeros are $z = -2, -4, -6, \ldots$.)

n	$\zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n}$	$\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^n}$	$\sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^n}$	$\sum_{k=0}^{\infty} \frac{1}{(2k+1)^n}$
1	∞	0.6931471805	0.7853981633	∞
2	1.6449340668	0.8224670334	0.9159655941	1.2337005501
3	1.2020569032	0.9015426773	0.9689461463	1.0517997903
4	1.0823232337	0.9470328294	0.9889445517	1.0146780316
5	1.0369277551	0.9721197705	0.9961578281	1.0045237628
6	1.0173430620	0.9855510912	0.9986852222	1.0014470766
7	1.0083492774	0.9925938199	0.9995545079	1.0004715487
8	1.0040773562	0.9962330018	0.9998499902	1.0001551790
9	1.0020083928	0.9980942975	0.9999496842	1.0000513452
10	1.0009945751	0.9990395075	0.9999831640	1.0000170414

$\beta(1) = \pi/4$	$\zeta(2) = \pi^2/6$
$\beta(3) = \pi^3/32$	$\zeta(4) = \pi^4/90$
$\beta(5) = 5\pi^5/1536$	$\zeta(6) = \pi^6/945$
$\beta(7) = 61\pi^7/184320$	$\zeta(8) = \pi^8/9450$
$\beta(9) = 277\pi^9/8257536$	$\zeta(10) = \pi^{10}/93555$

1.3.10 INTEGER SEQUENCES

These sequences are in numerical order (disregarding leading zeros or ones).

1. 1, −1, −1, 0, −1, 1, −1, 0, 0, 1, −1, 0, −1, 1, 1, 0, −1, 0, −1, 0, 1, 1, −1, 0, 0, 1, 0, 0, −1, −1, −1, 0, 1, 1, 1, 0, −1, 1, 1, 0, −1, −1, −1, 0, 0, 1, −1, 0, 0, 0, 1, 0, −1, 0, 1, 0 **Möbius function $\mu(n)$, $n \geq 1$**
2. 1, 1, 0, 1, 1, 0, 0, 1, 1, 1, 0, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0, 2, 1, 0, 0, 1, 0, 0, 1, 0, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 0, 0, 0, 1, 2, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 1, 0, 0, 1, 2, 0, 0, 1, 0 **Number of ways of writing n as a sum of 2 squares, $n \geq 0$**
3. 1, 1, 1, 2, 1, 1, 1, 3, 2, 1, 1, 2, 1, 1, 1, 5, 1, 2, 1, 2, 1, 1, 1, 3, 2, 1, 3, 2, 1, 1, 1, 7, 1, 1, 1, 4, 1, 1, 1, 3, 1, 1, 1, 2, 2, 1, 1, 5, 2, 2, 1, 2, 1, 3, 1, 3, 1, 1, 1, 2, 1, 1, 2, 11, 1, 1, 1, 2 **Number of Abelian groups of order n, $n \geq 1$**
4. 1, 1, 1, 2, 1, 2, 1, 5, 2, 2, 1, 5, 1, 2, 1, 14, 1, 5, 1, 5, 2, 2, 1, 15, 2, 2, 5, 4, 1, 4, 1, 51, 1, 2, 1, 14, 1, 2, 2, 14, 1, 6, 1, 4, 2, 2, 1, 52, 2, 5, 1, 5, 1, 15, 2, 13, 2, 2, 1, 13, 1, 2, 4, 267 **Number of groups of order n, $n \geq 1$**
5. 0, 1, 1, 2, 1, 2, 2, 3, 1, 2, 2, 3, 2, 3, 3, 4, 1, 2, 2, 3, 2, 3, 3, 4, 2, 3, 3, 4, 3, 4, 4, 5, 1, 2, 2, 3, 2, 3, 3, 4, 2, 3, 3, 4, 3, 4, 4, 5, 2, 3, 3, 4, 3, 4, 4, 5, 3, 4, 4, 5, 4, 5, 5, 6, 1, 2, 2, 3, 2 **Number of 1's in binary expansion of n, $n \geq 0$**
6. 1, 2, 1, 2, 3, 6, 9, 18, 30, 56, 99, 186, 335, 630, 1161, 2182, 4080, 7710, 14532, 27594, 52377, 99858, 190557, 364722, 698870, 1342176, 2580795, 4971008 **Number of binary irreducible polynomials of degree n, or n-bead necklaces, $n \geq 0$**
7. 1, 2, 2, 3, 2, 4, 2, 4, 3, 4, 2, 6, 2, 4, 4, 5, 2, 6, 2, 6, 4, 4, 2, 8, 3, 4, 4, 6, 2, 8, 2, 6, 4, 4, 4, 9, 2, 4, 4, 8, 2, 8, 2, 6, 6, 4, 2, 10, 3, 6, 4, 6, 2, 8, 4, 8, 4, 4, 2, 12, 2, 4, 6, 7, 4, 8, 2, 6 **$d(n)$, the number of divisors of n, $n \geq 1$**
8. 1, 1, 2, 2, 3, 4, 5, 6, 8, 10, 12, 15, 18, 22, 27, 32, 38, 46, 54, 64, 76, 89, 104, 122, 142, 165, 192, 222, 256, 296, 340, 390, 448, 512, 585, 668, 760, 864, 982, 1113, 1260, 1426 **Number of partitions of n into distinct parts, $n \geq 1$**
9. 1, 1, 2, 2, 4, 2, 6, 4, 6, 4, 10, 4, 12, 6, 8, 8, 16, 6, 18, 8, 12, 10, 22, 8, 20, 12, 18, 12, 28, 8, 30, 16, 20, 16, 24, 12, 36, 18, 24, 16, 40, 12, 42, 20, 24, 22, 46, 16, 42 **Euler totient function $\phi(n)$: count numbers $\leq n$ and prime to n, for $n \geq 1$**
10. 1, 2, 3, 4, 5, 7, 8, 9, 11, 13, 16, 17, 19, 23, 25, 27, 29, 31, 32, 37, 41, 43, 47, 49, 53, 59, 61, 64, 67, 71, 73, 79, 81, 83, 89, 97, 101, 103, 107, 109, 113, 121, 125, 127, 128, 131 **Powers of prime numbers**
11. 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 60, 61, 67, 71, 73, 79, 83, 89, 97, 101, 103, 107, 109, 113, 127, 131, 137, 139, 149, 151, 157, 163, 167, 168, 173 **Orders of simple groups**
12. 1, 2, 3, 5, 7, 11, 15, 22, 30, 42, 56, 77, 101, 135, 176, 231, 297, 385, 490, 627, 792, 1002, 1255, 1575, 1958, 2436, 3010, 3718, 4565, 5604, 6842, 8349, 10143, 12310, 14883 **Number of partitions of n, $n \geq 1$**
13. 2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497, 86243, 110503, 132049, 216091, 756839, 859433 **Mersenne primes: n such that $2^n - 1$ is prime**
14. 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, 377, 610, 987, 1597, 2584, 4181, 6765, 10946, 17711, 28657, 46368, 75025, 121393, 196418, 317811, 514229, 832040, 1346269 **Fibonacci numbers: $F(n) = F(n-1) + F(n-2)$**

15. 1, 2, 3, 6, 10, 20, 35, 70, 126, 252, 462, 924, 1716, 3432, 6435, 12870, 24310, 48620, 92378, 184756, 352716, 705432, 1352078, 2704156, 5200300, 10400600, 20058300
Central binomial coefficients: $C(n, \lfloor n/2 \rfloor)$**,** $n \geq 1$

16. 1, 1, 1, 2, 3, 6, 11, 23, 47, 106, 235, 551, 1301, 3159, 7741, 19320, 48629, 123867, 317955, 823065, 2144505, 5623756, 14828074, 39299897, 104636890, 279793450
Number of trees with n **unlabeled nodes,** $n \geq 1$

17. 0, 0, 1, 1, 2, 3, 7, 21, 49, 165, 552, 2176, 9988
Number of prime knots with n **crossings,** $n \geq 1$

18. 0, 1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, 25, 26, 29, 32, 34, 36, 37, 40, 41, 45, 49, 50, 52, 53, 58, 61, 64, 65, 68, 72, 73, 74, 80, 81, 82, 85, 89, 90, 97, 98, 100, 101, 104, 106
Numbers that are sums of 2 squares

19. 1, 2, 4, 6, 10, 14, 20, 26, 36, 46, 60, 74, 94, 114, 140, 166, 202, 238, 284, 330, 390, 450, 524, 598, 692, 786, 900, 1014, 1154, 1294, 1460, 1626, 1828, 2030, 2268, 2506
Binary partitions (partitions of $2n$ **into powers of 2),** $n \geq 0$

20. 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048, 4096, 8192, 16384, 32768, 65536, 131072, 262144, 524288, 1048576, 2097152, 4194304, 8388608, 16777216, 33554432, 67108864
Powers of 2

21. 1, 1, 2, 4, 9, 20, 48, 115, 286, 719, 1842, 4766, 12486, 32973, 87811, 235381, 634847, 1721159, 4688676, 12826228, 35221832, 97055181, 268282855, 743724984, 2067174645
Number of rooted trees with n **unlabeled nodes,** $n \geq 1$

22. 1, 1, 2, 4, 9, 22, 59, 167, 490, 1486, 4639, 14805, 48107, 158808, 531469, 1799659, 6157068, 21258104, 73996100, 259451116, 951695102, 3251073303
Number of different scores in n**-team round-robin tournament,** $n \geq 1$

23. 1, 1, 2, 5, 12, 35, 108, 369, 1285, 4655, 17073, 63600, 238591, 901971, 3426576, 13079255, 50107909, 192622052, 742624232, 2870671950, 11123060678, 43191857688
Polyominoes with n **cells,** $n \geq 1$

24. 1, 1, 2, 5, 14, 38, 120, 353, 1148, 3527, 11622, 36627, 121622, 389560, 1301140, 4215748, 13976335, 46235800, 155741571, 512559185, 1732007938, 5732533570
Number of ways to fold a strip of n **blank stamps,** $n \geq 1$

25. 0, 1, 2, 9, 44, 265, 1854, 14833, 133496, 1334961, 14684570, 176214841, 2290792932, 32071101049, 481066515734, 7697064251745, 130850092279664
Derangements: permutations of n **elements with no fixed points**

26. 1, 3, 4, 7, 6, 12, 8, 15, 13, 18, 12, 28, 14, 24, 24, 31, 18, 39, 20, 42, 32, 36, 24, 60, 31, 42, 40, 56, 30, 72, 32, 63, 48, 54, 48, 91, 38, 60, 56, 90, 42, 96, 44, 84, 78, 72, 48, 124
$\sigma(n)$**, sum of the divisors of** n**,** $n \geq 1$

27. 1, 3, 4, 7, 9, 12, 13, 16, 19, 21, 25, 27, 28, 31, 36, 37, 39, 43, 48, 49, 52, 57, 61, 63, 64, 67, 73, 75, 76, 79, 81, 84, 91, 93, 97, 100, 103, 108, 109, 111, 112, 117, 121, 124, 127
Numbers of the form $x^2 + xy + y^2$

28. 1, 20, 400, 8902, 197281, 4865609, 119060324, 3195901860, 84998978956, 2439530234167
Number of possible chess games at the end of the n^{th} **ply.**

29. 1, 362, 130683, 47046243, 16889859363, 6046709375131
Number of Go games with n **moves.**

For information on these and hundreds of thousands of other sequences, see "The On-Line Encyclopedia of Integer Sequences," at `oeis.org`.

1.3.11 p-ADIC NUMBERS

Given a prime p, a non-zero rational number x can be written as $x = \frac{a}{b}p^n$ where n is an integer and p does not divide a or b. Define the p-adic norm of x as $|x|_p = p^{-n}$ and also define $|0|_p = 0$. The p-adic norm has the properties:

1. $|x|_p \geq 0$ for all non-negative rational numbers x
2. $|x|_p = 0$ if and only if $x = 0$
3. For all non-negative rational numbers x and y
   (a) $|xy|_p = |x|_p|y|_p$
   (b) $|x + y|_p \leq \max(|x|_p, |y|_p) \leq |x|_p + |y|_p$

Note the *product formula*: $|x| \prod_{p\in\{2,3,5,7,11,\dots\}} |x|_p = 1$.

Let $\mathbb{Q}_p$ be the topological completion of $\mathbb{Q}$ with respect to $|\cdot|_p$. Then $\mathbb{Q}_p$ is the field of *p-adic numbers*. The elements of $\mathbb{Q}_p$ can be viewed as infinite series: the series $\sum_{n=0}^{\infty} a_n$ converges to a point in $\mathbb{Q}_p$ if and only if $\lim_{n\to\infty} |a_n|_p = 0$.

EXAMPLE The number $\frac{140}{297} = 2^2 \cdot 3^{-3} \cdot 5 \cdot 7 \cdot 11^{-1}$ has the different p-adic norms:

- $\left|\frac{140}{297}\right|_2 = 2^{-2} = \frac{1}{4}$
- $\left|\frac{140}{297}\right|_5 = 5^{-1} = \frac{1}{5}$
- $\left|\frac{140}{297}\right|_{11} = 11^1 = 11$
- $\left|\frac{140}{297}\right|_3 = 3^3 = 27$
- $\left|\frac{140}{297}\right|_7 = 7^{-1} = \frac{1}{7}$

1.3.12 DE BRUIJN SEQUENCES

A sequence of length q^n over an alphabet of size q is a *de Bruijn sequence* if every possible n-tuple occurs in the sequence (allowing wraparound to the start of the sequence). There are de Bruijn sequences for any q and n. (In fact, there are $q!^{q^{n-1}}/q!$ distinct sequences.) The table below contains some small examples.

q	n	Length	Sequence
2	1	2	01
2	2	4	0110
2	3	8	01110100
2	4	16	0101001101111000
3	2	9	001220211
3	3	27	000100201101202102211121222
4	2	16	0011310221203323

1.4 INTERVAL ANALYSIS

1. Definitions

   (a) An *interval* x is a subset of the real line:

   $$x = [\underline{x}, \overline{x}] = \{z \in \mathbb{R} \mid \underline{x} \le z \le \overline{x}\}.$$

   (b) A *thin interval* is a real number: x is thin if $\underline{x} = \overline{x}$

   (c) $\operatorname{mid}(x) = \dfrac{\overline{x} + \underline{x}}{2}$

   (d) $\operatorname{rad}(x) = \dfrac{\overline{x} - \underline{x}}{2}$

   (e) $|x| = \operatorname{mag}(x) = \max_{z \in x} |z|$

   (f) $\langle x \rangle = \operatorname{mig}(x) = \min_{z \in x} |z|$

   The MATLAB® package INTLAB performs interval computations.

2. Interval arithmetic rules

Operation	Rule
$x + y$	$[\underline{x} + \underline{y}, \overline{x} + \overline{y}]$
$x - y$	$[\underline{x} - \overline{y}, \overline{x} - \underline{y}]$
xy	$[\min(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y}), \max(\underline{x}\underline{y}, \underline{x}\overline{y}, \overline{x}\underline{y}, \overline{x}\overline{y})]$
$\frac{x}{y}$	$\left[\min\left(\frac{\underline{x}}{\underline{y}}, \frac{\underline{x}}{\overline{y}}, \frac{\overline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right), \max\left(\frac{\underline{x}}{\underline{y}}, \frac{\underline{x}}{\overline{y}}, \frac{\overline{x}}{\underline{y}}, \frac{\overline{x}}{\overline{y}}\right)\right]$ if $0 \notin y$

3. Interval arithmetic properties

Property	$+$ and $-$	$*$ and $/$
commutative	$x + y = y + x$	$xy = yx$
associative	$x + (y + z) = (x + y) + z$	$x(yz) = (xy)z$
identity elements	$0 + x = x + 0 = x$	$1 * y = y * 1 = y$
sub-distributivity	$x(y \pm z) \subseteq xy \pm xz$ (equality holds if x is thin)	
sub-cancellation	$x - y \subseteq (x + z) - (y + z)$	$\frac{x}{y} \subseteq \frac{xz}{yz}$
	$0 \in x - x$	$1 \in \frac{y}{y}$

4. Examples

   (a) $[1, 2] + [-2, 1] = [-1, 3]$

   (b) $[1, 2] - [1, 2] = [-1, 1]$

   (c) $[1, 2] * [-2, 1] = [-4, 2]$

   (d) $[1, 2]/[1, 2] = \left[\frac{1}{2}, 2\right]$

   (e) If $f(a, b; x) = ax + b$ then (when $a = [1, 2]$, $b = [5, 7]$, and $x = [2, 3]$);

   $$f([1, 2], [5, 7]; [2, 3]) = [1, 2] \cdot [2, 3] + [5, 7] = [1 \cdot 2, 2 \cdot 3] + [5, 7] = [7, 13]$$

1.5 FRACTAL ARITHMETIC

Given a real-valued bijection f with $f(0) = 0$ and $f(1) = 1$ define

$$x \oplus y = f^{-1}\left(f(x) + f(y)\right) \qquad x \ominus y = f^{-1}\left(f(x) - f(y)\right)$$
$$x \odot y = f^{-1}\left(f(x)f(y)\right) \qquad x \oslash y = f^{-1}\left(f(x)/f(y)\right) \tag{1.5.1}$$

For example, if $f(x) = x^q$ then $x \odot y = xy$ and $x \oplus y = (x^q + y^q)^{1/q}$. Note that the following hold:

- Associativity: $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ and $(x \odot y) \odot z = x \odot (y \odot z)$
- Commutativity: $x \oplus y = y \oplus x$ and $x \odot y = y \odot x$
- Distributivity: $(x \oplus y) \odot z = (x \odot z) \oplus (y \odot z)$

The elements 0 and 1 satisfy $0 \oplus x = x$ and $1 \odot x = x$.

1. $x \ominus x = 0$ and $x \oslash x = 1$.
2. In general, $x \oplus x \neq 2 \odot x$.
3. If $0 \ominus x$ exists, it is denoted as $\ominus x$

Define derivative and integration operations:

$$\frac{d_f A(x)}{d_f x} = \lim_{h \to 0} \left(A(x \oplus h) \ominus A(x)\right) \oslash h$$
$$\int_a^b F_f(x) \odot d_f x = f^{-1}\left(\int_{f(a)}^{f(b)} F(y)\, dy\right) \tag{1.5.2}$$

so that

1. $\dfrac{d_f\left(A(x) \odot B(x)\right)}{d_f x} = \dfrac{d_f\left(A(x)\right)}{d_f x} \odot B(x) \oplus A(x) \odot \dfrac{d_f\left(B(x)\right)}{d_f x}$
2. $\dfrac{d_f\left(A(x) \oplus B(x)\right)}{d_f x} = \dfrac{d_f\left(A(x)\right)}{d_f x} \oplus \dfrac{d_f\left(B(x)\right)}{d_f x}$
3. $\dfrac{d_f\left(A[B(x)]\right)}{d_f x} = \dfrac{d_f\left(A[B(x)]\right)}{d_f B(x)} \odot \dfrac{d_f\left(B(x)\right)}{d_f x}$
4. $\displaystyle\int_a^b \frac{d_f A(x)}{d_f x} \odot d_f x = A(b) \ominus A(a)$
5. $\displaystyle\frac{d_f}{d_f x} \int_a^b A(x') \odot d_f x' = A(x)$
6. $\dfrac{d_f F_f(x)}{d_f x} = f^{-1}\left(F'(f(x))\right)$

Special functions Given a function F define F_f as $F_f(x) = f^{-1}\left(F\left(f(x)\right)\right)$.

1. The $\exp_f$ function satisfies $\exp_f(x \oplus y) = \exp_f x \odot \exp_f y$ and is the unique solution to the differential equation $\frac{d_f A(x)}{d_f x} = A(x)$ with $A(0) = 1$.
2. The $\ln_f$ function satisfies $\ln_f(x \odot y) = \ln_f x \oplus \ln_f y$

1.6 MAX-PLUS ALGEBRA

The max-plus algebra is an algebraic structure with real numbers and two operations (called "plus" and "times"); "plus" is the operation of taking a maximum, and "times" is the "standard addition" operation. Mathematically:

The max-plus semiring $R_{\max}$ is the set $R \cup \{-\infty\}$ with two operations called "plus" ($\oplus$) and "times" ($\otimes$) defined as follows:

- $a \oplus b = \max(a, b)$
- $a \otimes b = a + b$

For a and b in $R_{\max}$ the following laws apply:

1. Associativity
   - $(a \oplus b) \oplus c = a \oplus (b \oplus c)$
   - $(a \otimes b) \otimes c = a \otimes (b \otimes c)$
2. Commutativity
   - $a \oplus b = b \oplus a$
   - $a \otimes b = b \otimes a$
3. Distributivity: $(a \oplus b) \otimes c = a \otimes c \oplus b \otimes c$
4. Idempotency of $\oplus$: $a \oplus a = a$

Special elements are the "zero element" $\epsilon = -\infty$ and and the "unit element" $e = 0$. These have the properties:

- $\epsilon \oplus a = a$
- $\epsilon \otimes a = \epsilon$
- $e \otimes a = a$

Let A and B be matrices and let $[C]_{ij}$ be the ij^{th} element of matrix C. Then

- $[A \oplus B]_{ij} = [A]_{ij} \oplus [B]_{ij} = \max([A]_{ij}, [B]_{ij})$ when A and B have the same size
- $[A \otimes B]_{ij} = \bigoplus_{k=1}^{p}([A]_{ik} \otimes [B]_{kj}) = \max([A]_{i1} + [B]_{1j}, \ldots, [A]_{ip} + [B]_{pj})$ when A has p columns and B has p rows.

Notes

1. $R_{\max}$ is not a group since not all elements have an additive inverse. (Only one element has an additive inverse, it is $-\infty$.)
2. The equation $a \oplus x = b$ need not have a unique solution.
   - If $a < b$, then $x = b$
   - If $a = b$, then x can be any value with $x \leq b$
   - If $a > b$, then there is no solution
3. In $R_{\max}$ matrix multiplication is associative.
4. Defining exponential as $a^n = \underbrace{a \otimes a \otimes \cdots \otimes a}_{n \text{ times}} = \underbrace{a + a + \cdots + a}_{n \text{ times}} = na$ we find $a^x \otimes a^y = a^{x+y}$ and $(a^x)^y = a^{xy}$, as in ordinary arithmetic.
5. The *Kleene star* of the matrix A is the matrix $A^* = I + A + A^2 + \cdots$.
6. The *completed max-plus semiring* $\overline{R}_{\max}$ is the same as $R_{\max}$ with the additional element $+\infty$ and the convention $(-\infty) + (+\infty) = (+\infty) + (-\infty) = -\infty$

EXAMPLE

Let $A = \begin{bmatrix} 10 & -\infty \\ 5 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} 8 & 2 \\ 7 & 0 \end{bmatrix}$. Then

- Scalar multiplication of a matrix $5 \otimes A = \begin{bmatrix} 5 \otimes 10 & 5 \otimes (-\infty) \\ 5 \otimes 5 & 5 \otimes 3 \end{bmatrix} = \begin{bmatrix} 15 & -\infty \\ 10 & 8 \end{bmatrix}$
- Matrix addition $A \oplus B = \begin{bmatrix} 10 \oplus 8 & -\infty \oplus 2 \\ 5 \oplus 7 & 3 \oplus 0 \end{bmatrix} = \begin{bmatrix} 10 & 2 \\ 7 & 3 \end{bmatrix}$
- Matrix multiplication

$$\begin{aligned} A \otimes B &= \begin{bmatrix} 10 \otimes 8 \oplus (-\infty) \otimes 7 & 10 \otimes 2 \oplus (-\infty) \otimes 0 \\ 5 \otimes 8 \oplus 3 \otimes 7 & 5 \otimes 2 \oplus 3 \otimes 0 \end{bmatrix} \\ &= \begin{bmatrix} 18 \oplus (-\infty) & 12 \oplus (-\infty) \\ 13 \oplus 10 & 7 \oplus 3 \end{bmatrix} = \begin{bmatrix} 18 & 12 \\ 13 & 7 \end{bmatrix} \end{aligned}$$

1.7 COUPLED-ANALOGUES OF FUNCTIONS

1. The coupled-logarithm, for $x > 0$, is $\ln_\kappa(x) = \dfrac{x^\kappa - 1}{\kappa}$

   Note that:

$$\lim_{\kappa \to 0} \ln_\kappa(x) = \ln x \qquad \ln_\kappa\left(x^a\right) = a \ln_{a\kappa}(x) \qquad \ln_\kappa\left(e_\kappa^x\right) = x \tag{1.7.1}$$

2. The coupled-exponential is $e_\kappa^x = \begin{cases} [1 + \kappa x]^{1/\kappa} & \text{when } 1 + \kappa x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

   Note that:

$$\begin{aligned} &\lim_{\kappa \to 0} e_\kappa^x = e^x \qquad \left(e_\kappa\right)^a = e_{\kappa/a}^{ax} \qquad \exp_\kappa\left(\ln_\kappa(x)\right) = x \\ &\frac{d}{dx} e_\kappa^{ax} = a \exp_{\frac{\kappa}{1-\kappa}}[(1-\kappa)ax] \qquad \text{when } \kappa \neq 1 \\ &\int e_\kappa^{ax}\, dx = \frac{1}{a(1+\kappa)} \exp_{\frac{\kappa}{1+\kappa}}[(1+\kappa)ax] + c_1 \quad \text{when } \kappa \neq -1 \end{aligned} \tag{1.7.2}$$

1.7.1 COUPLED-OPERATIONS

1. Coupled-addition is defined by: $x \oplus_\kappa y = x + y + \kappa xy$. Note that:

$$\begin{aligned} e_\kappa^x e_\kappa^y &= e_\kappa^{x \oplus_\kappa y} \\ \ln_\kappa(xy) &= \ln_\kappa(x) \oplus \ln_\kappa(y) \\ x \oplus y \oplus z &= x + y + z + \kappa(xy + xz + yz) + \kappa^2 xyz \end{aligned} \tag{1.7.3}$$

2. Coupled-subtraction is defined by: $x \ominus_\kappa y = (x - y)/(1 + \kappa y)$
3. Coupled-division is defined by: $x \oslash_\kappa y = \left(x^q - y^q + 1\right)^{1/\kappa}$
4. Coupled-multiplication is defined by: $x \otimes_\kappa y = \left(x^\kappa + y^\kappa - 1\right)^{1/\kappa}$. Note that:

$$e_\kappa^x \otimes_\kappa e_\kappa^y = e_\kappa^{x+y} \qquad \ln_\kappa(x \otimes_\kappa y) = \ln_\kappa(x) + \ln_\kappa(y)$$

$$x_1 \otimes_\kappa x_1 \otimes_\kappa \cdots \otimes_\kappa x_n = \prod_{i=1}^{n}{}_\kappa\, x_i = \left(x_1^\kappa + x_2^\kappa + \cdots + x_n^\kappa - n + 1\right)^{1/\kappa}$$

1.8 NUMBER THEORY

Divisibility The notation "$a|b$" means that the number a evenly divides the number b. That is, the ratio $\frac{b}{a}$ is an integer.

1.8.1 CONGRUENCES

1. If the integers a and b leave the same remainder when divided by the number n, then a and b are *congruent* modulo n. This is written $a \equiv b \pmod{n}$.
2. If the congruence $x^2 \equiv a \pmod{p}$ has a solution, then a is a *quadratic residue* of p. Otherwise, a is a *quadratic non-residue* of p.
   (a) Let p be an odd prime. *Legendre's symbol* $\left(\frac{a}{p}\right)$ has the value $+1$ if a is a quadratic residue of p, and the value -1 if a is a quadratic non-residue of p. This can be written $\left(\frac{a}{p}\right) \equiv a^{(p-1)/2} \pmod{p}$.
   (b) The *Jacobi symbol* generalizes the Legendre symbol to non-prime moduli. If $n = \prod_{i=1}^{k} p_i^{b_i}$ then the Jacobi symbol can be written in terms of the Legendre symbol as follows

$$\left(\frac{a}{n}\right) = \prod_{i=1}^{k} \left(\frac{a}{p_i}\right)^{b_i}. \tag{1.8.1}$$

3. An *exact covering* sequence is a set of non-negative ordered pairs $\{(a_i, b_i)\}_{i=1,\ldots,k}$ such that every non-negative integer n satisfies $n \equiv a_i \pmod{b_i}$ for exactly one i. An exact covering sequence satisfies

$$\sum_{i=1}^{k} \frac{x^{a_i}}{1 - x^{b_i}} = \frac{1}{1-x}. \tag{1.8.2}$$

   For example, every positive integer n is either congruent to 1 mod 2, or 0 mod 4, or 2 mod 4. Hence, the three pairs $\{(1,2), (0,4), (2,4)\}$ of residues and moduli *exactly cover* the positive integers. Note that

$$\frac{x}{1-x^2} + \frac{1}{1-x^4} + \frac{x^2}{1-x^4} = \frac{1}{1-x}. \tag{1.8.3}$$

4. Carmichael numbers are composite numbers $\{n\}$ that satisfy $a^{n-1} \equiv 1 \pmod{n}$ for every a (with $1 < a < n$) that is relatively prime to n. There are infinitely many Carmichael numbers. Every Carmichael number has at least three prime factors. If $n = \prod_i p_i$ is a Carmichael number, then $(p_i - 1)$ divides $(n-1)$ for each i.
   There are 43 Carmichael numbers less than 10^6 and 105,212 less than 10^{15}. The Carmichael numbers less than ten thousand are 561, 1105, 1729, 2465, 2821, 6601, and 8911.

1.8.1.1 Properties of congruences

1. If $a \equiv b \pmod{n}$, then $b \equiv a \pmod{n}$.
2. If $a \equiv b \pmod{n}$, and $b \equiv c \pmod{n}$, then $a \equiv c \pmod{n}$.
3. If $a \equiv a' \pmod{n}$, and $b \equiv b' \pmod{n}$, then $a \pm b \equiv a' \pm b' \pmod{n}$.
4. If $a \equiv a' \pmod{n}$, then $a^2 \equiv (a')^2 \pmod{n}$, $a^3 \equiv (a')^3 \pmod{n}$, etc.
5. If $\text{GCD}(k, m) = d$, then the congruence $kx \equiv n \pmod{m}$ is solvable if and only if d divides n. It then has d solutions.
6. If p is a prime, then $a^p \equiv a \pmod{p}$.
7. If p is a prime, and p does not divide a, then $a^{p-1} \equiv 1 \pmod{p}$.
8. If $\text{GCD}(a, m) = 1$, then $a^{\phi(m)} \equiv 1 \pmod{m}$. (See Section 1.8.12 for $\phi(m)$.)
9. If p is an odd prime and a is not a multiple of p, then Wilson's theorem states $(p-1)! \equiv -\left(\frac{a}{p}\right) a^{(p-1)/2} \pmod{p}$.
10. If p and q are odd primes, then Gauss' *law of quadratic reciprocity* states that $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{(p-1)(q-1)/4}$. Therefore, if a and b are relatively prime odd integers and $b \geq 3$, then $\left(\frac{a}{b}\right) = (-1)^{(a-1)(b-1)/4}\left(\frac{b}{a}\right)$.
11. The number -1 is a quadratic residue of primes of the form $4k+1$ and a non-residue of primes of the form $4k+3$. That is
$$\left(\frac{-1}{p}\right) = (-1)^{(p-1)/2} = \begin{cases} +1 & \text{when } p \equiv 1 \pmod{4} \\ -1 & \text{when } p \equiv 3 \pmod{4} \end{cases}$$
12. The number 2 is a quadratic residue of primes of the form $8k \pm 1$ and a non-residue of primes of the form $8k \pm 3$. That is
$$\left(\frac{2}{p}\right) = (-1)^{(p^2-1)/8} = \begin{cases} +1 & \text{when } p \equiv \pm 1 \pmod{8} \\ -1 & \text{when } p \equiv \pm 3 \pmod{8} \end{cases}$$

1.8.2 CHINESE REMAINDER THEOREM

Let $m_1, m_2, \ldots, m_r$ be pairwise relatively prime integers. The system of congruences

$$\begin{aligned} x &\equiv a_1 \pmod{m_1} \\ x &\equiv a_2 \pmod{m_2} \\ &\vdots \\ x &\equiv a_r \pmod{m_r} \end{aligned} \tag{1.8.4}$$

has a unique solution modulo $M = m_1 m_2 \cdots m_r$. This unique solution can be written as

$$x = a_1 M_1 y_1 + a_2 M_2 y_2 + \cdots + a_r M_r y_r \tag{1.8.5}$$

where $M_k = M/m_k$, and y_k is the inverse of M_k (modulo m_k).

EXAMPLE For the system of congruences

$$\begin{aligned} x &\equiv 1 \pmod 3 \\ x &\equiv 2 \pmod 5 \\ x &\equiv 3 \pmod 7 \end{aligned}$$

we have $M = 3{\cdot}5{\cdot}7 = 105$ with $M_1 = 35$, $M_2 = 21$, and $M_3 = 15$. The equation for y_1 is $M_1 y_1 = 35 y_1 \equiv 1 \pmod 3$ with solution $y_1 \equiv 2 \pmod 3$. Likewise, $y_2 \equiv 1 \pmod 5$ and $y_3 \equiv 1 \pmod 7$. This results in $x = 1\cdot 35\cdot 2 + 2\cdot 21\cdot 1 + 3\cdot 15\cdot 1 \equiv 52 \pmod{105}$.

1.8.3 CONTINUED FRACTIONS

The symbol $[a_0, a_1, \dots, a_N]$, with $a_i > 0$, represents the simple *continued fraction*,

$$[a_0, a_1, \dots, a_N] = a_0 + \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cfrac{1}{a_4 + \cfrac{\cdots}{\cdots + \cfrac{1}{a_N}}}}}}. \tag{1.8.6}$$

The n^{th} *convergent* (with $0 < n < N$) of $[a_0, a_1, \dots, a_N]$ is defined to be $[a_0, a_1, \dots, a_n]$. If $\{p_n\}$ and $\{q_n\}$ are defined by

$$\begin{aligned} p_0 &= a_0, \quad & p_1 &= a_1 a_0 + 1, \quad & p_n &= a_n p_{n-1} + p_{n-2} \quad (2 \le n \le N) \\ q_0 &= 1, & q_1 &= a_1, & q_n &= a_n q_{n-1} + q_{n-2} \quad (2 \le n \le N) \end{aligned}$$

then $[a_0, a_1, \dots, a_n] = p_n / q_n$. The continued fraction is convergent if and only if the infinite series $\sum_i^\infty a_i$ is divergent.

If the positive rational number x can be represented by a simple continued fraction with an odd (even) number of terms, then it is also representable by one with an even (odd) number of terms. (Specifically, if $a_n = 1$ then $[a_0, a_1, \dots, a_{n-1}, 1] = [a_0, a_1, \dots, a_{n-1} + 1]$, and if $a_n \ge 2$, then $[a_0, a_1, \dots, a_n] = [a_0, a_1, \dots, a_n - 1, 1]$.) Aside from this indeterminacy, the simple continued fraction of x is unique. The error in approximating by a convergent is bounded by

$$\left| x - \frac{p_n}{q_n} \right| \le \frac{1}{q_n q_{n+1}} < \frac{1}{q_n^2}. \tag{1.8.7}$$

The algorithm for finding a continued fraction expansion of a number is to remove the integer part of the number (this becomes a_i), take the reciprocal, and repeat.

For example, for the number π:

$$\begin{aligned}
&\beta_0 = \pi \approx 3.14159 && a_0 = \lfloor \beta_0 \rfloor = 3\\
&\beta_1 = 1/(\beta_0 - a_0) \approx 7.062 && a_1 = \lfloor \beta_1 \rfloor = 7\\
&\beta_2 = 1/(\beta_1 - a_1) \approx 15.997 && a_2 = \lfloor \beta_2 \rfloor = 15\\
&\beta_3 = 1/(\beta_2 - a_2) \approx 1.0034 && a_3 = \lfloor \beta_3 \rfloor = 1\\
&\beta_4 = 1/(\beta_3 - a_3) \approx 292.6 && a_4 = \lfloor \beta_4 \rfloor = 292
\end{aligned}$$

Since $\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, \dots]$ approximations to π may be found from the convergents: $\frac{22}{7} \approx 3.14\underline{2}$, $\frac{333}{106} \approx 3.1415\underline{0}$, $\frac{355}{113} \approx 3.141592\underline{9}$.

Since $e = [2, 1, 2, 1, 1, 4, 1, 1, 6, \dots, 1, 1, 2n, \dots]$ approximations to e may be found from the convergents: $\frac{8}{3} \approx 2.\underline{6}$, $\frac{11}{4} \approx 2.7\underline{5}$, $\frac{19}{7} \approx 2.71\underline{4}$, $\frac{87}{32} \approx 2.718\underline{7}, \dots$.

A *periodic continued fraction* is an infinite continued fraction in which $a_l = a_{l+k}$ for all $l \geq L$. The set of partial quotients $a_L, a_{L+1}, \dots, a_{L+k-1}$ is the *period*. A periodic continued fraction may be written as

$$[a_0, a_1, \dots, a_{L-1}, \overline{a_L, a_{L+1}, \dots, a_{L+k-1}}]\,. \tag{1.8.8}$$

For example,

$$\begin{array}{llll}
\sqrt{2} = [1, \overline{2}] & \sqrt{6} = [2, \overline{2, 4}] & \sqrt{10} = [3, \overline{6}] & \sqrt{14} = [3, \overline{1, 2, 1, 6}]\\
\sqrt{3} = [1, \overline{1, 2}] & \sqrt{7} = [2, \overline{1, 1, 1, 4}] & \sqrt{11} = [3, \overline{3, 6}] & \sqrt{15} = [3, \overline{1, 6}]\\
\sqrt{4} = [2] & \sqrt{8} = [2, \overline{1, 4}] & \sqrt{12} = [3, \overline{2, 6}] & \sqrt{16} = [4]\\
\sqrt{5} = [2, \overline{4}] & \sqrt{9} = [3] & \sqrt{13} = [3, \overline{1, 1, 1, 1, 6}] & \sqrt{17} = [4, \overline{8}]
\end{array}$$

If $x = [\overline{b, a}]$ then $x = \frac{1}{2}(b + \sqrt{b^2 + \frac{4b}{a}})$. For example, $[\overline{1}] = [\overline{1, 1}] = (1 + \sqrt{5})/2$, $[\overline{2}] = [\overline{2, 2}] = 1 + \sqrt{2}$, and $[\overline{2, 1}] = 1 + \sqrt{3}$.

Functions can be represented as continued fractions. Using the notation

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cfrac{a_4}{b_4 + \dots}}}} \equiv b_0 + \frac{a_1}{b_1+}\,\frac{a_2}{b_2+}\,\frac{a_3}{b_3+}\,\frac{a_4}{b_4+} \dots \tag{1.8.9}$$

we have (allowable values of z may be restricted in the following)

(a) $\ln(1 + z) = \frac{z}{1+}\,\frac{z}{2+}\,\frac{z}{3+}\,\frac{4z}{4+}\,\frac{4z}{5+}\,\frac{9z}{6+} \dots$

(b) $e^z = \frac{1}{1-}\,\frac{z}{1+}\,\frac{z}{2-}\,\frac{z}{3+}\,\frac{z}{2-}\,\frac{z}{5+}\,\frac{z}{2-} \dots = 1 + \frac{z}{1-}\,\frac{z}{2+}\,\frac{z}{3-}\,\frac{z}{2+}\,\frac{z}{5-}\,\frac{z}{2+}\,\frac{z}{7-} \dots$

(c) $\tan z = \frac{z}{1-}\,\frac{z^2}{3-}\,\frac{z^2}{5-}\,\frac{z^2}{7-} \dots$

(d) $\tanh z = \frac{z}{1+}\,\frac{z^2}{3+}\,\frac{z^2}{5+}\,\frac{z^2}{7+} \dots$

1.8.4 DIOPHANTINE EQUATIONS

A *diophantine equation* is one whose solutions are integers.

1. Fermat's last theorem states that there are no integer solutions to $x^n + y^n = z^n$, when $n > 2$. This was proved by Andrew Wiles in 1995.
2. The similar, but slightly different, equation $x^n + y^n = z^{n+1}$ has parametric solutions given by $x = a\left(a^m + b^m\right)^u$, $y = b\left(a^m + b^m\right)^u$, $z = \left(a^m + b^m\right)^v$ where $\gcd(m, n) = 1$ and $nv - mu = 1$.
3. The Hurwitz equation, $x_1^2 + x_2^2 + \cdots + x_n^2 = a x_1 x_2 \cdots x_n$, has no integer solutions for $a > n$.
4. Bachet's equation, $y^2 = x^3 + k$, has no solutions when k is: -144, -105, -78, -69, -42, -34, -33, -31, -24, -14, -5, 7, 11, 23, 34, 45, 58, 70.
5. Ramanujan's "square equation," $2^n = 7 + x^2$, has solutions for $n = 3, 4, 5, 7$, and 15 corresponding to $x = 1, 3, 5, 11$, and 181.
6. For given k and m consider $a_1^k + a_2^k + \cdots + a_m^k = b_1^k + b_2^k + \cdots + b_n^k$ with $a_1 \geq a_2 \geq \cdots \geq a_m$, $b_1 \geq b_2 \geq \cdots \geq b_n$, $a_1 > 1$, and $m \leq n$. For example:

$$\begin{aligned} 5^2 &= 4^2 + 3^2 \\ 12^3 + 1^3 &= 10^3 + 9^3 \\ 158^4 + 59^4 &= 134^4 + 133^4 \\ 422481^4 &= 414560^4 + 217519^4 + 95800^4 \\ 144^5 &= 133^5 + 110^5 + 84^5 + 27^5 \end{aligned}$$

   Given k and m the least value of n for which a solution is known is as follows:

	$m = 1$	2	3	4	5	6
$k = 2$	2					
3	3	2				
4	3	2				
5	4	3				
6	7	5	3			
7	7	6	5	4		
8	8	7	5	4		
9	10	8	8	6	5	
10	12	12	11	9	7	6

7. Cannonball problem: If n^2 cannonballs can be stacked to form a square pyramid of height k, what are n and k? The Diophantine equation is $\sum_{i=1}^{k} i^2 = \frac{1}{6}k(k+1)(2k+1) = n^2$ with solutions $(k, n) = (1, 1)$ and $(k, n) = (24, 70)$.
8. The Euler equation $2^n = 7x^2 + y^2$ has a unique solution, with x and y odd, for $n \geq 3$:

$$x = \frac{2^{n/2+1}}{\sqrt{7}} \left|\sin \alpha_n\right| \qquad y = 2^{n/2+1} \left|\cos \alpha_n\right|$$

   where $\alpha_n = \left([n-2] \tan^{-1} \sqrt{7}\right)$.
   The solutions are $(n, x, y) = \{(3, 1, 1), (4, 1, 3), (5, 1, 5), (6, 3, 1), \ldots\}$

9. Apart from the trivial solutions (with $x = y = 0$ or $x = u$), the general solution to the equation $x^3 + y^3 = u^3 + v^3$ is given parametrically by

$$x = \lambda \left[1 - (a - 3b)(a^2 + 3b^2)\right] \qquad y = \lambda \left[(a + 3b)(a^2 + 3b^2) - 1\right]$$
$$u = \lambda \left[(a + 3b) - (a^2 + 3b^2)^2\right] \qquad v = \lambda \left[(a^2 + 3b^2)^2 - (a - 3b)\right]$$

where $\{\lambda, a, b\}$ are any rational numbers except that $\lambda \neq 0$.

10. A parametric solution to $x^4 + y^4 = u^4 + v^4$ is given by

$$\begin{aligned} x &= a^7 + a^5b^2 - 2a^3b^4 + 3a^2b^5 + ab^6 \\ y &= a^6b - 3a^5b^2 - 2a^4b^3 + a^2b^5 + b^7 \\ u &= a^7 + a^5b^2 - 2a^3b^4 - 3a^2b^5 + ab^6 \\ v &= a^6b + 3a^5b^2 - 2a^4b^3 + a^2b^5 + b^7 \end{aligned}$$

11. Parametric solutions to the equation $(A^2+B^2)(C^2+D^2) = E^2+F^2$ are given by the Fibonacci identity $(a^2+b^2)(c^2+d^2) = (ac\pm bd)^2+(bc\mp ad)^2 = e^2+f^2$. A similar identity is the Euler four-square identity $(a_1^2 + a_2^2 + a_3^2 + a_4^2)(b_1^2 + b_2^2 + b_3^2 + b_4^2) = (a_1b_1 - a_2b_2 - a_3b_3 - a_4b_4)^2 + (a_1b_2 + a_2b_1 + a_3b_4 - a_4b_3)^2 + (a_1b_3 - a_2b_4 + a_3b_1 + a_4b_2)^2 + (a_1b_4 + a_2b_3 - a_3b_2 - a_4b_1)^2$.

12. The only integer solutions to the equation $x^y = y^x$ are $(2, 4)$ and $x = y$. Non-integral solutions are given by $\left\{x = \left(1 + \frac{1}{u}\right)^u, y = \left(1 + \frac{1}{u}\right)^{u+1}\right\}$. Setting $u = 2, 3, \dots$ yields the rational solutions $\left(\frac{9}{4}, \frac{27}{8}\right), \left(\frac{64}{27}, \frac{256}{81}\right), \dots$

1.8.4.1 Pythagorean triples

If the positive integers A, B, and C satisfy $A^2+B^2 = C^2$, then the triplet (A, B, C) is a *Pythagorean triple*. A right triangle can be constructed with sides of length A and B and a hypotenuse of C. There are infinitely many Pythagorean triples.

A general solution to $A^2 + B^2 = C^2$, with $\text{GCD}(A, B) = 1$ and A even, is

$$A = 2xy, \qquad B = x^2 - y^2, \qquad C = x^2 + y^2, \tag{1.8.10}$$

where x and y are relatively prime integers of opposite parity (i.e., one is even and the other is odd) with $x > y > 0$. The table below left shows some Pythagorean triples with the associated (x, y) values.

x	y	$A = 2xy$	$B = x^2 - y^2$	$C = x^2 + y^2$
2	1	4	3	5
4	1	8	15	17
6	1	12	35	37
8	1	16	63	65
10	1	20	99	101
3	2	12	5	13
5	2	20	21	29

n	p	q	A	B	C
6	1	18	7	24	25
6	2	9	8	15	17
6	3	6	9	12	15

A different general solution is obtained by factoring even squares as $n^2 = 2pq$. Here $A = n+p$, $B = n+q$, and $C = n+p+q$. The table above right shows the (p, q) and (A, B, C) values obtained from the factorizations $36 = 2 \cdot 1 \cdot 18 = 2 \cdot 2 \cdot 9 = 2 \cdot 3 \cdot 6$.

1.8.4.2 Pell's equation

Pell's equation is $x^2 - dy^2 = 1$. The solutions, integral values of (x, y), arise from continued fraction convergents of $\sqrt{d}$. If (x, y) is the least positive solution to Pell's equation (with d square-free), then every positive solution (x_k, y_k) is given by

$$x_k + y_k\sqrt{d} = (x + y\sqrt{d})^k \tag{1.8.11}$$

The following tables contain the least positive solutions to Pell's equation with d square-free and $d < 100$.

d	x	y
2	3	2
3	2	1
5	9	4
6	5	2
7	8	3
10	19	6
11	10	3
13	649	180
14	15	4
15	4	1
17	33	8
19	170	39
21	55	12
22	197	42
23	24	5
26	51	10
29	9,801	1,820
30	11	2
31	1,520	273
33	23	4

d	x	y
35	6	1
37	73	12
38	37	6
39	25	4
41	2,049	320
42	13	2
43	3,482	531
46	24,335	3,588
47	48	7
51	50	7
53	66,249	9,100
55	89	12
57	151	20
58	19,603	2,574
59	530	69
61	1,766,319,049	226,153,980
62	63	8
65	129	16
66	65	8
67	48,842	5,967

d	x	y
69	7,775	936
70	251	30
71	3,480	413
73	2,281,249	267,000
74	3,699	430
77	351	40
78	53	6
79	80	9
82	163	18
83	82	9
85	285,769	30,996
86	10,405	1,122
87	28	3
89	500,001	53,000
91	1,574	165
93	12,151	1,260
94	2,143,295	221,064
95	39	4
97	62,809,633	6,377,352

EXAMPLES

1. The number $\sqrt{2}$ has the continued fraction expansion $[1, 2, 2, 2, 2, \ldots]$, with convergents $\frac{3}{2}, \frac{7}{5}, \frac{17}{12}, \frac{41}{29}, \frac{99}{70}, \ldots$. In this case, every second convergent represents a solution:
$$3^2 - 2 \cdot 2^2 = 1, \qquad 17^2 - 2 \cdot 12^2 = 1, \qquad 99^2 - 2 \cdot 70^2 = 1$$
2. The least positive solution for $d = 11$ is $(x, y) = (10, 3)$. Since $(10 + 3\sqrt{11})^2 = 199 + 60\sqrt{11}$, another solution is given by $(x_2, y_2) = (199, 60)$.

1.8.4.3 Lagrange's theorem

Lagrange's theorem states: "Every positive integer is the sum of four squares." After showing every prime can be written as the sum of four squares, the following identity can be used to show how products can be written as the sum of four squares:

$$\begin{aligned}(x_1^2 + x_2^2 + x_3^2 + x_4^2)&(y_1^2 + y_2^2 + y_3^2 + y_4^2) = \\ &(x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4)^2 + (x_1y_2 - x_2y_1 + x_3y_4 - x_4y_3)^2 \\ &+ (x_1y_3 - x_3y_1 + x_4y_2 - x_2y_4)^2 + (x_1y_4 - x_4y_1 + x_2y_3 - x_3y_2)^2 \end{aligned} \tag{1.8.12}$$

Note also the identity for sums of two squares:

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) = (x_1y_2 + x_2y_1)^2 + (x_2y_2 - x_1y_1)^2 \tag{1.8.13}$$

1.8.5 GREATEST COMMON DIVISOR

The *greatest common divisor* of the integers n and m is the largest integer that evenly divides both n and m; this is written as GCD(n, m) or (n, m). The Euclidean algorithm is frequently used for computing the GCD of two numbers; it utilizes the fact that $m = \left\lfloor \frac{m}{n} \right\rfloor n + p$ where $0 \leq p < n$.

Given the integers m and n, two integers a and b can always be found so that $am + bn =$ GCD(n, m).

Two numbers, m and n, are said to be *relatively prime* if they have no divisors in common; i.e., if GCD$(a, b) = 1$. The probability that two integers chosen randomly are relatively prime is $\pi/6$.

EXAMPLE Consider 78 and 21. Since $78 = 3 \cdot 21 + 15$, the largest integer that evenly divides both 78 and 21 is also the largest integer that evenly divides both 21 and 15. Iterating results in:

$$\begin{aligned} 78 &= 3 \cdot 21 + 15 \\ 21 &= 1 \cdot 15 + 6 \\ 15 &= 2 \cdot 6 + 3 \\ 6 &= 2 \cdot 3 + 0 \end{aligned}$$

Hence GCD$(78, 21) =$ GCD$(21, 15) =$ GCD$(15, 6) =$ GCD$(6, 3) = 3$. Note that $78 \cdot (-4) + 21 \cdot 15 = 3$.

1.8.6 LEAST COMMON MULTIPLE

The *least common multiple* of the integers a and b (denoted LCM(a, b)) is the least integer r that is divisible by both a and b. The simplest way to find the LCM of a and b is via the formula LCM$(a, b) = ab/$GCD(a, b). For example, LCM$(10, 4) = \frac{10 \cdot 4}{\text{GCD}(10,4)} = \frac{10 \cdot 4}{2} = 20$.

1.8.7 MÖBIUS FUNCTION

The *Möbius function* is defined by

1. $\mu(1) = 1$
2. $\mu(n) = 0$ if n has a squared factor
3. $\mu(p_1 p_2 \dots p_k) = (-1)^k$ if all the primes $\{p_1, \dots, p_k\}$ are distinct

Its properties include:

1. If GCD$(m, n) = 1$ then $\mu(mn) = \mu(m)\,\mu(n)$
2. $\displaystyle\sum_{d|n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \\ 0 & \text{if } n > 1 \end{cases}$
3. Generating function: $\displaystyle\sum_{n=0}^{\infty} \mu(n) n^{-s} = \frac{1}{\zeta(s)}$
4. The Möbius inversion formula states that, if $g(n) = \sum_{d|n} f(d)$, then

$$f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right). \tag{1.8.14}$$

For example, the Möbius inversion of $n = \sum_{d|n} \phi(d)$ is $\phi(n) = n \sum_{d|n} \frac{\mu(d)}{d}$.

The table below has the value of $\mu(10n + k)$ in row n_ and column _k. For example, $\mu(2) = -1$, $\mu(4) = 0$, and $\mu(6) = 1$.

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
0_		1	−1	−1	0	−1	1	−1	0	0
1_	1	−1	0	−1	1	1	0	−1	0	−1
2_	0	1	1	−1	0	0	1	0	0	−1
3_	−1	−1	0	1	1	1	0	−1	1	1
4_	0	−1	−1	−1	0	0	1	−1	0	0
5_	0	1	0	−1	0	1	0	1	1	−1
6_	0	−1	1	0	0	1	−1	−1	0	1
7_	−1	−1	0	−1	1	0	0	1	−1	−1
8_	0	0	1	−1	0	1	1	1	0	−1
9_	0	1	0	1	1	1	0	−1	0	0
10_	0	−1	−1	−1	0	−1	1	−1	0	−1
11_	−1	1	0	−1	−1	1	0	0	1	1
12_	0	0	1	1	0	0	0	−1	0	1
13_	−1	−1	0	1	1	0	0	−1	−1	−1
14_	0	1	1	1	0	1	1	0	0	−1
15_	0	−1	0	0	−1	1	0	−1	1	1

1.8.8 PRIME NUMBERS

1. A *prime number* is a positive integer greater than 1 with no positive, integral divisors other than 1 and itself. There are infinitely many prime numbers, $2, 3, 5, 7, \ldots$. The sum of the reciprocals of the prime numbers diverges: $\sum_n \frac{1}{p_n} = \frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \ldots = \infty$.
2. *Twin primes* are prime numbers that differ by two: $(3, 5)$, $(5, 7)$, $(11, 13)$, $(17, 19), \ldots$. It is not known whether there are infinitely many twin primes. The sum of the reciprocals of the twin primes converges; the value
$$B = \left(\frac{1}{3} + \frac{1}{5}\right) + \left(\frac{1}{5} + \frac{1}{7}\right) + \left(\frac{1}{11} + \frac{1}{13}\right) + \ldots + \left(\frac{1}{p} + \frac{1}{p+2}\right) + \ldots$$
known as Brun's constant is approximately $B \approx 1.90216054$.
3. For every integer $n \geq 2$, the numbers $\{n! + 2, n! + 3, \ldots, n! + n\}$ are a sequence of $n - 1$ consecutive composite (i.e., not prime) numbers.
4. *Dirichlet's theorem on primes in arithmetic progressions*: Let a and b be relatively prime positive integers. Then the arithmetic progression $an + b$ (for $n = 1, 2, \ldots$) contains infinitely many primes.
5. *Goldbach conjecture*: every even number is the sum of two prime numbers.
6. The function $\pi(x)$ represents the number of primes less than x. The prime number theorem states that $\pi(x) \sim x/\log x$ as $x \to \infty$. For $x \geq 17$, we have $\frac{x}{\log x} \leq \pi(x) \leq 1.26\frac{x}{\log x}$. The number of primes less than a given number is:

x	100	1000	10,000	10^5	10^6	10^7	10^8
$\pi(x)$	25	168	1,229	9,592	78,498	664,579	5,761,455

x	10^{10}	10^{15}	10^{21}
$\pi(x)$	455,052,511	29,844,570,422,669	21,127,269,486,018,731,928

1.8.8.1 Prime formulas

The polynomial $x^2 - x + 41$ yields prime numbers for $x = 0, 1, 2, \ldots, 39$.

The set of prime numbers is identical with the set of positive values taken on by the polynomial of degree 25 in the 26 variables $\{a, b, \ldots, z\}$:

$$\begin{aligned}(k+2)\{1-[wz+h+j-q]^2-[(gk+2g+k+1)(h+j)+h-z]^2-[2n+p+q+z-e]^2\\ -[16(k+1)^3(k+2)(n+1)^2+1-f^2]^2-[e^3(e+2)(a+1)^2+1-o^2]^2-[(a^2-1)y^2+1-x^2]^2\\ -[16r^2y^4(a^2-1)+1-u^2]^2-[((a+u^2(u^2-a))^2-1)(n+4dy)^2+1-(x+cu)^2]^2\\ -[n+l+v-y]^2-[(a^2-1)l^2+1-m^2]^2-[ai+k+1-l-i]^2-[p+l(a-n-1)\\ +b(2an+2a-n^2-2n-2)-m]^2-[q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2\\ -[z+pl(a-p)+t(2ap-p^2-1)-pm]^2\}\end{aligned}$$

Although this polynomial appears to factor, the factors are improper, $P = P \cdot 1$. Note that this formula will also take on negative values, such as -76. There also exists a prime representing polynomial with 12 variables of degree 13697, and one of 10 variables and degree about 10^{45}.

1.8.8.2 Lucas–Lehmer primality test

Define the sequence $r_{m+1} = r_m^2 - 2$ with $r_1 = 3$. If p is a prime of the form $4n+3$ and $M_p = 2^p - 1$, then M_p will be prime (called a *Mersenne prime*) if and only if M_p divides r_{p-1}.

This simple test is the reason that the largest known prime numbers are Mersenne primes. For example, consider $p = 7$ and $M_7 = 127$. The $\{r_n\}$ sequence is $\{3, 7, 47, 2207 \equiv 48, 2302 \equiv 16, 254 \equiv 0\}$; hence M_7 is prime.

1.8.8.3 Primality test certificates

A *primality certificate* is an easily verifiable statement (easier than it was to determine that a number is prime) that proves that a specific number is prime. There are several types of certificates. The Atkin–Morain certificate uses elliptic curves.

To show that the number p is prime, Pratt's certificate consists of a number a and the factorization of the number $p-1$. The number p will be prime if there exists a primitive root a in the field GF$[p]$ that satisfies the conditions $a^{p-1} = 1 \pmod{p}$ and $a^{(p-1)/q} \neq 1 \pmod{p}$ for any prime q that divides $p-1$.

EXAMPLE The number $p = 31$ has $p - 1 = 30 = 2 \cdot 3 \cdot 5$, and a primitive root is given by $a = 3$. Hence, to verify that $p = 31$ is prime, we compute

$$\begin{aligned}
3^{(31-1)/2} &= 3^{15} \equiv 14348907 \equiv -1 \neq 1 &\pmod{31},\\
3^{(31-1)/3} &= 3^{10} \equiv 59049 \equiv 25 \neq 1 &\pmod{31},\\
3^{(31-1)/5} &= 3^{6} \equiv 729 \equiv 16 \neq 1 &\pmod{31},\\
3^{(31-1)} &= \left(3^{(31-1)/2}\right)^2 \equiv (-1)^2 = 1 &\pmod{31}.
\end{aligned}$$

1.8.8.4 Probabilistic primality test

Let n be a number whose primality is to be determined. Probabilistic primality tests can return one of two results: either a proof that the number n is composite or a statement of the form, "The probability that the number n is not prime is less than ϵ," where ϵ can be specified by the user. Typically, we take $\epsilon = 2^{-200} < 10^{-60}$.

From Fermat's theorem, if $b \neq 0$, then $b^{n-1} = 1 \pmod{n}$ whenever n is prime. If this holds, then *n is a probable prime to the base b*. Given a value of n, if a value of b can be found such that this does not hold, then n cannot be prime. It can happen, however, that a probable prime is not prime.

Let $P(x)$ be the probability that n is composite under the hypotheses:

1. n is an odd integer chosen randomly from the range $[2, x]$;
2. b is an integer chosen randomly from the range $[2, n-2]$;
3. n is a probable prime to the base b.

Then $P(x) \leq (\log x)^{-197}$ for $x \geq 10^{10000}$.

A different test can be obtained from the following theorem. Given the number n, find s and t with $n - 1 = 2^s t$, with t odd. Then choose a random integer b from the range $[2, n-2]$. If either

$$b^t = 1 \pmod{n} \qquad \text{or} \qquad b^{2^i t} = -1 \pmod{n}, \quad \text{for some } i < s,$$

then n *is a strong probable prime to the base* b. Every odd prime must pass this test. If $n > 1$ is an odd composite, then the probability that it is a strong probable prime to the base b, when b is chosen randomly, is less than $1/4$.

A stronger test can be obtained by choosing k independent values for b in the range $[2, n-2]$ and checking the above relation for each value of b. Let $P_k(x)$ be the probability that n is found to be a strong probable prime to each base b. Then $P_k(x) \leq 4^{-(k-1)} P(x)/(1 - P(x))$.

1.8.9 PRIME NUMBERS OF SPECIAL FORMS

1. The largest known prime numbers, in descending order, are

	Prime number	Number of digits
(1)	$2^{74207281} - 1$	22,338,618
(2)	$2^{57885161} - 1$	17,425,170
(3)	$2^{43112609} - 1$	12,978,189
(4)	$2^{42643801} - 1$	12,837,064
(5)	$2^{37156667} - 1$	11,185,272
(6)	$2^{32582657} - 1$	9,808,358
(7)	$10223 \cdot 2^{31172165} + 1$	9,383,761
(8)	$2^{30402457} - 1$	9,152,052
(9)	$2^{25964951} - 1$	7,816,230
(10)	$2^{24036583} - 1$	7,235,733

2. The largest known twin primes are $2996863034895 \cdot 2^{1290000} \pm 1$ (with 388,342 digits).
3. There exist constants $\theta \approx 1.30637788$ and $\omega \approx 1.9287800$ such that $\left\lfloor \theta^{3^n} \right\rfloor$ and $\left\lfloor \underbrace{2^{2^{\cdot^{\cdot^{\cdot^{2^\omega}}}}}}_{n} \right\rfloor$ are prime for every $n \geq 1$.
4. Primes with special properties
   (a) A *Sophie Germain prime* p has the property that $2p + 1$ is also prime. Sophie Germain primes include: 2, 3, 5, 11, 23, 29, 41, 53, 83, 89, 113, 131, $\ldots$, $3714089895285 \cdot 2^{60000} - 1, \ldots$.
   (b) An odd prime p is called a *Wieferich prime* if $2^{p-1} \equiv 1 \pmod{p^2}$. Wieferich primes include 1093 and 3511.
   (c) A *Wilson prime* satisfies $(p-1)! \equiv -1 \pmod{p^2}$. Wilson primes include 5, 13, and 563.
5. For each n below $\{a + md \mid m = 0, 1, \ldots, n-1\}$ is an arithmetic sequence of n prime numbers. (Note that $23\# = 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23$.)

n	a	d
3	3	2
4	61	6
5	11	30
10	199	210
26	43142746595714191	23681770·23#

6. Define $p\#$ to be the product of the prime numbers less than or equal to p.

Form	Values of n or p for which the form is prime
$2^{2^n}+1$	0, 1, 2, 3, 4 … (Fermat primes)
2^n-1	2, 3, 5, 7, 13, 17, 19, 31, 61, 89, 107, 127, 521, 607, 1279, 2203, 2281, 3217, 4253, 4423, 9689, 9941, 11213, 19937, 21701, 23209, 44497, 86243, 110503, 132049, 216091, 756839, 859433, …, 74207281, … (Mersenne primes)
$n!-1$	3, 4, 6, 7, 12, 14, 30, 32, 33, 38, 94, 166, 324, 379, 469, 546, 974, 1963, 3507, 3610, 6917, … (factorial primes)
$n!+1$	1, 2, 3, 11, 27, 37, 41, 73, 77, 116, 154, 320, 340, 399, 427, 872, 1477, 6380, … (factorial primes)
$p\#-1$	3, 5, 11, 13, 41, 89, 317, 337, 991, 1873, 2053, 2377, 4093, 4297, 4583, 6569, 13033, 15877, … (primorial or Euclid primes)
$p\#+1$	2, 3, 5, 7, 11, 31, 379, 1019, 1021, 2657, 3229, 4547, 4787, 11549, 13649, 18523, 23801, 24029, 42209, …, 145823, 366439, 392113, … (primorial or Euclid primes)
$n2^n+1$	1, 141, 4713, 5795, 6611, 18496, 32292, 32469, 59656, 90825, 262419, 361275, …, 481899, … (Cullen primes)
$n2^n-1$	2, 3, 6, 30, 75, 81, 115, 123, 249, 362, 384, 462, 512, 751, 822, 5312, 7755, 9531, 12379, 15822, 18885, … 143018, 151023, 667071, … (Woodall primes)

7. Prime numbers of the form $\dfrac{a^n-1}{a-1}$ (called repunits).

Form	Values of n for which the form is prime
$\dfrac{2^n-1}{1}$	These are Mersenne primes; see the previous table.
$\dfrac{3^n-1}{2}$	3, 7, 13, 71, 103, 541, 1091, 1367, 1627, 4177, 9011, 9551, …
$\dfrac{5^n-1}{4}$	3, 7, 11, 13, 47, 127, 149, 181, 619, 929, 3407, 10949, 13241, …
$\dfrac{6^n-1}{5}$	2, 3, 7, 29, 71, 127, 271, 509, 1049, 6389, 6883, 10613, 19889, …
$\dfrac{7^n-1}{6}$	5, 13, 131, 149, 1699, 14221, 35201, 126037, 371669, …
$\dfrac{10^n-1}{9}$	2, 19, 23, 317, 1031, 49081, 86453, 109297, 270343, …

8. Prime numbers of the forms $2^n \pm a$, $10^n \pm b$ and $16^n \pm c$

In the following table, for a given value of n, the quantities $a_\pm$, $b_\pm$, and $c_\pm$ are the least values such that $2^n + a_\pm$, $10^n + b_\pm$, and $16^n + c_\pm$ are probably primes. (A probabilistic primality test was used.)

For example, for $n = 3$, the numbers $2^3 - 1 = 7$, $2^3 + 3 = 11$, $10^3 - 3 = 997$, $10^3 + 9 = 1009$, $16^3 - 3 = 4093$, and $16^3 + 3 = 4099$ are all prime.

	$2^n + a$		$10^n + b$		$16^n + c$	
n	a_-	a_+	b_-	b_+	c_-	c_+
2	−1	1	−3	1	−5	1
3	−1	3	−3	9	−3	3
4	−3	1	−27	7	−15	1
5	−1	5	−9	3	−3	7
6	−3	3	−17	3	−3	43
7	−1	3	−9	19	−57	3
8	−5	1	−11	7	−5	15
9	−3	9	−63	7	−5	31
10	−3	7	−33	19	−87	15
11	−9	5	−23	3	−17	7
12	−3	3	−11	39	−59	21
13	−1	17	−29	37	−47	21
14	−3	27	−27	31	−5	81
15	−19	3	−11	37	−93	33
16	−15	1	−63	61	−59	13
17	−1	29	−3	3	−23	33
18	−5	3	−11	3	−93	15
19	−1	21	−39	51	−15	15
20	−3	7	−11	39	−65	13
50	−27	55	−57	151	−75	235
100	−15	277	−797	267	−593	181
150	−3	147	−273	67	−95	187
200	−75	235	−189	357	−105	25
300	−153	157	−69	331	−305	1515
400	−593	181	−513	69	−2273	895
500	−863	55	−1037	961	−2217	841
600	−95	187	−1791	543	−5	255
700	−1113	535	−2313	7	−909	2823
800	−105	25	−1007	1537	−1683	751
900	−207	693	−773	1873	−1193	8767
1000	−1245	297	−1769	453	−2303	63

1.8.10 PRIME NUMBERS LESS THAN 7,000

The prime number p_{10n+k} is found by looking at row n_ and column _k. For example, the 9th prime number is 23 and the 18th prime number is 61.

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
0_		2	3	5	7	11	13	17	19	23
1_	29	31	37	41	43	47	53	59	61	67
2_	71	73	79	83	89	97	101	103	107	109
3_	113	127	131	137	139	149	151	157	163	167
4_	173	179	181	191	193	197	199	211	223	227
5_	229	233	239	241	251	257	263	269	271	277
6_	281	283	293	307	311	313	317	331	337	347
7_	349	353	359	367	373	379	383	389	397	401
8_	409	419	421	431	433	439	443	449	457	461
9_	463	467	479	487	491	499	503	509	521	523
10_	541	547	557	563	569	571	577	587	593	599
11_	601	607	613	617	619	631	641	643	647	653
12_	659	661	673	677	683	691	701	709	719	727
13_	733	739	743	751	757	761	769	773	787	797
14_	809	811	821	823	827	829	839	853	857	859
15_	863	877	881	883	887	907	911	919	929	937
16_	941	947	953	967	971	977	983	991	997	1009
17_	1013	1019	1021	1031	1033	1039	1049	1051	1061	1063
18_	1069	1087	1091	1093	1097	1103	1109	1117	1123	1129
19_	1151	1153	1163	1171	1181	1187	1193	1201	1213	1217
20_	1223	1229	1231	1237	1249	1259	1277	1279	1283	1289
21_	1291	1297	1301	1303	1307	1319	1321	1327	1361	1367
22_	1373	1381	1399	1409	1423	1427	1429	1433	1439	1447
23_	1451	1453	1459	1471	1481	1483	1487	1489	1493	1499
24_	1511	1523	1531	1543	1549	1553	1559	1567	1571	1579
25_	1583	1597	1601	1607	1609	1613	1619	1621	1627	1637
26_	1657	1663	1667	1669	1693	1697	1699	1709	1721	1723
27_	1733	1741	1747	1753	1759	1777	1783	1787	1789	1801
28_	1811	1823	1831	1847	1861	1867	1871	1873	1877	1879
29_	1889	1901	1907	1913	1931	1933	1949	1951	1973	1979
30_	1987	1993	1997	1999	2003	2011	2017	2027	2029	2039
31_	2053	2063	2069	2081	2083	2087	2089	2099	2111	2113
32_	2129	2131	2137	2141	2143	2153	2161	2179	2203	2207
33_	2213	2221	2237	2239	2243	2251	2267	2269	2273	2281
34_	2287	2293	2297	2309	2311	2333	2339	2341	2347	2351
35_	2357	2371	2377	2381	2383	2389	2393	2399	2411	2417
36_	2423	2437	2441	2447	2459	2467	2473	2477	2503	2521
37_	2531	2539	2543	2549	2551	2557	2579	2591	2593	2609
38_	2617	2621	2633	2647	2657	2659	2663	2671	2677	2683
39_	2687	2689	2693	2699	2707	2711	2713	2719	2729	2731
40_	2741	2749	2753	2767	2777	2789	2791	2797	2801	2803
41_	2819	2833	2837	2843	2851	2857	2861	2879	2887	2897
42_	2903	2909	2917	2927	2939	2953	2957	2963	2969	2971
43_	2999	3001	3011	3019	3023	3037	3041	3049	3061	3067
44_	3079	3083	3089	3109	3119	3121	3137	3163	3167	3169
45_	3181	3187	3191	3203	3209	3217	3221	3229	3251	3253

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
46_	3257	3259	3271	3299	3301	3307	3313	3319	3323	3329
47_	3331	3343	3347	3359	3361	3371	3373	3389	3391	3407
48_	3413	3433	3449	3457	3461	3463	3467	3469	3491	3499
49_	3511	3517	3527	3529	3533	3539	3541	3547	3557	3559
50_	3571	3581	3583	3593	3607	3613	3617	3623	3631	3637
51_	3643	3659	3671	3673	3677	3691	3697	3701	3709	3719
52_	3727	3733	3739	3761	3767	3769	3779	3793	3797	3803
53_	3821	3823	3833	3847	3851	3853	3863	3877	3881	3889
54_	3907	3911	3917	3919	3923	3929	3931	3943	3947	3967
55_	3989	4001	4003	4007	4013	4019	4021	4027	4049	4051
56_	4057	4073	4079	4091	4093	4099	4111	4127	4129	4133
57_	4139	4153	4157	4159	4177	4201	4211	4217	4219	4229
58_	4231	4241	4243	4253	4259	4261	4271	4273	4283	4289
59_	4297	4327	4337	4339	4349	4357	4363	4373	4391	4397
60_	4409	4421	4423	4441	4447	4451	4457	4463	4481	4483
61_	4493	4507	4513	4517	4519	4523	4547	4549	4561	4567
62_	4583	4591	4597	4603	4621	4637	4639	4643	4649	4651
63_	4657	4663	4673	4679	4691	4703	4721	4723	4729	4733
64_	4751	4759	4783	4787	4789	4793	4799	4801	4813	4817
65_	4831	4861	4871	4877	4889	4903	4909	4919	4931	4933
66_	4937	4943	4951	4957	4967	4969	4973	4987	4993	4999
67_	5003	5009	5011	5021	5023	5039	5051	5059	5077	5081
68_	5087	5099	5101	5107	5113	5119	5147	5153	5167	5171
69_	5179	5189	5197	5209	5227	5231	5233	5237	5261	5273
70_	5279	5281	5297	5303	5309	5323	5333	5347	5351	5381
71_	5387	5393	5399	5407	5413	5417	5419	5431	5437	5441
72_	5443	5449	5471	5477	5479	5483	5501	5503	5507	5519
73_	5521	5527	5531	5557	5563	5569	5573	5581	5591	5623
74_	5639	5641	5647	5651	5653	5657	5659	5669	5683	5689
75_	5693	5701	5711	5717	5737	5741	5743	5749	5779	5783
76_	5791	5801	5807	5813	5821	5827	5839	5843	5849	5851
77_	5857	5861	5867	5869	5879	5881	5897	5903	5923	5927
78_	5939	5953	5981	5987	6007	6011	6029	6037	6043	6047
79_	6053	6067	6073	6079	6089	6091	6101	6113	6121	6131
80_	6133	6143	6151	6163	6173	6197	6199	6203	6211	6217
81_	6221	6229	6247	6257	6263	6269	6271	6277	6287	6299
82_	6301	6311	6317	6323	6329	6337	6343	6353	6359	6361
83_	6367	6373	6379	6389	6397	6421	6427	6449	6451	6469
84_	6473	6481	6491	6521	6529	6547	6551	6553	6563	6569
85_	6571	6577	6581	6599	6607	6619	6637	6653	6659	6661
86_	6673	6679	6689	6691	6701	6703	6709	6719	6733	6737
87_	6761	6763	6779	6781	6791	6793	6803	6823	6827	6829
88_	6833	6841	6857	6863	6869	6871	6883	6899	6907	6911
89_	6917	6947	6949	6959	6961	6967	6971	6977	6983	6991
90_	6997	7001	7013	7019	7027	7039	7043	7057	7069	7079
91_	7103	7109	7121	7127	7129	7151	7159	7177	7187	7193
92_	7207	7211	7213	7219	7229	7237	7243	7247	7253	7283
93_	7297	7307	7309	7321	7331	7333	7349	7351	7369	7393
94_	7411	7417	7433	7451	7457	7459	7477	7481	7487	7489
95_	7499	7507	7517	7523	7529	7537	7541	7547	7549	7559

1.8.11 FACTORIZATION TABLE

The following table has the factors of all numbers up to 1029.
When a number is prime, it is shown in a boldface font.

0_			**2**	**3**	2^2	**5**	$2\cdot 3$	**7**	2^3	3^2
1_	$2\cdot 5$	**11**	$2^2\cdot 3$	**13**	$2\cdot 7$	$3\cdot 5$	2^4	**17**	$2\cdot 3^2$	**19**
2_	$2^2\cdot 5$	$3\cdot 7$	$2\cdot 11$	**23**	$2^3\cdot 3$	5^2	$2\cdot 13$	3^3	$2^2\cdot 7$	**29**
3_	$2\cdot 3\cdot 5$	**31**	2^5	$3\cdot 11$	$2\cdot 17$	$5\cdot 7$	$2^2\cdot 3^2$	**37**	$2\cdot 19$	$3\cdot 13$
4_	$2^3\cdot 5$	**41**	$2\cdot 3\cdot 7$	**43**	$2^2\cdot 11$	$3^2\cdot 5$	$2\cdot 23$	**47**	$2^4\cdot 3$	7^2
5_	$2\cdot 5^2$	$3\cdot 17$	$2^2\cdot 13$	**53**	$2\cdot 3^3$	$5\cdot 11$	$2^3\cdot 7$	$3\cdot 19$	$2\cdot 29$	**59**
6_	$2^2\cdot 3\cdot 5$	**61**	$2\cdot 31$	$3^2\cdot 7$	2^6	$5\cdot 13$	$2\cdot 3\cdot 11$	**67**	$2^2\cdot 17$	$3\cdot 23$
7_	$2\cdot 5\cdot 7$	**71**	$2^3\cdot 3^2$	**73**	$2\cdot 37$	$3\cdot 5^2$	$2^2\cdot 19$	$7\cdot 11$	$2\cdot 3\cdot 13$	**79**
8_	$2^4\cdot 5$	3^4	$2\cdot 41$	**83**	$2^2\cdot 3\cdot 7$	$5\cdot 17$	$2\cdot 43$	$3\cdot 29$	$2^3\cdot 11$	**89**
9_	$2\cdot 3^2\cdot 5$	$7\cdot 13$	$2^2\cdot 23$	$3\cdot 31$	$2\cdot 47$	$5\cdot 19$	$2^5\cdot 3$	**97**	$2\cdot 7^2$	$3^2\cdot 11$
10_	$2^2\cdot 5^2$	**101**	$2\cdot 3\cdot 17$	**103**	$2^3\cdot 13$	$3\cdot 5\cdot 7$	$2\cdot 53$	**107**	$2^2\cdot 3^3$	**109**
11_	$2\cdot 5\cdot 11$	$3\cdot 37$	$2^4\cdot 7$	**113**	$2\cdot 3\cdot 19$	$5\cdot 23$	$2^2\cdot 29$	$3^2\cdot 13$	$2\cdot 59$	$7\cdot 17$
12_	$2^3\cdot 3\cdot 5$	11^2	$2\cdot 61$	$3\cdot 41$	$2^2\cdot 31$	5^3	$2\cdot 3^2\cdot 7$	**127**	2^7	$3\cdot 43$
13_	$2\cdot 5\cdot 13$	**131**	$2^2\cdot 3\cdot 11$	$7\cdot 19$	$2\cdot 67$	$3^3\cdot 5$	$2^3\cdot 17$	**137**	$2\cdot 3\cdot 23$	**139**
14_	$2^2\cdot 5\cdot 7$	$3\cdot 47$	$2\cdot 71$	$11\cdot 13$	$2^4\cdot 3^2$	$5\cdot 29$	$2\cdot 73$	$3\cdot 7^2$	$2^2\cdot 37$	**149**
15_	$2\cdot 3\cdot 5^2$	**151**	$2^3\cdot 19$	$3^2\cdot 17$	$2\cdot 7\cdot 11$	$5\cdot 31$	$2^2\cdot 3\cdot 13$	**157**	$2\cdot 79$	$3\cdot 53$
16_	$2^5\cdot 5$	$7\cdot 23$	$2\cdot 3^4$	**163**	$2^2\cdot 41$	$3\cdot 5\cdot 11$	$2\cdot 83$	**167**	$2^3\cdot 3\cdot 7$	13^2
17_	$2\cdot 5\cdot 17$	$3^2\cdot 19$	$2^2\cdot 43$	**173**	$2\cdot 3\cdot 29$	$5^2\cdot 7$	$2^4\cdot 11$	$3\cdot 59$	$2\cdot 89$	**179**
18_	$2^2\cdot 3^2\cdot 5$	**181**	$2\cdot 7\cdot 13$	$3\cdot 61$	$2^3\cdot 23$	$5\cdot 37$	$2\cdot 3\cdot 31$	$11\cdot 17$	$2^2\cdot 47$	$3^3\cdot 7$
19_	$2\cdot 5\cdot 19$	**191**	$2^6\cdot 3$	**193**	$2\cdot 97$	$3\cdot 5\cdot 13$	$2^2\cdot 7^2$	**197**	$2\cdot 3^2\cdot 11$	**199**
20_	$2^3\cdot 5^2$	$3\cdot 67$	$2\cdot 101$	$7\cdot 29$	$2^2\cdot 3\cdot 17$	$5\cdot 41$	$2\cdot 103$	$3^2\cdot 23$	$2^4\cdot 13$	$11\cdot 19$
21_	$2\cdot 3\cdot 5\cdot 7$	**211**	$2^2\cdot 53$	$3\cdot 71$	$2\cdot 107$	$5\cdot 43$	$2^3\cdot 3^3$	$7\cdot 31$	$2\cdot 109$	$3\cdot 73$
22_	$2^2\cdot 5\cdot 11$	$13\cdot 17$	$2\cdot 3\cdot 37$	**223**	$2^5\cdot 7$	$3^2\cdot 5^2$	$2\cdot 113$	**227**	$2^2\cdot 3\cdot 19$	**229**
23_	$2\cdot 5\cdot 23$	$3\cdot 7\cdot 11$	$2^3\cdot 29$	**233**	$2\cdot 3^2\cdot 13$	$5\cdot 47$	$2^2\cdot 59$	$3\cdot 79$	$2\cdot 7\cdot 17$	**239**
24_	$2^4\cdot 3\cdot 5$	**241**	$2\cdot 11^2$	3^5	$2^2\cdot 61$	$5\cdot 7^2$	$2\cdot 3\cdot 41$	$13\cdot 19$	$2^3\cdot 31$	$3\cdot 83$
25_	$2\cdot 5^3$	**251**	$2^2\cdot 3^2\cdot 7$	$11\cdot 23$	$2\cdot 127$	$3\cdot 5\cdot 17$	2^8	**257**	$2\cdot 3\cdot 43$	$7\cdot 37$
26_	$2^2\cdot 5\cdot 13$	$3^2\cdot 29$	$2\cdot 131$	**263**	$2^3\cdot 3\cdot 11$	$5\cdot 53$	$2\cdot 7\cdot 19$	$3\cdot 89$	$2^2\cdot 67$	**269**
27_	$2\cdot 3^3\cdot 5$	**271**	$2^4\cdot 17$	$3\cdot 7\cdot 13$	$2\cdot 137$	$5^2\cdot 11$	$2^2\cdot 3\cdot 23$	**277**	$2\cdot 139$	$3^2\cdot 31$
28_	$2^3\cdot 5\cdot 7$	**281**	$2\cdot 3\cdot 47$	**283**	$2^2\cdot 71$	$3\cdot 5\cdot 19$	$2\cdot 11\cdot 13$	$7\cdot 41$	$2^5\cdot 3^2$	17^2
29_	$2\cdot 5\cdot 29$	$3\cdot 97$	$2^2\cdot 73$	**293**	$2\cdot 3\cdot 7^2$	$5\cdot 59$	$2^3\cdot 37$	$3^3\cdot 11$	$2\cdot 149$	$13\cdot 23$
30_	$2^2\cdot 3\cdot 5^2$	$7\cdot 43$	$2\cdot 151$	$3\cdot 101$	$2^4\cdot 19$	$5\cdot 61$	$2\cdot 3^2\cdot 17$	**307**	$2^2\cdot 7\cdot 11$	$3\cdot 103$
31_	$2\cdot 5\cdot 31$	**311**	$2^3\cdot 3\cdot 13$	**313**	$2\cdot 157$	$3^2\cdot 5\cdot 7$	$2^2\cdot 79$	**317**	$2\cdot 3\cdot 53$	$11\cdot 29$
32_	$2^6\cdot 5$	$3\cdot 107$	$2\cdot 7\cdot 23$	$17\cdot 19$	$2^2\cdot 3^4$	$5^2\cdot 13$	$2\cdot 163$	$3\cdot 109$	$2^3\cdot 41$	$7\cdot 47$
33_	$2\cdot 3\cdot 5\cdot 11$	**331**	$2^2\cdot 83$	$3^2\cdot 37$	$2\cdot 167$	$5\cdot 67$	$2^4\cdot 3\cdot 7$	**337**	$2\cdot 13^2$	$3\cdot 113$
34_	$2^2\cdot 5\cdot 17$	$11\cdot 31$	$2\cdot 3^2\cdot 19$	7^3	$2^3\cdot 43$	$3\cdot 5\cdot 23$	$2\cdot 173$	**347**	$2^2\cdot 3\cdot 29$	**349**
35_	$2\cdot 5^2\cdot 7$	$3^3\cdot 13$	$2^5\cdot 11$	**353**	$2\cdot 3\cdot 59$	$5\cdot 71$	$2^2\cdot 89$	$3\cdot 7\cdot 17$	$2\cdot 179$	**359**
36_	$2^3\cdot 3^2\cdot 5$	19^2	$2\cdot 181$	$3\cdot 11^2$	$2^2\cdot 7\cdot 13$	$5\cdot 73$	$2\cdot 3\cdot 61$	**367**	$2^4\cdot 23$	$3^2\cdot 41$
37_	$2\cdot 5\cdot 37$	$7\cdot 53$	$2^2\cdot 3\cdot 31$	**373**	$2\cdot 11\cdot 17$	$3\cdot 5^3$	$2^3\cdot 47$	$13\cdot 29$	$2\cdot 3^3\cdot 7$	**379**
38_	$2^2\cdot 5\cdot 19$	$3\cdot 127$	$2\cdot 191$	**383**	$2^7\cdot 3$	$5\cdot 7\cdot 11$	$2\cdot 193$	$3^2\cdot 43$	$2^2\cdot 97$	**389**
39_	$2\cdot 3\cdot 5\cdot 13$	$17\cdot 23$	$2^3\cdot 7^2$	$3\cdot 131$	$2\cdot 197$	$5\cdot 79$	$2^2\cdot 3^2\cdot 11$	**397**	$2\cdot 199$	$3\cdot 7\cdot 19$
40_	$2^4\cdot 5^2$	**401**	$2\cdot 3\cdot 67$	$13\cdot 31$	$2^2\cdot 101$	$3^4\cdot 5$	$2\cdot 7\cdot 29$	$11\cdot 37$	$2^3\cdot 3\cdot 17$	**409**
41_	$2\cdot 5\cdot 41$	$3\cdot 137$	$2^2\cdot 103$	$7\cdot 59$	$2\cdot 3^2\cdot 23$	$5\cdot 83$	$2^5\cdot 13$	$3\cdot 139$	$2\cdot 11\cdot 19$	**419**
42_	$2^2\cdot 3\cdot 5\cdot 7$	**421**	$2\cdot 211$	$3^2\cdot 47$	$2^3\cdot 53$	$5^2\cdot 17$	$2\cdot 3\cdot 71$	$7\cdot 61$	$2^2\cdot 107$	$3\cdot 11\cdot 13$
43_	$2\cdot 5\cdot 43$	**431**	$2^4\cdot 3^3$	**433**	$2\cdot 7\cdot 31$	$3\cdot 5\cdot 29$	$2^2\cdot 109$	$19\cdot 23$	$2\cdot 3\cdot 73$	**439**
44_	$2^3\cdot 5\cdot 11$	$3^2\cdot 7^2$	$2\cdot 13\cdot 17$	**443**	$2^2\cdot 3\cdot 37$	$5\cdot 89$	$2\cdot 223$	$3\cdot 149$	$2^6\cdot 7$	**449**
45_	$2\cdot 3^2\cdot 5^2$	$11\cdot 41$	$2^2\cdot 113$	$3\cdot 151$	$2\cdot 227$	$5\cdot 7\cdot 13$	$2^3\cdot 3\cdot 19$	**457**	$2\cdot 229$	$3^3\cdot 17$
46_	$2^2\cdot 5\cdot 23$	**461**	$2\cdot 3\cdot 7\cdot 11$	**463**	$2^4\cdot 29$	$3\cdot 5\cdot 31$	$2\cdot 233$	**467**	$2^2\cdot 3^2\cdot 13$	$7\cdot 67$
47_	$2\cdot 5\cdot 47$	$3\cdot 157$	$2^3\cdot 59$	$11\cdot 43$	$2\cdot 3\cdot 79$	$5^2\cdot 19$	$2^2\cdot 7\cdot 17$	$3^2\cdot 53$	$2\cdot 239$	**479**
48_	$2^5\cdot 3\cdot 5$	$13\cdot 37$	$2\cdot 241$	$3\cdot 7\cdot 23$	$2^2\cdot 11^2$	$5\cdot 97$	$2\cdot 3^5$	**487**	$2^3\cdot 61$	$3\cdot 163$

49_	$2\cdot5\cdot7^2$	**491**	$2^2\cdot3\cdot41$	$17\cdot29$	$2\cdot13\cdot19$	$3^2\cdot5\cdot11$	$2^4\cdot31$	$7\cdot71$	$2\cdot3\cdot83$	**499**
50_	$2^2\cdot5^3$	$3\cdot167$	$2\cdot251$	**503**	$2^3\cdot3^2\cdot7$	$5\cdot101$	$2\cdot11\cdot23$	$3\cdot13^2$	$2^2\cdot127$	**509**
51_	$2\cdot3\cdot5\cdot17$	$7\cdot73$	2^9	$3^3\cdot19$	$2\cdot257$	$5\cdot103$	$2^2\cdot3\cdot43$	$11\cdot47$	$2\cdot7\cdot37$	$3\cdot173$
52_	$2^3\cdot5\cdot13$	**521**	$2\cdot3^2\cdot29$	**523**	$2^2\cdot131$	$3\cdot5^2\cdot7$	$2\cdot263$	$17\cdot31$	$2^4\cdot3\cdot11$	23^2
53_	$2\cdot5\cdot53$	$3^2\cdot59$	$2^2\cdot7\cdot19$	$13\cdot41$	$2\cdot3\cdot89$	$5\cdot107$	$2^3\cdot67$	$3\cdot179$	$2\cdot269$	$7^2\cdot11$
54_	$2^2\cdot3^3\cdot5$	**541**	$2\cdot271$	$3\cdot181$	$2^5\cdot17$	$5\cdot109$	$2\cdot3\cdot7\cdot13$	**547**	$2^2\cdot137$	$3^2\cdot61$
55_	$2\cdot5^2\cdot11$	$19\cdot29$	$2^3\cdot3\cdot23$	$7\cdot79$	$2\cdot277$	$3\cdot5\cdot37$	$2^2\cdot139$	**557**	$2\cdot3^2\cdot31$	$13\cdot43$
56_	$2^4\cdot5\cdot7$	$3\cdot11\cdot17$	$2\cdot281$	**563**	$2^2\cdot3\cdot47$	$5\cdot113$	$2\cdot283$	$3^4\cdot7$	$2^3\cdot71$	**569**
57_	$2\cdot3\cdot5\cdot19$	**571**	$2^2\cdot11\cdot13$	$3\cdot191$	$2\cdot7\cdot41$	$5^2\cdot23$	$2^6\cdot3^2$	**577**	$2\cdot17^2$	$3\cdot193$
58_	$2^2\cdot5\cdot29$	$7\cdot83$	$2\cdot3\cdot97$	$11\cdot53$	$2^3\cdot73$	$3^2\cdot5\cdot13$	$2\cdot293$	**587**	$2^2\cdot3\cdot7^2$	$19\cdot31$
59_	$2\cdot5\cdot59$	$3\cdot197$	$2^4\cdot37$	**593**	$2\cdot3^3\cdot11$	$5\cdot7\cdot17$	$2^2\cdot149$	$3\cdot199$	$2\cdot13\cdot23$	**599**
60_	$2^3\cdot3\cdot5^2$	**601**	$2\cdot7\cdot43$	$3^2\cdot67$	$2^2\cdot151$	$5\cdot11^2$	$2\cdot3\cdot101$	**607**	$2^5\cdot19$	$3\cdot7\cdot29$
61_	$2\cdot5\cdot61$	$13\cdot47$	$2^2\cdot3^2\cdot17$	**613**	$2\cdot307$	$3\cdot5\cdot41$	$2^3\cdot7\cdot11$	**617**	$2\cdot3\cdot103$	**619**
62_	$2^2\cdot5\cdot31$	$3^3\cdot23$	$2\cdot311$	$7\cdot89$	$2^4\cdot3\cdot13$	5^4	$2\cdot313$	$3\cdot11\cdot19$	$2^2\cdot157$	$17\cdot37$
63_	$2\cdot3^2\cdot5\cdot7$	**631**	$2^3\cdot79$	$3\cdot211$	$2\cdot317$	$5\cdot127$	$2^2\cdot3\cdot53$	$7^2\cdot13$	$2\cdot11\cdot29$	$3^2\cdot71$
64_	$2^7\cdot5$	**641**	$2\cdot3\cdot107$	**643**	$2^2\cdot7\cdot23$	$3\cdot5\cdot43$	$2\cdot17\cdot19$	**647**	$2^3\cdot3^4$	$11\cdot59$
65_	$2\cdot5^2\cdot13$	$3\cdot7\cdot31$	$2^2\cdot163$	**653**	$2\cdot3\cdot109$	$5\cdot131$	$2^4\cdot41$	$3^2\cdot73$	$2\cdot7\cdot47$	**659**
66_	$2^2\cdot3\cdot5\cdot11$	**661**	$2\cdot331$	$3\cdot13\cdot17$	$2^3\cdot83$	$5\cdot7\cdot19$	$2\cdot3^2\cdot37$	$23\cdot29$	$2^2\cdot167$	$3\cdot223$
67_	$2\cdot5\cdot67$	$11\cdot61$	$2^5\cdot3\cdot7$	**673**	$2\cdot337$	$3^3\cdot5^2$	$2^2\cdot13^2$	**677**	$2\cdot3\cdot113$	$7\cdot97$
68_	$2^3\cdot5\cdot17$	$3\cdot227$	$2\cdot11\cdot31$	**683**	$2^2\cdot3^2\cdot19$	$5\cdot137$	$2\cdot7^3$	$3\cdot229$	$2^4\cdot43$	$13\cdot53$
69_	$2\cdot3\cdot5\cdot23$	**691**	$2^2\cdot173$	$3^2\cdot7\cdot11$	$2\cdot347$	$5\cdot139$	$2^3\cdot3\cdot29$	$17\cdot41$	$2\cdot349$	$3\cdot233$
70_	$2^2\cdot5^2\cdot7$	**701**	$2\cdot3^3\cdot13$	$19\cdot37$	$2^6\cdot11$	$3\cdot5\cdot47$	$2\cdot353$	$7\cdot101$	$2^2\cdot3\cdot59$	**709**
71_	$2\cdot5\cdot71$	$3^2\cdot79$	$2^3\cdot89$	$23\cdot31$	$2\cdot3\cdot7\cdot17$	$5\cdot11\cdot13$	$2^2\cdot179$	$3\cdot239$	$2\cdot359$	**719**
72_	$2^4\cdot3^2\cdot5$	$7\cdot103$	$2\cdot19^2$	$3\cdot241$	$2^2\cdot181$	$5^2\cdot29$	$2\cdot3\cdot11^2$	**727**	$2^3\cdot7\cdot13$	3^6
73_	$2\cdot5\cdot73$	$17\cdot43$	$2^2\cdot3\cdot61$	**733**	$2\cdot367$	$3\cdot5\cdot7^2$	$2^5\cdot23$	$11\cdot67$	$2\cdot3^2\cdot41$	**739**
74_	$2^2\cdot5\cdot37$	$3\cdot13\cdot19$	$2\cdot7\cdot53$	**743**	$2^3\cdot3\cdot31$	$5\cdot149$	$2\cdot373$	$3^2\cdot83$	$2^2\cdot11\cdot17$	$7\cdot107$
75_	$2\cdot3\cdot5^3$	**751**	$2^4\cdot47$	$3\cdot251$	$2\cdot13\cdot29$	$5\cdot151$	$2^2\cdot3^3\cdot7$	**757**	$2\cdot379$	$3\cdot11\cdot23$
76_	$2^3\cdot5\cdot19$	**761**	$2\cdot3\cdot127$	$7\cdot109$	$2^2\cdot191$	$3^2\cdot5\cdot17$	$2\cdot383$	$13\cdot59$	$2^8\cdot3$	**769**
77_	$2\cdot5\cdot7\cdot11$	$3\cdot257$	$2^2\cdot193$	**773**	$2\cdot3^2\cdot43$	$5^2\cdot31$	$2^3\cdot97$	$3\cdot7\cdot37$	$2\cdot389$	$19\cdot41$
78_	$2^2\cdot3\cdot5\cdot13$	$11\cdot71$	$2\cdot17\cdot23$	$3^3\cdot29$	$2^4\cdot7^2$	$5\cdot157$	$2\cdot3\cdot131$	**787**	$2^2\cdot197$	$3\cdot263$
79_	$2\cdot5\cdot79$	$7\cdot113$	$2^3\cdot3^2\cdot11$	$13\cdot61$	$2\cdot397$	$3\cdot5\cdot53$	$2^2\cdot199$	**797**	$2\cdot3\cdot7\cdot19$	$17\cdot47$
80_	$2^5\cdot5^2$	$3^2\cdot89$	$2\cdot401$	$11\cdot73$	$2^2\cdot3\cdot67$	$5\cdot7\cdot23$	$2\cdot13\cdot31$	$3\cdot269$	$2^3\cdot101$	**809**
81_	$2\cdot3^4\cdot5$	**811**	$2^2\cdot7\cdot29$	$3\cdot271$	$2\cdot11\cdot37$	$5\cdot163$	$2^4\cdot3\cdot17$	$19\cdot43$	$2\cdot409$	$3^2\cdot7\cdot13$
82_	$2^2\cdot5\cdot41$	**821**	$2\cdot3\cdot137$	**823**	$2^3\cdot103$	$3\cdot5^2\cdot11$	$2\cdot7\cdot59$	**827**	$2^2\cdot3^2\cdot23$	**829**
83_	$2\cdot5\cdot83$	$3\cdot277$	$2^6\cdot13$	$7^2\cdot17$	$2\cdot3\cdot139$	$5\cdot167$	$2^2\cdot11\cdot19$	$3^3\cdot31$	$2\cdot419$	**839**
84_	$2^3\cdot3\cdot5\cdot7$	29^2	$2\cdot421$	$3\cdot281$	$2^2\cdot211$	$5\cdot13^2$	$2\cdot3^2\cdot47$	$7\cdot11^2$	$2^4\cdot53$	$3\cdot283$
85_	$2\cdot5^2\cdot17$	$23\cdot37$	$2^2\cdot3\cdot71$	**853**	$2\cdot7\cdot61$	$3^2\cdot5\cdot19$	$2^3\cdot107$	**857**	$2\cdot3\cdot11\cdot13$	**859**
86_	$2^2\cdot5\cdot43$	$3\cdot7\cdot41$	$2\cdot431$	**863**	$2^5\cdot3^3$	$5\cdot173$	$2\cdot433$	$3\cdot17^2$	$2^2\cdot7\cdot31$	$11\cdot79$
87_	$2\cdot3\cdot5\cdot29$	$13\cdot67$	$2^3\cdot109$	$3^2\cdot97$	$2\cdot19\cdot23$	$5^3\cdot7$	$2^2\cdot3\cdot73$	**877**	$2\cdot439$	$3\cdot293$
88_	$2^4\cdot5\cdot11$	**881**	$2\cdot3^2\cdot7^2$	**883**	$2^2\cdot13\cdot17$	$3\cdot5\cdot59$	$2\cdot443$	**887**	$2^3\cdot3\cdot37$	$7\cdot127$
89_	$2\cdot5\cdot89$	$3^4\cdot11$	$2^2\cdot223$	$19\cdot47$	$2\cdot3\cdot149$	$5\cdot179$	$2^7\cdot7$	$3\cdot13\cdot23$	$2\cdot449$	$29\cdot31$
90_	$2^2\cdot3^2\cdot5^2$	$17\cdot53$	$2\cdot11\cdot41$	$3\cdot7\cdot43$	$2^3\cdot113$	$5\cdot181$	$2\cdot3\cdot151$	**907**	$2^2\cdot227$	$3^2\cdot101$
91_	$2\cdot5\cdot7\cdot13$	**911**	$2^4\cdot3\cdot19$	$11\cdot83$	$2\cdot457$	$3\cdot5\cdot61$	$2^2\cdot229$	$7\cdot131$	$2\cdot3^3\cdot17$	**919**
92_	$2^3\cdot5\cdot23$	$3\cdot307$	$2\cdot461$	$13\cdot71$	$2^2\cdot3\cdot7\cdot11$	$5^2\cdot37$	$2\cdot463$	$3^2\cdot103$	$2^5\cdot29$	**929**
93_	$2\cdot3\cdot5\cdot31$	$7^2\cdot19$	$2^2\cdot233$	$3\cdot311$	$2\cdot467$	$5\cdot11\cdot17$	$2^3\cdot3^2\cdot13$	**937**	$2\cdot7\cdot67$	$3\cdot313$
94_	$2^2\cdot5\cdot47$	**941**	$2\cdot3\cdot157$	$23\cdot41$	$2^4\cdot59$	$3^3\cdot5\cdot7$	$2\cdot11\cdot43$	**947**	$2^2\cdot3\cdot79$	$13\cdot73$
95_	$2\cdot5^2\cdot19$	$3\cdot317$	$2^3\cdot7\cdot17$	**953**	$2\cdot3^2\cdot53$	$5\cdot191$	$2^2\cdot239$	$3\cdot11\cdot29$	$2\cdot479$	$7\cdot137$
96_	$2^6\cdot3\cdot5$	31^2	$2\cdot13\cdot37$	$3^2\cdot107$	$2^2\cdot241$	$5\cdot193$	$2\cdot3\cdot7\cdot23$	**967**	$2^3\cdot11^2$	$3\cdot17\cdot19$
97_	$2\cdot5\cdot97$	**971**	$2^2\cdot3^5$	$7\cdot139$	$2\cdot487$	$3\cdot5^2\cdot13$	$2^4\cdot61$	**977**	$2\cdot3\cdot163$	$11\cdot89$
98_	$2^2\cdot5\cdot7^2$	$3^2\cdot109$	$2\cdot491$	**983**	$2^3\cdot3\cdot41$	$5\cdot197$	$2\cdot17\cdot29$	$3\cdot7\cdot47$	$2^2\cdot13\cdot19$	$23\cdot43$
99_	$2\cdot3^2\cdot5\cdot11$	**991**	$2^5\cdot31$	$3\cdot331$	$2\cdot7\cdot71$	$5\cdot199$	$2^2\cdot3\cdot83$	**997**	$2\cdot499$	$3^3\cdot37$
100_	$2^3\cdot5^3$	$7\cdot11\cdot13$	$2\cdot3\cdot167$	$17\cdot59$	$2^2\cdot251$	$3\cdot5\cdot67$	$2\cdot503$	$19\cdot53$	$2^4\cdot3^2\cdot7$	**1009**
101_	$2\cdot5\cdot101$	$3\cdot337$	$2^2\cdot11\cdot23$	**1013**	$2\cdot3\cdot13^2$	$5\cdot7\cdot29$	$2^3\cdot127$	$3^2\cdot113$	$2\cdot509$	**1019**
102_	$2^2\cdot3\cdot5\cdot17$	**1021**	$2\cdot7\cdot73$	$3\cdot11\cdot31$	2^{10}	$5^2\cdot41$	$2\cdot3^3\cdot19$	$13\cdot79$	$2^2\cdot257$	$3\cdot7^3$

1.8.12 EULER TOTIENT FUNCTION

1.8.12.1 Definitions

$\phi(n)$ the *totient function*, is the number of integers not exceeding and relatively prime to n.

$\sigma(n)$ is the *sum of the divisors* of n.

$\tau(n)$ is the *number of divisors* of n. (Also called the $d(n)$ function.)

Define $\sigma_k(n)$ to be the k^{th} divisor function, the sum of the k^{th} powers of the divisors of n. Then $\tau(n) = \sigma_0(n)$ and $\sigma(n) = \sigma_1(n)$.

EXAMPLE The numbers that are less than 6 and relatively prime to 6 are $\{1, 5\}$. Hence $\phi(6) = 2$. The divisors of 6 are $\{1, 2, 3, 6\}$. There are $\tau(6) = 4$ divisors. The sum of these numbers is $\sigma(6) = 1 + 2 + 3 + 6 = 12$.

1.8.12.2 Properties of the totient function

1. ϕ is a multiplicative function: if $(n, m) = 1$, then $\phi(nm) = \phi(m)\phi(n)$.
2. If p is prime, then $\phi(p) = p - 1$. In general, $\phi(n) = n \prod_{p|n} \left(1 - \frac{1}{p}\right)$.
3. Gauss' theorem states: $n = \sum_{d|n} \phi(d)$.
4. When $n = \prod_i p_i^{\alpha_i}$, and the $\{p_i\}$ are prime

$$\sigma_k(n) = \sum_{d|n} d^k = \prod_i \frac{p_i^{k(\alpha_i+1)} - 1}{p_i^k - 1} \tag{1.8.15}$$

5. Generating functions

$$\sum_{n=1}^{\infty} \frac{\sigma_k(n)}{n^s} = \zeta(s)\zeta(s-k) \qquad \sum_{n=1}^{\infty} \frac{\phi(n)}{n^s} = \frac{\zeta(s-1)}{\zeta(s)} \tag{1.8.16}$$

6. A *perfect number* n satisfies $\sigma(n) = 2n$. The integer n is an even perfect number if and only if $n = 2^{m-1}(2^m - 1)$, where m is a positive integer such that $M_m = 2^m - 1$ is a Mersenne prime. The sequence of perfect numbers is $\{6, 28, 496, \dots\}$, corresponding to $m = 2, 3, 5, \dots$. It is not known whether there exists an odd perfect number.

1.8.12.3 Table of totient function values

n	$\phi(n)$	$\tau(n)$	$\sigma(n)$	n	$\phi(n)$	$\tau(n)$	$\sigma(n)$	n	$\phi(n)$	$\tau(n)$	$\sigma(n)$	n	$\phi(n)$	$\tau(n)$	$\sigma(n)$
1	0	1	1	2	1	2	3	3	2	2	4	4	2	3	7
5	4	2	6	6	2	4	12	7	6	2	8	8	4	4	15
9	6	3	13	10	4	4	18	11	10	2	12	12	4	6	28
13	12	2	14	14	6	4	24	15	8	4	24	16	8	5	31
17	16	2	18	18	6	6	39	19	18	2	20	20	8	6	42
21	12	4	32	22	10	4	36	23	22	2	24	24	8	8	60
25	20	3	31	26	12	4	42	27	18	4	40	28	12	6	56
29	28	2	30	30	8	8	72	31	30	2	32	32	16	6	63
33	20	4	48	34	16	4	54	35	24	4	48	36	12	9	91
37	36	2	38	38	18	4	60	39	24	4	56	40	16	8	90
41	40	2	42	42	12	8	96	43	42	2	44	44	20	6	84
45	24	6	78	46	22	4	72	47	46	2	48	48	16	10	124
49	42	3	57	50	20	6	93	51	32	4	72	52	24	6	98
53	52	2	54	54	18	8	120	55	40	4	72	56	24	8	120

1.9 SERIES AND PRODUCTS

1.9.1 DEFINITIONS

If $\{a_n\}$ is a sequence of numbers or functions, then

1. $S_N = \sum_{n=1}^{N} a_n = a_1 + a_2 + ... + a_N$ is the N^{th} *partial sum*.
2. For an infinite series: $S = \lim_{N\to\infty} S_N = \sum_{n=1}^{\infty} a_n$ (when the limit exists). Then S is called the sum of the series.
3. The series is said to *converge* if the limit exists and *diverge* if it does not.
4. If $a_n = b_n x^n$, where b_n is independent of x, then S is called a *power series*.
5. If $a_n = (-1)^n |a_n|$, then S is called an *alternating series*.
6. If $\sum |a_n|$ converges, then the series *converges absolutely*.
7. If S_N converges, but not absolutely, then it *converges conditionally*.

EXAMPLES

1. The *harmonic series* $S = 1 + \frac{1}{2} + \frac{1}{3} + \dots$ diverges. The corresponding alternating series (called the *alternating harmonic series*) $S = 1 - \frac{1}{2} + \frac{1}{3} - \dots + (-1)^{n-1}\frac{1}{n} + \dots$ converges (conditionally) to $\log 2$.
2. The harmonic numbers are $H_n = \displaystyle\sum_{k=1}^{n} \frac{1}{k}$. The first few value are $\{1, \frac{3}{2}, \frac{11}{6}, \frac{25}{12}, \dots\}$. Asymptotically, $H_n \sim \ln n + \gamma + \frac{1}{2n}$.
3. $S_N = \displaystyle\sum_{n=0}^{N} x^n = \frac{1 - x^{N+1}}{1 - x}$ if $x \neq 1$.

1.9.2 GENERAL PROPERTIES

1. Adding or removing a finite number of terms does not affect the convergence or divergence of an infinite series.
2. The terms of an absolutely convergent series may be rearranged in any manner without affecting its value.
3. A conditionally convergent series can be made to diverge or to converge to any (finite) value by suitable rearranging of its terms.
4. If the component series are convergent, then

$$\sum(\alpha a_n + \beta b_n) = \alpha \sum a_n + \beta \sum b_n.$$

5. $\displaystyle\left(\sum_{n=0}^{\infty} a_n\right)\left(\sum_{n=0}^{\infty} b_n\right) = \sum_{n=0}^{\infty} c_n$ where $c_n = a_0 b_n + a_1 b_{n-1} + \dots + a_n b_0$.

6. Summation by parts: let $\sum a_n$ and $\sum b_n$ converge. Then

$$\sum a_n b_n = \sum S_n(b_n - b_{n+1}) + \text{(boundary terms)}$$

where $S_n = \sum a_n$ is the n^{th} partial sum of the $\{a_n\}$.

7. A power series may be integrated and differentiated term-by-term within its interval of convergence.

8. *Schwartz inequality*:

$$\sum |a_n|\,|b_n| \leq \left(\sum |a_n|^2\right)^{1/2} \left(\sum |b_n|^2\right)^{1/2}$$

9. *Holder inequality*: when $1/p + 1/q = 1$ and $p, q > 1$

$$\sum |a_n b_n| \leq \left(\sum |a_n|^p\right)^{1/p} \left(\sum |b_n|^q\right)^{1/q}$$

10. *Minkowski inequality*: when $p \geq 1$

$$\left(\sum |a_n + b_n|^p\right)^{1/p} \leq \left(\sum |a_n|^p\right)^{1/p} + \left(\sum |b_n|^p\right)^{1/p}$$

11. *Arithmetic mean—geometric mean inequality*: If $a_i \geq 0$ then

$$\frac{a_1 + a_2 + \ldots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}$$

12. *Kantorovich inequality*: Suppose that $0 < x_1 < x_2 < \ldots < x_n$. If $\lambda_i \geq 0$ and $\sum_{i=1}^{n} \lambda_i = 1$ then

$$\left(\sum \lambda_i x_i\right) \left(\sum \frac{\lambda_i}{x_i}\right) \leq A^2 G^{-2}$$

where $A = \frac{1}{2}(x_1 + x_n)$ and $G = \sqrt{x_1 x_n}$.

EXAMPLES

1. Let T be the alternating harmonic series S rearranged so that each positive term is followed by the next two negative terms. By combining each positive term of T with the succeeding negative term, we find that $T_{3N} = \frac{1}{2} S_{2N}$. Hence, $T = \frac{1}{2} \log 2$.

2. The series $1 + \frac{1}{2} - \frac{1}{3} + \frac{1}{4} + \frac{1}{5} - \frac{1}{6} + \frac{1}{7} + \frac{1}{8} - \frac{1}{9} + \ldots$ diverges, whereas

$$\left(1 + \frac{1}{3} - \frac{1}{2}\right) + \left(\frac{1}{5} + \frac{1}{7} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{4n-3} + \frac{1}{4n-1} - \frac{1}{2n}\right) + \ldots$$

converges to $\log(2\sqrt{2})$.

1.9.3 CONVERGENCE TESTS

1. *Comparison test*
   If $|a_n| \leq b_n$ and $\sum b_n$ converges, then $\sum a_n$ converges.
2. *Limit test*
   If $\lim_{n\to\infty} a_n \neq 0$, or the limit does not exist, then $\sum a_n$ is divergent.
3. *Ratio test*
   Let $\rho = \lim_{n\to\infty} |\frac{a_{n+1}}{a_n}|$. If $\rho < 1$, the series converges absolutely. If $\rho > 1$, the series diverges.
4. *Cauchy root test*
   Let $\sigma = \lim_{n\to\infty} |a_n|^{1/n}$. If $\sigma < 1$, the series converges. If $\sigma > 1$, it diverges.
5. *Integral test*
   Let $|a_n| = f(n)$ with $f(x)$ being monotone decreasing, and $\lim_{x\to\infty} f(x) = 0$. Then $\int_A^\infty f(x)\,dx$ and $\sum a_n$ both converge or both diverge for any $A > 0$.
6. *Gauss test*
   If $\left|\frac{a_n}{a_{n+1}}\right| = 1 + \frac{p}{n} + \frac{A_n}{n^q}$, for sufficiently large n, where $q > 1$ and the sequence $\{A_n\}$ is bounded, then the series is absolutely convergent if and only if $p > 1$.
7. *Alternating series test*
   If $|a_n|$ tends monotonically to 0, then $\sum(-1)^n|a_n|$ converges.
8. If $a_i > 0$ and $\alpha = \overline{\lim}\frac{\log a_n}{\log n}$ and $\beta = \underline{\lim}\frac{\log a_n}{\log n}$, then we can apply the test
   - if $\alpha < -1$ then $\sum a_n$ converges
   - if $\beta > 1$ then $\sum a_n$ diverges

EXAMPLES

1. For $S = \sum_{n=1}^\infty n^c x^n$, $\rho = \lim_{n\to\infty}(1 + \frac{1}{n})^c x = x$. Hence, using the ratio test, S converges for $|x| < 1$ and any value of c.
2. For $S = \sum_{n=1}^\infty \frac{5^n}{n^{20}}$, $\sigma = \lim_{n\to\infty}(\frac{5^n}{n^{20}})^{1/n} = 5$. Therefore the series diverges.
3. For $S = \sum_{n=1}^\infty n^{-t}$, consider $f(x) = x^{-t}$. Then

   $$\int_1^\infty f(x)\,dx = \int_1^\infty \frac{dx}{x^t} = \begin{cases} \dfrac{1}{t-1} & \text{for } t > 1 \\ \text{diverges} & \text{for } t \leq 1 \end{cases}$$

   Hence, S converges for $t > 1$.
4. The sum $\sum_{n=2}^\infty \frac{1}{n(\log n)^s}$ converges for $s > 1$ by the integral test.
5. Let $a_n = \frac{(c)_n}{n!} = \frac{c(c+1)\dots(c+n-1)}{n!}$ where c is not 0 or a negative integer. Then $\left|\frac{a_n}{a_{n+1}}\right| = 1 + \frac{1-c}{n} - \frac{1}{n^2}\left(\frac{cn(1-c)}{n+c}\right)$. By Gauss' test, the series is absolutely convergent if and only if $c < 0$.
6. The series $\sum_{n=2}^\infty \frac{(-1)^n}{\sqrt{n} + (-1)^n}$ diverges while the series $\sum_{n=2}^\infty \frac{(-1)^n}{\sqrt{n} - 1}$ converges.

1.9.4 TYPES OF SERIES

1.9.4.1 Power series

1. The values of x, for which the power series $\sum_{n=0}^{\infty} a_n x^n$ converges, form an interval (*interval of convergence*) which may or may not include one or both endpoints.
2. A power series may be integrated and differentiated term-by-term within its interval of convergence.
3. Note that $[1 + \sum_{n=1}^{\infty} a_n x^n]^{-1} = 1 - \sum_{n=1}^{\infty} b_n x^n$, where $b_1 = a_1$ and $b_n = a_n + \sum_{k=1}^{n-1} b_{n-k} a_k$ for $n \geq 2$.
4. *Inversion of power series*: If $s = \sum_{n=1}^{\infty} a_n x^n$, then $x = \sum_{n=1}^{\infty} A_n s^n$, where $A_1 = 1/a_1$, $A_2 = -a_2/a_1^3$, $A_3 = (2a_2^2 - a_1 a_3)/a_1^5$, $A_4 = (5a_1 a_2 a_3 - a_1^2 a_4 - 5a_2^3)/a_1^7$, $A_5 = (6a_1^2 a_2 a_4 + 3a_1^2 a_3^2 + 14a_2^4 - a_1^3 a_5 - 21 a_1 a_2^2 a_3)/a_1^9$.

1.9.4.2 Taylor series

1. Taylor series in 1 variable:
$$f(a+x) = \sum_{n=0}^{N} \frac{f^{(n)}(a)}{n!} x^n + R_N.$$
or
$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \frac{f'''(a)}{3!}(x-a)^3 + \dots$$
or, specializing to $a = 0$, results in the MacLaurin series
$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \frac{f'''(0)}{3!}x^3 + \dots$$
2. Lagrange's form of the remainder:
$$R_N = \frac{x^{N+1}}{(N+1)!} f^{(N+1)}(a + \theta x), \qquad \text{for some } 0 < \theta < 1.$$
3. Taylor series in 2 variables:
$$f(a+x, b+y) = f(a,b) + x f_x(a,b) + y f_y(a,b) + \frac{1}{2!}\left[x^2 f_{xx}(a,b) + 2xy f_{xy}(a,b) + y^2 f_{yy}(a,b)\right] + \dots.$$
4. Taylor series for vectors:
$$f(\mathbf{a}+\mathbf{x}) = \sum_{n=0}^{N} \frac{[(\mathbf{x}\cdot\nabla)^n f](\mathbf{a})}{n!} + R_N(\mathbf{a}) = f(\mathbf{a}) + \mathbf{x}\cdot\nabla f(\mathbf{a}) + \dots.$$

EXAMPLE

- *Binomial series*: $(x+y)^{\nu} = \sum_{n=0}^{\infty} \frac{\Gamma(\nu+1)}{\Gamma(\nu-n+1)} \frac{x^n y^{\nu-n}}{n!}$.

  When ν is a positive integer, this series terminates at $n = \nu$.

1.9.4.3 Telescoping series

If $\lim_{n\to\infty} F(n) = 0$, then $\sum_{n=1}^{\infty}[F(n) - F(n+1)] = F(1)$. For example,

$$\sum_{n=1}^{\infty} \frac{1}{(n+1)(n+2)} = \sum_{n=1}^{\infty} \left[\frac{1}{n+1} - \frac{1}{n+2}\right] = \frac{1}{2}. \tag{1.9.1}$$

The Wilf–Zeilberger algorithm expresses a proposed identity in the form of a telescoping series $\sum_k [F(n+1,k) - F(n,k)] = 0$, then searches for a $G(n,k)$ that satisfies $F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k)$ and $G(n,\pm\infty) = 0$. The search assumes that $G(n,k) = R(n,k)F(n,k-1)$ where $R(n,k)$ is a rational expression in n and k. When R is found, the proposed identity is verified.

- The identity $\displaystyle\sum_{k=-\infty}^{\infty} (-1)^k \binom{n}{k}\binom{2k}{k} 4^{n-k} = \binom{2n}{n}$ has the proof $R(n,k) = (2k-1)/(2n-1)$. This is equivalent to $F(n,k) = (-1)^m \binom{n}{k}\binom{2k}{k} 4^{n-K}/\binom{2n}{n}$ and $G(n,k) = R(n,K)F(n,k-1)$.
- The identity $\displaystyle\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$ has the proof $R(n,k) = -\frac{k^2(3n+3-2k)}{2(n+1-k)^2(2n+1)}$
- The Pfaff–Saalschutz identity:

$$\sum_{k=-\infty}^{\infty} \frac{(a+k)!(b+k)!(c-a-b+n-1-k)!}{(k+1)!(n-k)!(c+k)!} = \frac{(c-a+n)!(c-b+n)!}{(n+1)!(c+n)!},$$

  has the proof $R(n,k) = -\dfrac{(b+k)(a+k)}{(c-b+n+1)(c-a+n+1)}$.

1.9.4.4 Dirichlet series

These are series of the form $\sum_{n=1}^{\infty} \frac{a_n}{n^x}$. They converge for $x > x_0$, where x_0 is the *abscissa of convergence*. Assuming the limits exist:

1. If $\sum a_n$ diverges, then $x_0 = \displaystyle\lim_{n\to\infty} \frac{\log|a_1 + \cdots + a_n|}{\log n}$.
2. If $\sum a_n$ converges, then $x_0 = \displaystyle\lim_{n\to\infty} \frac{\log|a_{n+1} + a_{n+2} + \dots|}{\log n}$.

EXAMPLES

1. *Riemann zeta function*: $\zeta(x) = \displaystyle\sum_{n=1}^{\infty} \frac{1}{n^x}$, $x_0 = 1$
2. $\displaystyle\sum_{n=1}^{\infty} \frac{\mu(n)}{n^x} = \frac{1}{\zeta(x)}$, $x_0 = 1$ ($\mu(n)$ denotes the Möbius function; see page 36)
3. $\displaystyle\sum_{n=1}^{\infty} \frac{d(n)}{n^x} = \zeta^2(x)$, $x_0 = 1$ ($d(n)$ is the number of divisors of n; see page 46)

1.9.4.5 Hypergeometric series

The hypergeometric function is

$$ {}_pF_q\left(\begin{matrix} a_1 & a_2 & \dots & a_p \\ b_1 & b_2 & \dots & b_q \end{matrix}\middle| x\right) = \sum_{n=0}^{\infty} \frac{(a_1)_n(a_2)_n\dots(a_p)_n}{(b_1)_n(b_2)_n\dots(b_q)_n}\frac{x^n}{n!} \tag{1.9.2} $$

where $(a)_n = \Gamma(a+n)/\Gamma(a)$ is the shifted factorial. Any infinite series $\sum A_n$ with A_{n+1}/A_n a rational function of n is a hypergeometric series. These include series of products and quotients of binomial coefficients.

EXAMPLES

- ${}_2F_1\left(\begin{matrix} a, & b \\ c & \end{matrix}\middle| 1\right) = \dfrac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}$ (Gauss)
- ${}_3F_2\left(\begin{matrix} -n, & a, & b \\ c, & 1+a+b-c-n & \end{matrix}\middle| 1\right) = \dfrac{(c-a)_n(c-b)_n}{(c)_n(c-a-b)_n}$ (Saalschutz)

1.9.4.6 Other types of series

1. *Arithmetic series*: $\displaystyle\sum_{n=1}^{N}(a+nd) = Na + \frac{1}{2}N(N+1)d.$
2. *Arithmetic power series*:

$$ \sum_{n=0}^{N}(a+nb)x^n = \frac{a-(a+bN)x^{N+1}}{1-x} + \frac{bx(1-x^N)}{(1-x)^2}, \qquad (x \neq 1). $$

3. *Geometric series*: $1 + x + x^2 + x^3 + \cdots = \dfrac{1}{1-x}$ $(|x| < 1)$
4. *Arithmetic–geometric series*:
   $a+(a+b)x+(a+2b)x^2+(a+3b)x^3+\cdots = \dfrac{a}{1-x} + \dfrac{bx}{(1-x)^2}$ for $(|x| < 1)$.
5. *Combinatorial sums* (x *can be complex*):

   (a) $\displaystyle\sum_{k=0}^{n}\binom{x-k}{n-k} = \binom{x+1}{n}$

   (b) $\displaystyle\sum_{k=-\infty}^{m}(-1)^k\binom{x}{k} = (-1)^m\binom{x-1}{m}$

   (c) $\displaystyle\sum_{k=0}^{n}\binom{k+m}{k} = \binom{m+n+1}{n}$

   (d) $\displaystyle\sum_{k=-\infty}^{m}(-1)^k\binom{x+m}{k} = \binom{-x}{m}$

   (e) $\displaystyle\sum_{k=-\infty}^{\infty}\binom{x}{m+k}\binom{y}{n-k} = \binom{x+y}{m+n}$

(f) $\sum_{k=-\infty}^{\infty} \binom{l}{m+k}\binom{x}{n+k} = \binom{l+x}{l-m+n}$

(g) $\sum_{k=0}^{l} \binom{l-k}{m}\binom{q+k}{n} = \binom{l+q+1}{m+n+1}$ (for $m \geq q$)

6. *Generating functions*:

   (a) Bessel functions: $\sum_{k=-\infty}^{\infty} J_k(x)z^k = \exp(\frac{1}{2}x\frac{z^2-1}{z})$

   (b) Chebyshev polynomials: $\sum_{n=1}^{\infty} T_n(x)z^n = \frac{z(z+2x)}{2xz-z^2-1}$

   (c) Hermite polynomials: $\sum_{n=0}^{\infty} \frac{H_n(x)}{n!}z^n = \exp(2xz - z^2)$

   (d) Laguerre polynomials: $\sum_{n=0}^{\infty} L_n^{(\alpha)}(x)z^n = (1-z)^{-\alpha-1}\exp[\frac{xz}{z-1}]$

   (e) Legendre polynomials: $\sum_{n=0}^{\infty} P_n(z)x^n = \frac{1}{\sqrt{1-xz+x^2}}$, for $|x| < 1$

7. *Multiple series*:

   (a) $\sum \frac{(-1)^{l+m+n}}{\sqrt{(l+1/6)^2+(m+1/6)^2+(n+1/6)^2}} = \sqrt{3}$
   where $-\infty < l, m, n < \infty$ and they are not all zero

   (b) $\sum \frac{1}{(m^2+n^2)^z} = 4\beta(z)\zeta(z)$ for $-\infty < m, n < \infty$ not both zero

   (c) $\sum_{m,n=0}^{\infty} \frac{(-1)^n}{n!}\frac{\Gamma(n+1/2)}{\Gamma(m+n+1/2)}z^{m+n} = \sqrt{\pi}e^z\frac{\operatorname{erf}(\sqrt{z}-z)}{\sqrt{z}-z}$ for $z > 0$

   (d) $\sum_{m,n=1}^{\infty} \frac{m^2-n^2}{(m^2+n^2)^2} = \frac{\pi}{4}$

   (e) $\sum \frac{1}{k_1^2k_2^2\dots k_n^2} = \frac{\pi^{2n}}{(2n+1)!}$ for $1 \leq k_1 < \cdots < k_n < \infty$

8. *Lagrange series*: If $f(z)$ is analytic at $z = z_0$, $f(z_0) = w_0$, and $f'(z_0) \neq 0$, then the equation $w = f(z)$ can be inverted to obtain the unique solution $z = F(w)$. If both functions are expanded

$$\begin{aligned} f(z) &= f_0 + f_1(z-z_0) + f_2(z-z_0)^2 + \dots \\ F(w) &= F_0 + F_1(w-w_0) + F_2(w-w_0)^2 + \dots \end{aligned} \tag{1.9.3}$$

with $F_0 = F(w_0) = z$, then

$$F_j = \frac{1}{j!}\left[\frac{d^{j-1}}{dz^{j-1}}\left\{\frac{z-z_0}{f(z)-f_0}\right\}^j\right]\Bigg|_{z=z_0} \tag{1.9.4}$$

For example: $F_1 = \frac{1}{f_1}$, $F_2 = -\frac{f_2}{f_1^3}$, $F_3 = \frac{2f_2^2 - f_1f_3}{f_1^5}, \dots$.

1.9.5 FOURIER SERIES

If $f(x)$ is a bounded periodic function of period $2L$ (that is, $f(x+2L) = f(x)$) and satisfies the Dirichlet conditions,

1. In any period, $f(x)$ is continuous, except possibly for a finite number of jump discontinuities.
2. In any period $f(x)$ has only a finite number of maxima and minima.

Then $f(x)$ may be represented by the Fourier series,

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right), \tag{1.9.5}$$

where $\{a_n\}$ and $\{b_n\}$ are determined as follows:

$$\begin{aligned} a_n &= \frac{1}{L} \int_{\alpha}^{\alpha+2L} f(x) \cos \frac{n\pi x}{L}\, dx \qquad \text{for } n = 0, 1, 2, \dots, \\ &= \frac{1}{L} \int_{0}^{2L} f(x) \cos \frac{n\pi x}{L}\, dx, \\ &= \frac{1}{L} \int_{-L}^{L} f(x) \cos \frac{n\pi x}{L}\, dx; \end{aligned} \tag{1.9.6}$$

$$\begin{aligned} b_n &= \frac{1}{L} \int_{\alpha}^{\alpha+2L} f(x) \sin \frac{n\pi x}{L}\, dx \qquad \text{for } n = 1, 2, 3, \dots, \\ &= \frac{1}{L} \int_{0}^{2L} f(x) \sin \frac{n\pi x}{L}\, dx, \\ &= \frac{1}{L} \int_{-L}^{L} f(x) \sin \frac{n\pi x}{L}\, dx, \end{aligned} \tag{1.9.7}$$

where α is any real number (the second and third lines of each formula represent $\alpha = 0$ and $\alpha = -L$, respectively).

The series in Equation (1.9.5) will converge (in the Cesaro sense) to every point where $f(x)$ is continuous, and to $\dfrac{f(x^+) + f(x^-)}{2}$ (i.e., the average of the left-hand and right-hand limits) at every point where $f(x)$ has a jump discontinuity.

1.9.5.1 Special cases

1. If, in addition to the Dirichlet conditions in Section 1.9.5, $f(x)$ is an even function (i.e., $f(x) = f(-x)$), then the Fourier series becomes

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos \frac{n\pi x}{L}. \tag{1.9.8}$$

That is, every $b_n = 0$. In this case, the $\{a_n\}$ may be determined from

$$a_n = \frac{2}{L} \int_0^L f(x) \cos \frac{n\pi x}{L}\, dx \qquad n = 0, 1, 2, \dots. \tag{1.9.9}$$

If, in addition to the above requirements, $f(x) = -f(L - x)$, then a_n will be zero for all even values of n. In this case the expansion becomes

$$f(x) = \sum_{m=1}^{\infty} a_{2m-1} \cos \frac{(2m-1)\pi x}{L}. \tag{1.9.10}$$

2. If, in addition to the Dirichlet conditions in Section 1.9.5, $f(x)$ is an odd function (i.e., $f(x) = -f(-x)$), then the Fourier series becomes

$$f(x) = \sum_{n=1}^{\infty} b_n \sin \frac{n\pi x}{L}. \tag{1.9.11}$$

That is, every $a_n = 0$. In this case, the $\{b_n\}$ may be determined from

$$b_n = \frac{2}{L} \int_0^L f(x) \sin \frac{n\pi x}{L}\, dx \qquad n = 1, 2, 3, \ldots. \tag{1.9.12}$$

If, in addition to the above requirements, $f(x) = f(L - x)$, then b_n will be zero for all even values of n. In this case the expansion becomes

$$f(x) = \sum_{m=1}^{\infty} b_{2m-1} \sin \frac{(2m-1)\pi x}{L}. \tag{1.9.13}$$

The series in Equation (1.9.10) and Equation (1.9.13) are known as odd harmonic series, since only the odd harmonics appear. Similar rules may be stated for even harmonic series, but when a series appears in even harmonic form, it means that $2L$ has not been taken to be the smallest period of $f(x)$. Since any integral multiple of a period is also a period, series obtained in this way will also work, but, in general, computation is simplified if $2L$ is taken as the least period.

Writing the trigonometric functions in terms of complex exponentials, we obtain the complex form of the Fourier series known as the *complex Fourier series* or as the *exponential Fourier series*. It is represented by

$$f(x) = \sum_{n=-\infty}^{\infty} c_n e^{i\omega_n x} \tag{1.9.14}$$

where $\omega_n = \dfrac{n\pi}{L}$ for $n = 0, \pm 1, \pm 2, \ldots$ and the $\{c_n\}$ are determined from

$$c_n = \frac{1}{2L} \int_{-L}^{L} f(x) e^{-i\omega_n x}\, dx. \tag{1.9.15}$$

The set of coefficients $\{c_n\}$ is often referred to as the *Fourier spectrum*.

1.9.5.2 Alternate forms

The Fourier series in Equation (1.9.5) may be represented in the alternate forms:

1. When $\phi_n = \tan^{-1}(-a_n/b_n)$, $a_n = c_n \sin\phi_n$, $b_n = -c_n \cos\phi_n$, and $c_n = \sqrt{a_n^2 + b_n^2}$, then
$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} c_n \sin\left(\frac{n\pi x}{L} + \phi_n\right). \qquad (1.9.16)$$

2. When $\phi_n = \tan^{-1}(a_n/b_n)$, $a_n = c_n \sin\phi_n$, $b_n = c_n \cos\phi_n$, and $c_n = \sqrt{a_n^2 + b_n^2}$, then
$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} c_n \cos\left(\frac{n\pi x}{L} + \phi_n\right). \qquad (1.9.17)$$

1.9.5.3 Fourier series

1. If $f(x)$ has the Laplace transform $F(k) = \int_0^\infty e^{-xk} f(x)\,dx$, then
$$\begin{aligned}\sum_{k=1}^{\infty} F(k)\cos(kt) &= \frac{1}{2}\int_0^\infty \frac{\cos(t) - e^{-x}}{\cosh(x) - \cos(t)} f(x)\,dx,\\ \sum_{k=1}^{\infty} F(k)\sin(kt) &= \frac{1}{2}\int_0^\infty \frac{\sin(t)\,f(x)}{\cosh(x) - \cos(t)}\,dx.\end{aligned} \qquad (1.9.18)$$

2. $\displaystyle\sum_{n=1}^{\infty} \frac{\sin(2n\pi x)}{n^{2k+1}} = \frac{(-1)^{k-1}}{2}\frac{(2\pi)^{2k+1}}{2k+1} B_{2k+1}(x)$, for $0 < x < \frac{1}{2}$
3. $\displaystyle\sum_{n=1}^{\infty} \frac{\cos(2n\pi x)}{n^{2k}} = \frac{(-1)^{k-1}}{2}\frac{(2\pi)^{2k}}{(2k)!} B_{2k}(x)$ for $0 < x < \frac{1}{2}$
4. $\displaystyle\sum_{n=1}^{\infty} a^n \sin(nx) = \frac{a\sin(x)}{1 - 2a\cos(x) + a^2}$ for $|a| < 1$
5. $\displaystyle\sum_{n=0}^{\infty} a^n \cos(nx) = \frac{1 - a\cos(x)}{1 - 2a\cos(x) + a^2}$ for $|a| < 1$

1.9.5.4 Useful series

(a) $1 = \dfrac{4}{\pi}\left[\sin\dfrac{\pi x}{L} + \dfrac{1}{3}\sin\dfrac{3\pi x}{L} + \dfrac{1}{5}\sin\dfrac{5\pi x}{L} + \dots\right]$ $\quad (0 < x < L)$

(b) $x = \dfrac{2L}{\pi}\left[\sin\dfrac{\pi x}{L} - \dfrac{1}{2}\sin\dfrac{2\pi x}{L} + \dfrac{1}{3}\sin\dfrac{3\pi x}{L} + \dots\right]$ $\quad (-L < x < L)$

(c) $x = \dfrac{L}{2} - \dfrac{4L}{\pi^2}\left[\cos\dfrac{\pi x}{L} + \dfrac{1}{3^2}\cos\dfrac{3\pi x}{L} + \dfrac{1}{5^2}\cos\dfrac{5\pi x}{L} + \dots\right]$ $\quad (0 < x < L)$

(d)
$$x^2 = \frac{2L^2}{\pi^3}\left[\left(\frac{\pi^2}{1} - \frac{4}{1}\right)\sin\frac{\pi x}{L} - \frac{\pi^2}{2}\sin\frac{2\pi x}{L} + \left(\frac{\pi^2}{3} - \frac{4}{3^3}\right)\sin\frac{3\pi x}{L} - \frac{\pi^2}{4}\sin\frac{4\pi x}{L} + \left(\frac{\pi^2}{5} - \frac{4}{5^3}\right)\sin\frac{5\pi x}{L}\dots\right] \quad (0 < x < L)$$

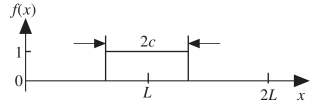
(e) $x^2 = \frac{L^2}{3} - \frac{4L^2}{\pi^2}\left[\cos\frac{\pi x}{L} - \frac{1}{2^2}\cos\frac{2\pi x}{L} + \frac{1}{3^2}\cos\frac{3\pi x}{L} - \frac{1}{4^2}\cos\frac{4\pi x}{L} + \dots\right]$
$(-L < x < L)$

(f) $\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \dots$

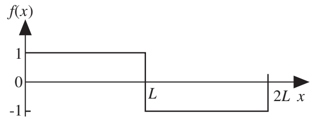
(g) $\frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \dots = 2\left(1 - \frac{1}{2^2} + \frac{1}{3^2} - \frac{1}{4^2} + \dots\right)$

1.9.5.5 Expansions of basic periodic functions

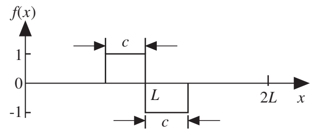
(a) $f(x) = \frac{c}{L} + \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\sin\frac{n\pi c}{L}\cos\frac{n\pi x}{L}$



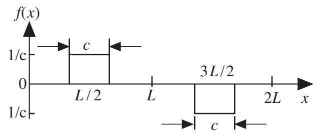
(b) $f(x) = \frac{4}{\pi}\sum_{n=1,3,5,\dots}\frac{1}{n}\sin\frac{n\pi x}{L}$



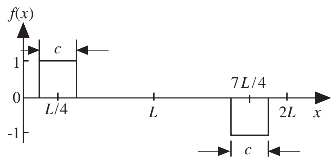
(c) $f(x) = \frac{2}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\left(\cos\frac{n\pi c}{L} - 1\right)\sin\frac{n\pi x}{L}$



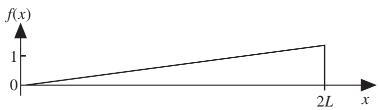
(d) $f(x) = \frac{2}{L}\sum_{n=1}^{\infty}\sin\frac{n\pi}{2}\,\frac{\sin(n\pi c/2L)}{n\pi c/2L}\,\sin\frac{n\pi x}{L}$



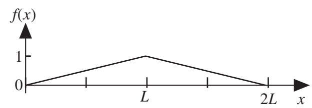
(e) $f(x) = \frac{4}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{n\pi}{4}\sin n\pi a\,\sin\frac{n\pi x}{L}$ $\qquad (a = \frac{c}{2L})$



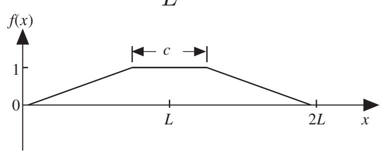
(f) $f(x) = \frac{1}{2} - \frac{1}{\pi}\sum_{n=1}^{\infty}\frac{1}{n}\sin\frac{n\pi x}{L}$



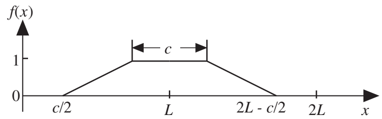
(g) $f(x) = \frac{1}{2} - \frac{4}{\pi^2}\sum_{n=1,3,5,\dots}\frac{1}{n^2}\cos\frac{n\pi x}{L}$



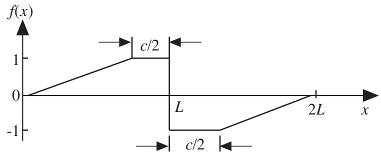
(h) $f(x) = \frac{1+a}{2} + \frac{2}{\pi^2(1-a)}\sum_{n=1}^{\infty}\frac{1}{n^2}\left[(-1)^n\cos n\pi a - 1\right]\cos\frac{n\pi x}{L}$ $\qquad (a = \frac{c}{2L})$



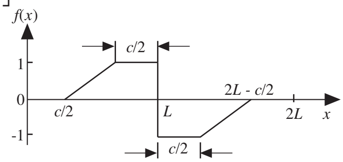
(i) $f(x) = \frac{1}{2} - \frac{4}{\pi^2(1-2a)} \sum_{n=1,3,5,\ldots} \frac{1}{n^2} \cos n\pi a \cos \frac{n\pi x}{L}$ $\quad (a = \frac{c}{2L})$



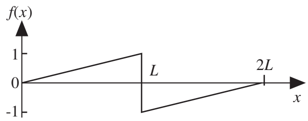
(j) $f(x) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left[1 + \frac{\sin n\pi a}{n\pi(1-a)}\right] \sin \frac{n\pi x}{L}$ $\quad (a = \frac{c}{2L})$



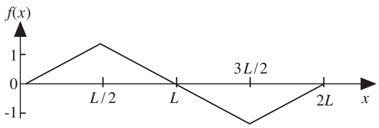
(k) $f(x) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n} \left[1 + \frac{1+(-1)^n}{n\pi(1-2a)} \sin n\pi a\right] \sin \frac{n\pi x}{L}$ $\quad (a = \frac{c}{2L})$



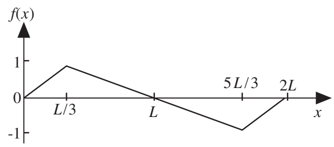
(l) $f(x) = \frac{2}{\pi} \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n} \sin \frac{n\pi x}{L}$



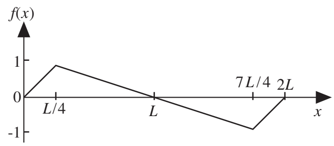
(m) $f(x) = \frac{8}{\pi^2} \sum_{n=1,3,5,\ldots} \frac{(-1)^{(n-1)/2}}{n^2} \sin \frac{n\pi x}{L}$



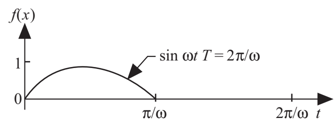
(n) $f(x) = \frac{9}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin \frac{n\pi}{3} \sin \frac{n\pi x}{L}$



(o) $f(x) = \frac{32}{3\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin \frac{n\pi}{4} \sin \frac{n\pi x}{L}$



(p) $f(x) = \frac{1}{\pi} + \frac{1}{2} \sin \omega t - \frac{2}{\pi} \sum_{n=2,4,6,\ldots} \frac{1}{n^2-1} \cos n\omega t$



1.9.6 SERIES EXPANSIONS OF SPECIAL FUNCTIONS

1.9.6.1 Algebraic functions

$$(x+y)^n = x^n + \binom{n}{1}x^{n-1}y + \binom{n}{2}x^{n-2}y^2 + \dots.$$

$$(1 \pm x)^n = 1 \pm \binom{n}{1}x + \binom{n}{2}x^2 \pm \binom{n}{3}x^3 + \dots, \qquad (x^2 < 1).$$

$$(1 \pm x)^{-n} = 1 \mp \binom{n}{1}x + \binom{n+1}{2}x^2 \mp \binom{n+2}{3}x^3 + \dots, \qquad (x^2 < 1).$$

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{8}x^2 + \frac{1}{16}x^3 - \frac{5}{128}x^4 + \dots, \qquad (x^2 < 1).$$

$$(1+x)^{-1/2} = 1 - \frac{1}{2}x + \frac{3}{8}x^2 - \frac{5}{16}x^3 + \frac{35}{128}x^4 + \dots, \qquad (x^2 < 1).$$

$$(1 \pm x)^{-1} = 1 \mp x + x^2 \mp x^3 + x^4 \mp x^5 + \dots, \qquad (x^2 < 1).$$

$$(1 \pm x)^{-2} = 1 \mp 2x + 3x^2 \mp 4x^3 + 5x^4 \mp 6x^5 + \dots, \qquad (x^2 < 1).$$

1.9.6.2 Exponential functions

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \dots + \frac{1}{n!} + \cdots.$$

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \dots + \frac{x^n}{n!} + \cdots, \qquad \text{(all real values of } x\text{)}$$

$$= \frac{1}{1-x} + \sum_{n=1}^{\infty} \frac{x^{n+1}}{n!(x-n)(n+1-x)} \qquad (x \text{ not a positive integer})$$

$$= e^a \left[1 + (x-a) + \frac{(x-a)^2}{2!} + \dots + \frac{(x-a)^n}{n!} + \dots\right].$$

$$a^x = 1 + x\log_e a + \frac{(x\log_e a)^2}{2!} + \dots + \frac{(x\log_e a)^n}{n!} + \cdots.$$

(all real values of x)

1.9.6.3 Logarithmic functions

$$\log x = \frac{x-1}{x} + \frac{1}{2}\left(\frac{x-1}{x}\right)^2 + \dots + \frac{1}{n}\left(\frac{x-1}{x}\right)^n + \cdots, \qquad (x > \tfrac{1}{2}),$$

$$= (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \dots, \qquad (2 \ge x > 0),$$

$$= 2\left[\frac{x-1}{x+1} + \frac{1}{3}\left(\frac{x-1}{x+1}\right)^3 + \frac{1}{5}\left(\frac{x-1}{x+1}\right)^5 + \dots\right], \qquad (x > 0).$$

$$= \log a + \frac{(x-a)}{a} - \frac{(x-a)^2}{2a^2} + \frac{(x-a)^3}{3a^2} - \dots, \qquad (0 < x \le 2a).$$

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \dots, \qquad (-1 < x \le 1).$$

$$\log(x+1) = \log(x-1) + 2\left[\frac{1}{x} + \frac{1}{3x^3} + \frac{1}{5x^5} + \dots\right]. \qquad (x > 1).$$

$$\log(a+x) = \log a + 2\left[\frac{x}{2a+x} + \frac{1}{3}\left(\frac{x}{2a+x}\right)^3 + \frac{1}{5}\left(\frac{x}{2a+x}\right)^5 + \dots\right],$$

$$(a > 0, -a < x).$$

$$\log\frac{1+x}{1-x} = 2\left[x + \frac{x^3}{3} + \dots + \frac{x^{2n-1}}{2n-1} + \dots\right], \qquad (-1 < x < 1).$$

1.9.6.4 Trigonometric functions

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots. \qquad \text{(all real values of } x\text{)}.$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \dots. \qquad \text{(all real values of } x\text{)}.$$

$$\tan x = x + \frac{x^3}{3} + \frac{2x^5}{15} + \dots 5! + \frac{(-1)^{n-1}2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}x^{2n-1} + \dots$$

$$(x^2 < \tfrac{\pi^2}{4} \text{ and } B_n \text{ is the } n^{\text{th}} \text{ Bernoulli number}).$$

$$\cot x = \frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \frac{2x^5}{945} - \frac{x^7}{4725} - \dots + \frac{(-1)^{n+1}2^{2n}B_{2n}}{(2n)!}x^{2n-1} + \dots$$

$$(x^2 < \pi^2 \text{ and } B_n \text{ is the } n^{\text{th}} \text{ Bernoulli number}).$$

$$\sec x = 1 + \frac{x^2}{2} + \frac{5}{24}x^4 + \frac{61}{720}x^6 + \frac{277}{8064}x^8 + \dots + \frac{(-1)^n E_{2n}}{(2n)!}x^{2n} + \dots$$

$$(x^2 < \tfrac{\pi^2}{4} \text{ and } E_n \text{ is the } n^{\text{th}} \text{ Euler number}).$$

$$\csc x = \frac{1}{x} + \frac{x}{6} + \frac{7x^3}{360} + \frac{31x^5}{15120} + \dots + \frac{(-1)^{n+1}2(2^{2n-1}-1)B_{2n}}{(2n)!}x^{2n-1} + \dots$$

$$(x^2 < \pi^2 \text{ and } B_n \text{ is the } n^{\text{th}} \text{ Bernoulli number}).$$

$$\log\sin x = \log x - \frac{x^2}{6} - \frac{x^4}{180} - \frac{x^6}{2835} - \dots \qquad (x^2 < \pi^2).$$

$$\log\cos x = -\frac{x^2}{2} - \frac{x^4}{12} - \frac{x^6}{45} - \frac{17x^6}{2520} - \dots \qquad (x^2 < \tfrac{\pi^2}{4}).$$

$$\log\tan x = \log x + \frac{x^2}{3} + \frac{7x^4}{90} + \frac{62x^6}{2835} + \dots \qquad (x^2 < \tfrac{\pi^2}{4}).$$

$$\sin x = \sin a + (x-a)\cos a - \frac{(x-a)^2}{2!}\sin a - \frac{(x-a)^3}{3!}\cos a + \dots.$$

1.9.6.5 Inverse trigonometric functions

$$\sin^{-1} x = x + \frac{1}{2\cdot 3}x^3 + \frac{1\cdot 3}{2\cdot 4\cdot 5}x^5 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}x^7 + \dots \qquad (x^2 < 1, -\tfrac{\pi}{2} < \sin^{-1} x < \tfrac{\pi}{2}).$$

$$\cos^{-1} x = \frac{\pi}{2} - \left(x + \frac{1}{2\cdot 3}x^3 + \frac{1\cdot 3}{2\cdot 4\cdot 5}x^5 + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}x^7 + \dots\right) \qquad (x^2 < 1, 0 < \cos^{-1} x < \pi).$$

$$\begin{aligned}
\tan^{-1} x &= x - \frac{x^3}{3} + \frac{x^5}{5} - \frac{x^7}{7} + \dots && (x^2 < 1),\\
&= \frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \frac{1}{7x^7} - \dots && (x > 1),\\
&= -\frac{\pi}{2} - \frac{1}{x} + \frac{1}{3x^3} - \frac{1}{5x^5} + \frac{1}{7x^7} - \dots && (x < -1).\\
\cot^{-1} x &= \frac{\pi}{2} - x + \frac{x^3}{3} - \frac{x^5}{5} + \frac{x^7}{7} - \dots && (x^2 < 1).
\end{aligned}$$

1.9.6.6 Hyperbolic functions

$$\begin{aligned}
\sinh x &= x + \frac{x^3}{3!} + \frac{x^5}{5!} + \frac{x^7}{7!} + \dots + \frac{x^{(2n+1)}}{(2n+1)!} + \dots.\\
\sinh ax &= \frac{2}{\pi}\sinh \pi a\left[\frac{\sin x}{a^2+1^2} - \frac{2\sin 2x}{a^2+2^2} + \frac{3\sin 3x}{a^2+3^2} + \dots\right] && (|x| < \pi).\\
\cosh x &= 1 + \frac{x^2}{2!} + \frac{x^4}{4!} + \frac{x^6}{6!} + \dots + \frac{x^{2n}}{(2n)!} + \dots.\\
\cosh ax &= \frac{2a}{\pi}\sinh \pi a\left[\frac{1}{2a^2} - \frac{\cos x}{a^2+1^2} + \frac{\cos 2x}{a^2+2^2} - \frac{\cos 3x}{a^2+3^2} + \dots\right] && (|x| < \pi).\\
\tanh x &= x - \frac{1}{3}x^3 + \frac{2}{15}x^5 - \dots + \frac{2^{2n}(2^{2n}-1)B_{2n}}{(2n)!}x^{2n-1} + \dots && (|x| < \tfrac{\pi}{2}),\\
&= 1 - 2e^{-2x} + 2e^{-4x} - 2e^{-6x} + \dots && (\text{Re } x > 0),\\
&= 2x\left[\frac{1}{\left(\frac{\pi}{2}\right)^2 + x^2} + \frac{1}{\left(\frac{3\pi}{2}\right)^2 + x^2} + \frac{1}{\left(\frac{5\pi}{2}\right)^2 + x^2} + \dots\right].\\
\coth x &= \frac{1}{x} + \frac{x}{3} - \frac{x^3}{45} + \frac{2x^5}{945} + \dots + \frac{2^{2n}B_{2n}}{(2n)!}x^{2n-1} + \dots && (0 < |x| < \pi),\\
&= 1 + 2e^{-2x} + 2e^{-4x} + 2e^{-6x} + \dots && (\text{Re } x > 0),
\end{aligned}$$

$$\operatorname{sech} x = 1 - \frac{1}{2!}x^2 + \frac{5}{4!}x^4 - \frac{61}{6!}x^6 + \cdots + \frac{E_{2n}}{(2n)!}x^{2n} + \ldots$$

$$(|x| < \tfrac{\pi}{2},\ E_n \text{ is the } n^{\text{th}} \text{ Euler number}),$$

$$= 2\left(e^{-x} - e^{-3x} + e^{-5x} - e^{-7x} + \ldots\right) \qquad (\operatorname{Re} x > 0),$$

$$\operatorname{csch} x = \frac{1}{x} - \frac{x}{6} + \frac{7x^3}{360} + \cdots + \frac{2(2^{2n-1}-1)B_{2n}}{(2n)!}x^{2n-1} + \ldots \qquad (0 < |x| < \pi),$$

$$= 2\left(e^{-x} + e^{-3x} + e^{-5x} + e^{-7x} + \ldots\right) \qquad (\operatorname{Re} x > 0),$$

$$\begin{aligned}\sinh nu = \sinh u \Big[&(2\cosh u)^{n-1} - \frac{(n-2)}{1!}(2\cosh u)^{n-3} \\ &+ \frac{(n-3)(n-4)}{2!}(2\cosh u)^{n-5} \\ &- \frac{(n-4)(n-5)(n-6)}{3!}(2\cosh u)^{n-7} + \ldots\Big].\end{aligned}$$

$$\begin{aligned}\cosh nu = \frac{1}{2}\Big[&(2\cosh u)^{n} - \frac{n}{1!}(2\cosh u)^{n-2} + \frac{n(n-3)}{2!}(2\cosh u)^{n-4} \\ &- \frac{n(n-4)(n-5)}{3!}(2\cosh u)^{n-6} + \ldots\Big].\end{aligned}$$

1.9.6.7 Inverse hyperbolic functions

$$\sinh^{-1} x = x - \frac{1}{2\cdot 3}x^3 + \frac{1\cdot 3}{2\cdot 4\cdot 5}x^5 - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 7}x^7 + \ldots \qquad (|x| < 1),$$

$$= \log(2x) + \frac{1}{2}\cdot\frac{1}{2x^2} + \frac{1\cdot 3}{2\cdot 4}\cdot\frac{1}{4x^4} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\cdot\frac{1}{6x^6} + \ldots \qquad (|x| > 1).$$

$$\cosh^{-1} x = \pm\left[\log(2x) - \frac{1}{2}\cdot\frac{1}{2x^2} - \frac{1\cdot 3}{2\cdot 4}\cdot\frac{1}{4x^4} + \ldots\right] \qquad (x > 1).$$

$$\operatorname{csch}^{-1} x = \frac{1}{x} - \frac{1}{2}\cdot\frac{1}{3x^3} - \frac{1\cdot 3}{2\cdot 4}\cdot\frac{1}{5x^5} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\cdot\frac{1}{7x^7} + \ldots \qquad (|x| > 1),$$

$$= \log\frac{2}{x} + \frac{1}{2}\cdot\frac{x^2}{2} - \frac{1\cdot 3}{2\cdot 4}\cdot\frac{x^4}{4} + \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\cdot\frac{x^6}{6} - \ldots \qquad (0 < x < 1).$$

$$\operatorname{sech}^{-1} x = \log\frac{2}{x} - \frac{1}{2}\cdot\frac{x^2}{2} - \frac{1\cdot 3}{2\cdot 4}\cdot\frac{x^4}{4} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\cdot\frac{x^6}{6} - \ldots \qquad (0 < x < 1).$$

$$\tanh^{-1} x = x + \frac{x^3}{3} + \frac{x^5}{5} + \frac{x^7}{7} + \cdots + \frac{x^{2n+1}}{2n+1} + \ldots \qquad (|x| < 1).$$

$$\coth^{-1} x = \frac{1}{x} + \frac{1}{3x^3} + \frac{1}{5x^5} + \frac{1}{7x^7} + \cdots + \frac{1}{(2n+1)x^{2n+1}} + \ldots \qquad (|x| > 1).$$

$$\operatorname{gd} x = x - \frac{1}{6}x^3 + \frac{1}{24}x^5 + \cdots + \frac{E_{2n}}{(2n+1)!}x^{2n+1} + \ldots \qquad (|x| < 1).$$

1.9.7 SUMMATION FORMULAS

1. *Euler–MacLaurin summation formula*: As $n \to \infty$,

$$\sum_{k=0}^{n} f(k) \sim \frac{1}{2} f(n) + \int_0^n f(x)\,dx + C + \sum_{j=1}^{\infty} (-1)^{j+1} B_{j+1} \frac{f^{(j)}(n)}{(j+1)!}$$

where B_j is the j^{th} Bernoulli number and

$$C = \lim_{m\to\infty} \left[\sum_{j=1}^{m} (-1)^j B_{j+1} \frac{f^{(j)}(0)}{(j+1)!} + \frac{1}{2} f(0) + \frac{(-1)^m}{(m+1)!} \int_0^\infty B_{m+1}(x - \lfloor x \rfloor) f^{(m+1)}(x)\,dx \right].$$

2. *Poisson summation formula*: If f is continuous,

$$\frac{1}{2} f(0) + \sum_{n=1}^{\infty} f(n) = \int_0^\infty f(x)\,dx + 2 \sum_{n=1}^{\infty} \left[\int_0^\infty f(x) \cos(2n\pi x) dx \right].$$

3. *Plana's formula*: If f is analytic,

$$\sum_{k=1}^{n} f(k) = \frac{1}{2} f(n) + \int_a^n f(x)\,dx + \sum_{j=1}^{\infty} (-1)^{j-1} B_j \frac{f^{(2j+1)}(n)}{2j!}$$

where a is a constant dependent on f and B_j is the j^{th} Bernoulli number.

EXAMPLES

1. $\displaystyle\sum_{k=1}^{n} \frac{1}{k} \sim \log n + \gamma + \frac{1}{2n} - \frac{B_2}{2n^2} - \dots$ where γ is Euler's constant.
2. $\displaystyle 1 + 2\sum_{n=1}^{\infty} e^{-n^2 x} = \sqrt{\frac{\pi}{x}} \left[1 + 2 \sum_{n=1}^{\infty} e^{-\pi^2 n^2 / x} \right]$ (Jacobi)

1.9.8 FASTER CONVERGENCE: SHANKS TRANSFORMATION

Let s_n be the n^{th} partial sum. The sequences $\{S(s_n)\},\ \{S(S(s_n))\},\dots$ often converge successively more rapidly to the same limit as $\{s_n\}$, where

$$S(s_n) = \frac{s_{n+1}s_{n-1} - s_n^2}{s_{n+1} + s_{n-1} - 2s_n}. \tag{1.9.19}$$

EXAMPLE For $s_n = \sum_{k=0}^{n} (-1)^k z^k$, we find $S(s_n) = \dfrac{1}{1+z}$ for all n.

1.9.9 SUMMABILITY METHODS

Unique values can be assigned to divergent series in a variety of ways which preserve the values of convergent series.

1. Abel summation: $\displaystyle\sum_{n=0}^{\infty} a_n = \lim_{r\to 1^-} \sum_{n=0}^{\infty} a_n r^n$.
2. Cesaro $(C,1)$-summation: $\displaystyle\sum_{n=0}^{\infty} a_n = \lim_{N\to\infty} \frac{s_0 + s_1 + \dots s_N}{N+1}$
   where $s_n = \sum_{m=0}^{n} a_m$ are the partial sums.
3. Ramanujan summation: $\displaystyle\sum_{k=1}^{\infty} f(k) = -\frac{1}{2} f(0) - \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} f^{(2k-1)}(0)$ where $f^{(2k-1)}$ is the $(2k-1)^{\text{th}}$ derivative of f and B_{2k} is the $2k^{\text{th}}$ Bernoulli number.

EXAMPLES

- $1 - 1 + 1 - 1 + \cdots = \frac{1}{2}$ (in the sense of Abel summation)
- $1 - 1 + 0 + 1 - 1 + 0 + 1 - \cdots = \frac{1}{3}$ (in the sense of Cesaro summation)
- $1 + 2 + 3 + \cdots = -\frac{B_2}{2!} = -\frac{1}{12}$ (using $f(x) = x$ in the Ramanujan formula)

1.9.10 OPERATIONS WITH POWER SERIES

Let $y = a_1 x + a_2 x^2 + a_3 x^3 + \dots$, and let $z = z(y) = b_1 x + b_2 x^2 + b_3 x^3 + \dots$.

$z(y)$	b_0	b_1	b_2	b_3
$1/(1-y)$	1	a_1	$a_1^2 + a_2$	$a_1^3 + 2a_1a_2 + a_3$
$\sqrt{1+y}$	1	$\frac{1}{2}a_1$	$-\frac{1}{8}a_1^2 + \frac{1}{2}a_2$	$\frac{1}{16}a_1^3 - \frac{1}{4}a_1a_2 + \frac{1}{2}a_3$
$(1+y)^{-1/2}$	1	$-\frac{1}{2}a_1$	$\frac{3}{8}a_1^2 - \frac{1}{2}a_2$	$-\frac{5}{16}a_1^3 + \frac{3}{4}a_1a_2 - \frac{1}{2}a_3$
e^y	1	a_1	$\frac{1}{2}a_1^2 + a_2$	$\frac{1}{6}a_1^3 + a_1a_2 + a_3$
$\log(1+y)$	0	a_1	$a_2 - \frac{1}{2}a_1^2$	$a_3 - a_1a_2 + \frac{1}{3}a_1^3$
$\sin y$	0	a_1	a_2	$-\frac{1}{6}a_1^3 + a_3$
$\cos y$	1	0	$-\frac{1}{2}a_1^2$	$-a_1a_2$
$\tan y$	0	a_1	a_2	$\frac{1}{3}a_1^3 + a_3$

1.9.11 MISCELLANEOUS SUMS

1. $\sum_{k=1}^{n} \frac{1}{k(k+1)(k+2)} = \frac{1}{4} - \frac{1}{2(n+1)(n+2)}$.
2. $\sum_{k=1}^{\infty} \frac{1}{k(k+1)\cdots(k+p)} = \frac{1}{p \cdot p!}$.
3. $\displaystyle\sum_{k=3}^{\infty} \frac{1}{k \log k (\log\log k)^2}$ converges to 38.4067680928... very slowly. Using naive summation more than $10^{3 \cdot 10^{86}}$ terms are needed for two-decimal places.
4. $\displaystyle\sum_{k=3}^{\infty} \frac{1}{k \log k (\log\log k)}$ diverges, but the partial sums exceed 10 only after a googolplex of terms have appeared.

1.9.12 INFINITE PRODUCTS

For the sequence of complex numbers $\{a_k\}$, an infinite product is $\prod_{k=1}^{\infty}(1+a_k)$. A necessary condition for convergence is that $\lim_{n\to\infty} a_n = 0$. A necessary and sufficient condition for convergence is that $\sum_{k=1}^{\infty}\log(1+a_k)$ converges. Examples:

(a) $\sin z = z\prod_{k=1}^{\infty}\cos\frac{z}{2^k}$

(b) $\sin \pi z = \pi z\prod_{k=1}^{\infty}\left(1-\frac{z^2}{k^2}\right)$

(c) $\cos \pi z = \prod_{k=1}^{\infty}\left(1-\frac{4z^2}{(2k-1)^2}\right)$

(d) $\sinh z = z\prod_{k=1}^{\infty}\left(1+\frac{z^2}{k^2\pi^2}\right)$

(e) $\cosh z = \prod_{k=0}^{\infty}\left(1+\frac{4z^2}{(2k+1)^2\pi^2}\right)$

(f) $z! = \prod_{k=1}^{\infty}\frac{\left(1+\frac{1}{k}\right)^z}{1+\frac{z}{k}}$

(g) $\sin(a+z) = (\sin a)\prod_{k=0,\pm1,\pm2,\ldots}^{\infty}\left(1+\frac{z}{a+k\pi}\right)$

(h) $\cos(a+z) = (\cos a)\prod_{k=\pm1,\pm3,\pm5,\ldots}^{\infty}\left(1+\frac{2z}{2a+k\pi}\right)$

1.9.12.1 Weierstrass theorem

Define $E(w,m) = (1-w)\exp\left(w+\frac{w^2}{2}+\cdots+\frac{w^m}{m}\right)$. For $k=1,2,\ldots$ let $\{b_k\}$ be a sequence of complex numbers such that $|b_k|\to\infty$. Then the infinite product $P(z) = \prod_{k=1}^{\infty}E\left(\frac{z}{b_k},k\right)$ is an entire function with zeros at b_k and at these points only. The multiplicity of the root at b_n is equal to the number of indices j such that $b_j = b_n$.

1.9.13 INFINITE PRODUCTS AND INFINITE SERIES

1. The Rogers–Ramanujan identities (for $a=0$ or $a=1$) are

$$1 + \sum_{k=1}^{\infty}\frac{q^{k^2+ak}}{(1-q)(1-q^2)\cdots(1-q^k)} = \prod_{j=0}^{\infty}\frac{1}{(1-q^{5j+a+1})(1-q^{5j-a+4})}. \tag{1.9.20}$$

2. Jacobi's triple product identity is

$$\sum_{k=-\infty}^{\infty}q^{\binom{k}{2}}x^k = \prod_{j=1}^{\infty}(1-q^j)(1+x^{-1}q^j)(1+xq^{j-1}). \tag{1.9.21}$$

Chapter 2

Algebra

2.1 ELEMENTARY ALGEBRA

2.1.1 BASIC ALGEBRA

2.1.1.1 Laws of exponents

Assuming all quantities are real, $a > 0$, $b > 0$, and no denominators are zero:

$$a^x a^y = a^{x+y}, \qquad \frac{a^x}{a^y} = a^{x-y}, \qquad (ab)^x = a^x b^x,$$

$$a^0 = 1, \qquad a^{-x} = \frac{1}{a^x}, \qquad \left(\frac{a}{b}\right)^x = \frac{a^x}{b^x},$$

$$(a^x)^y = a^{xy}, \qquad a^{\frac{1}{x}} = \sqrt[x]{a}, \qquad \sqrt[x]{ab} = \sqrt[x]{a}\,\sqrt[x]{b},$$

$$\sqrt[x]{\sqrt[y]{a}} = \sqrt[xy]{a}, \qquad a^{\frac{x}{y}} = \sqrt[y]{a^x} = (\sqrt[y]{a})^x, \qquad \sqrt[x]{\frac{a}{b}} = \frac{\sqrt[x]{a}}{\sqrt[x]{b}}.$$

2.1.1.2 Proportion

If $\frac{a}{b} = \frac{c}{d}$, then

- $\frac{a+b}{b} = \frac{c+d}{d}$
- $\frac{a-b}{a+b} = \frac{c-d}{c+d}$
- $\frac{a-b}{b} = \frac{c-d}{d}$
- $ad = bc$
- $\frac{a}{c} = \frac{b}{d}$

2.1.2 PROGRESSIONS

2.1.2.1 Arithmetic progression

An *arithmetic progression* is a sequence of numbers such that the difference of any two consecutive numbers is constant. If the sequence is $a_1, a_2, \dots a_n$, where $a_{i+1} - a_i = d$, then $a_k = a_1 + (k-1)d$ and

$$a_1 + a_2 + \dots + a_n = \frac{n}{2}\left(2a_1 + (n-1)d\right). \tag{2.1.1}$$

In particular, the arithmetic progression $1, 2, \dots, n$ has the sum $n(n+1)/2$.

2.1.2.2 Geometric progression

A *geometric progression* is a sequence of numbers such that the ratio of any two consecutive numbers is constant. If the sequence is $a_1, a_2, \dots a_n$, where $a_{i+1}/a_i = r$, then $a_k = a_1 r^{k-1}$ and

$$a_1 + a_2 + \dots + a_n = \begin{cases} a_1 \dfrac{1-r^n}{1-r} & r \neq 1 \\ na_1 & r = 1. \end{cases} \tag{2.1.2}$$

If $|r| < 1$, then the infinite geometric series $a_1(1 + r + r^2 + r^3 + \cdots)$ converges to $\frac{a_1}{1-r}$. For example, $1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \cdots = 2$ (since $a_1 = 1$ and $r = \frac{1}{2}$).

2.1.2.3 Means

1. The *arithmetic mean* of a and b is given by $\frac{a+b}{2}$. More generally, the arithmetic mean of $a_1, a_2, \ldots, a_n$ is given by $(a_1 + a_2 + \cdots + a_n)/n$.
2. The *geometric mean* of a and b is given by $\sqrt{ab}$. More generally, the geometric mean of $a_1, a_2, \ldots, a_n$ is given by $\sqrt[n]{a_1 a_2 \ldots a_n}$.
3. The *harmonic mean* of a and b is given by $\frac{1}{\frac{1}{2}\left(\frac{1}{a}+\frac{1}{b}\right)} = \frac{2ab}{a+b}$.

If A, G, and H represent the arithmetic, geometric, and harmonic means

1. of two positive values, then $AH = G^2$.
2. of any number of positive values, then $A \geq G \geq H$.

2.1.2.4 Algebraic equations

A *polynomial equation in one variable* has the form $f(x) = 0$ where $f(x)$ is a polynomial of degree n

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0. \tag{2.1.3}$$

and $a_n \neq 0$.

A complex number z is a *root* of the polynomial $f(x)$ if $f(z) = 0$. A complex number z is a *root of multiplicity* k if $f(z) = f'(z) = f''(z) = \cdots = f^{(k-1)}(z) = 0$, but $f^{(k)}(z) \neq 0$. A root of multiplicity 1 is called a *simple root*. A root of multiplicity 2 is called a *double root*, and a root of multiplicity 3 is called a *triple root*.

2.1.2.5 Roots of polynomials

1. *Fundamental theorem of algebra*

   A polynomial equation of degree n has exactly n complex roots, where a double root is counted twice, a triple root three times, and so on. If the n roots of the polynomial $f(x)$ are z_1, z_2, …, z_n (where a double root is listed twice, a triple root three times, and so on), then the polynomial can be written as

   $$f(x) = a_n (x - z_1)(x - z_2) \ldots (x - z_n). \tag{2.1.4}$$

2. If the coefficients of the polynomial, $\{a_0, a_1, \ldots, a_n\}$, are real numbers, then the polynomial will always have an even number of complex roots occurring in pairs. That is, if z is a complex root, then so is $\overline{z}$. If the polynomial has an odd degree and the coefficients are real, then it must have at least one real root.
3. Equations for roots of 2^{nd}, 3^{rd}, and 4^{th} order equations are on pages 73– 74.
4. The coefficients of a polynomial may be expressed as symmetric functions of the roots. For example, the elementary symmetric functions $\{s_i\}$, and their

values for a polynomial of degree n (known as *Vieta's formulas*), are:

$$\begin{aligned} s_1 = z_1 + z_2 + \cdots + z_n &= -\frac{a_{n-1}}{a_n}, \\ s_2 = z_1 z_2 + z_1 z_3 + z_2 z_3 + \cdots = \sum_{i>j} z_i z_j &= \frac{a_{n-2}}{a_n}, \\ &\vdots \\ s_n = z_1 z_2 z_3 \dots z_n &= (-1)^n \frac{a_0}{a_n}. \end{aligned} \tag{2.1.5}$$

where s_k is the sum of $\binom{n}{k}$ products, each product combining k factors of roots without repetition.

5. The *discriminant* of a polynomial is $\prod_{i>j}(z_i - z_j)^2$. (Note that the ordering of the roots is irrelevant.) The discriminant can always be written as a polynomial combination of $a_0, a_1, \dots, a_n$, divided by a_n.
   (a) For the quadratic equation $ax^2 + bx + c = 0$ the discriminant is $\frac{b^2-4ac}{a^2}$.
   (b) For the cubic equation $ax^3 + bx^2 + cx + d = 0$ the discriminant is
   $$\frac{b^2c^2 - 4b^3d - 4ac^3 + 18abcd - 27a^2d^2}{a^4}.$$
6. The number of roots of a polynomial in modular arithmetic is difficult to predict. For example
   (a) $y^4 + y + 1 = 0$ has one root modulo 51: $y = 37$
   (b) $y^4 + y + 2 = 0$ has no root modulo 51
   (c) $y^4 + y + 3 = 0$ has six roots modulo 51: $y = \{15, 27, 30, 32, 44, 47\}$

2.1.2.6 Algebraic identities

$$(a \pm b)^2 = a^2 \pm 2ab + b^2.$$
$$(a \pm b)^3 = a^3 \pm 3a^2b + 3ab^2 \pm b^3.$$
$$(a \pm b)^4 = a^4 \pm 4a^3b + 6a^2b^2 \pm 4ab^3 + b^4.$$
$$(a \pm b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k (\pm b)^{n-k} \quad \text{where } \binom{n}{k} = \frac{n!}{k!(n-k)!}.$$
$$a^2 + b^2 = (a + bi)(a - bi).$$
$$a^4 + b^4 = (a^2 + \sqrt{2}ab + b^2)(a^2 - \sqrt{2}ab + b^2).$$
$$a^2 - b^2 = (a - b)(a + b).$$
$$a^3 - b^3 = (a - b)(a^2 + ab + b^2).$$
$$a^n - b^n = (a - b)(a^{n-1} + a^{n-2}b + \cdots + ab^{n-2} + b^{n-1}).$$
$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$
$$(a + b + c)^3 = a^3 + b^3 + c^3 + 3(a^2b + ab^2 + a^2c + ac^2 + b^2c + bc^2) + 6abc.$$

2.1.3 PARTIAL FRACTIONS

The technique of partial fractions allows a quotient of two polynomials to be written as a sum of simpler terms.

Let $f(x)$ and $g(x)$ be polynomials and let the fraction be $\frac{f(x)}{g(x)}$. If the degree of $f(x)$ is greater than the degree of $g(x)$ then divide $f(x)$ by $g(x)$ to produce a quotient $q(x)$ and a remainder $r(x)$, where the degree of $r(x)$ is less than the degree of $g(x)$. That is, $\frac{f(x)}{g(x)} = q(x) + \frac{r(x)}{g(x)}$. Therefore, assume that the fraction has the form $\frac{r(x)}{g(x)}$, where the degree of the numerator is less than the degree of the denominator. The techniques used to find the partial fraction decomposition of $\frac{r(x)}{g(x)}$ depend on the factorization of $g(x)$.

2.1.3.1 Single linear factor

Suppose that $g(x) = (x-a)h(x)$, where $h(a) \neq 0$. Then

$$\frac{r(x)}{g(x)} = \frac{A}{x-a} + \frac{s(x)}{h(x)}, \tag{2.1.6}$$

where $s(x)$ can be computed and the number A is given by $r(a)/h(a)$. For example (here $r(x) = 2x$, $g(x) = x^2 - 1 = (x-1)(x+1)$, $h(x) = x+1$, and $a = 1$):

$$\frac{2x}{x^2-1} = \frac{1}{x-1} + \frac{1}{x+1}. \tag{2.1.7}$$

2.1.3.2 Repeated linear factor

Suppose that $g(x) = (x-a)^k h(x)$, where $h(a) \neq 0$. Then

$$\frac{r(x)}{g(x)} = \frac{A_1}{x-a} + \frac{A_2}{(x-a)^2} + \cdots + \frac{A_k}{(x-a)^k} + \frac{s(x)}{h(x)}, \tag{2.1.8}$$

for a computable $s(x)$ where

$$\begin{aligned} A_k &= \frac{r(a)}{h(a)}, & A_{k-1} &= \frac{d}{dx}\left(\frac{r(x)}{h(x)}\right)\bigg|_{x=a}, \\ A_{k-2} &= \frac{1}{2!}\frac{d^2}{dx^2}\left(\frac{r(x)}{h(x)}\right)\bigg|_{x=a}, & A_{k-j} &= \frac{1}{j!}\frac{d^j}{dx^j}\left(\frac{r(x)}{h(x)}\right)\bigg|_{x=a}. \end{aligned} \tag{2.1.9}$$

2.1.3.3 Single quadratic factor

Suppose that $g(x) = (x^2 + bx + c)h(x)$, where $b^2 - 4c < 0$ (so that $x^2 + bx + c$ does not factor into real linear factors) and $h(x)$ is relatively prime to $x^2 + bx + c$ (that is $h(x)$ and $x^2 + bx + c$ have no factors in common). Then

$$\frac{r(x)}{g(x)} = \frac{Ax+B}{x^2+bx+c} + \frac{s(x)}{h(x)}. \tag{2.1.10}$$

In order to determine A and B, multiply the equation by $g(x)$ so that there are no denominators remaining, and substitute any two values for x, yielding two equations for A and B.

When A and B are both real, if after multiplying the equation by $g(x)$ a root of x^2+bx+c is substituted for x, then the values of A and B can be inferred from this single complex equation by equating real and imaginary parts. (Since x^2+bx+c divides $g(x)$, there are no zeros in the denominator.) This technique can also be used for repeated quadratic factors (below).

2.1.3.4 Repeated quadratic factor

Suppose that $g(x) = (x^2+bx+c)^k h(x)$, where $b^2-4c<0$ (so that x^2+bx+c does not factor into real linear factors) and $h(x)$ is relatively prime to x^2+bx+c. Then

$$\frac{r(x)}{g(x)} = \frac{A_1x+B_1}{x^2+bx+c} + \frac{A_2x+B_2}{(x^2+bx+c)^2} + \frac{A_3x+B_3}{(x^2+bx+c)^3} + \cdots + \frac{A_kx+B_k}{(x^2+bx+c)^k} + \frac{s(x)}{h(x)}.$$

In order to determine A_i and B_i, multiply the equation by $g(x)$ so that there are no denominators remaining, and substitute any $2k$ values for x, yielding $2k$ equations for A_i and B_i.

2.2 POLYNOMIALS

All polynomials of degree 2, 3, or 4 are solvable by radicals. That is, their roots can be written in terms of algebraic operations ($+$, $-$, $\times$, and $\div$) and root-taking ($\sqrt[n]{\ }$). While some higher degree polynomials can be solved by radicals (e.g., $x^{10}=1$ can be solved), the general polynomial of degree 5 or higher cannot be solved by radicals.

2.2.1 QUADRATIC EQUATION

The solution of the quadratic equation $ax^2+bx+c=0$, where $a\neq 0$, is given by

$$x = \frac{-b\pm\sqrt{b^2-4ac}}{2a}. \tag{2.2.1}$$

The discriminant of this equation is $\Delta = (b^2-4ac)/a^2$. If a, b, and c are all real and

- $\Delta<0$, then the two roots are complex numbers which are conjugate.
- $\Delta>0$, then the two roots are unequal real numbers.
- $\Delta=0$, then the two roots are equal.

2.2.2 CUBIC POLYNOMIALS

Given the cubic equation $Ax^3 + Bx^2 + Cx + D = 0$, divide the coefficients by A so the coefficient of the cubic term is one:

$$x^3 + bx^2 + cx + d = 0 \tag{2.2.2}$$

To find the roots $\{x_1, x_2, x_3\}$ of this cubic define the following:

$$\begin{aligned} E &= \left(3c - b^2\right)/3 \\ F &= \left(2b^3 - 9bc + 27d\right)/27 \\ G &= \frac{F^2}{4} + \frac{E^3}{27} \end{aligned} \tag{2.2.3}$$

If the coefficients $\{b, c, d\}$ are real, then Equation (2.2.2) will have either 3 real roots or 1 real root and 2 imaginary roots. The next steps presume that $\{b, c, d\}$ are real, this allows computations only involving real numbers. If $\{b, c, d\}$ are not real then either of the following methods results in the same 3 roots.

1. If $G < 0$, then there will be 3 real roots. Continue with

   (a) $H = \sqrt{-E^3/27}$
   (b) $I = \sqrt[3]{H}$
   (c) $J = \cos^{-1}\left(-F/2H\right)$
   (d) $K = \cos\left(J/3\right)$
   (e) $L = \sqrt{3}\sin\left(J/3\right)$
   (f) $x_1 = 2IK - \frac{b}{3}$
   (g) $x_2 = -I(K + L) - \frac{b}{3}$
   (h) $x_3 = -I(K - L) - \frac{b}{3}$

2. If $G > 0$, then there will be 1 real root and 2 complex roots. Continue with

   (a) $M = \sqrt[3]{-\frac{F}{2} + \sqrt{G}}$
   (b) $N = -\sqrt[3]{\frac{F}{2} + \sqrt{G}}$
   (c) $x_1 = M + N - \frac{b}{3}$
   (d) $x_2 = -\frac{M+N}{2} - \frac{b}{3} + i\frac{(M-N)\sqrt{3}}{2}$
   (e) $x_3 = -\frac{M+N}{2} - \frac{b}{3} - i\frac{(M-N)\sqrt{3}}{2}$

2.2.3 QUARTIC POLYNOMIALS

Given the quartic equation $b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0 = 0$, divide the coefficients by b_4 so the coefficient of the quartic term is one:

$$x^4 + a_3x^3 + a_2x^2 + a_1x + a_0 = 0 \tag{2.2.4}$$

To find the roots $\{x_1, x_2, x_3, x_4\}$ of this quartic do the following:

1. Define the *resolvant cubic* to be $z^3 + bz^2 + cz + d = 0$ where
$$b = -a_2, \quad c = a_1 a_3 - 4a_0, \quad d = 4a_2 a_0 - a_1^2 - a_3^2 a_0$$
2. Use Section 2.2.2 to find the solutions $\{z_1, z_2, z_3\}$. If the coefficients $\{a_3, a_2, a_1, a_0\}$ are real, then there will be at least one real root z, which we select. If there are no real roots, then select any root z.
3. $R = \sqrt{\dfrac{a_3^2}{4} - a_2 + z}$
4. $S = \dfrac{3a_3^2}{4} - R^2 - 2a_2$
5. $T = \begin{cases} 2\sqrt{z^2 - 4a_0} & \text{if } R = 0 \\ \dfrac{4a_3a_2 - 8a_1 - a_3^3}{4R} & \text{if } R \neq 0 \end{cases}$
6. $x_{1,2} = -\dfrac{1}{4}a_3 + \dfrac{1}{2}R \pm \dfrac{1}{2}\sqrt{S+T}$
7. $x_{3,4} = -\dfrac{1}{4}a_3 + \dfrac{1}{2}R \pm \dfrac{1}{2}\sqrt{S-T}$

2.2.4 RESULTANTS

Let $f(x) = a_n x^n + a_{n-1}x^{n-1} + \cdots + a_0$ and $g(x) = b_m x^m + b_{m-1}x^{m-1} + \cdots + b_0$, where $a_n \neq 0$ and $b_m \neq 0$. The *resultant* of f and g is the determinant of the $(m+n) \times (m+n)$ matrix

$$\det \begin{bmatrix} a_n & a_{n-1} & \cdots & \cdots & a_0 & 0 & \cdots & 0 \\ 0 & a_n & a_{n-1} & \cdots & a_1 & a_0 & & 0 \\ \vdots & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & \cdots & 0 & a_n & a_{n-1} & \cdots & a_1 & a_0 \\ \hdashline 0 & \cdots & \cdots & 0 & b_m & \cdots & b_1 & b_0 \\ \vdots & & & \iddots & \iddots & & & \\ 0 & b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & 0 \\ b_m & b_{m-1} & \cdots & b_0 & 0 & \cdots & \cdots & 0 \end{bmatrix} \tag{2.2.5}$$

The resultant of f and g is 0 if and only if f and g have a common root.

EXAMPLES

1. For $f(x) = x^2 + 2x + 3$ and $g(x;\alpha) = 4x^3 + 5x^2 + 6x + (7+\alpha)$, the resultant is
$$\det \begin{bmatrix} 1 & 2 & 3 & 0 & 0 \\ 0 & 1 & 2 & 3 & 0 \\ 0 & 0 & 1 & 2 & 3 \\ 0 & 4 & 5 & 6 & 7+\alpha \\ 4 & 5 & 6 & 7+\alpha & 0 \end{bmatrix} = (16+\alpha)^2$$
Note that $g(x;-16) = (4x-3)(x^2+2x+3) = (4x-3)f(x)$.
2. The resultant of $ax+b$ and $cx+d$ is $da - bc$.
3. The resultant of $(x+a)^5$ and $(x+b)^5$ is $(b-a)^{25}$.

2.2.5 POLYNOMIAL NORMS

The polynomial $P(x) = \sum_{j=0}^{n} a_j x^j$ has the norms:

$$||P||_1 = \int_0^{2\pi} \left|P\left(e^{i\theta}\right)\right| \frac{d\theta}{2\pi} \qquad |P|_1 = \sum_{j=0}^{n} |a_j|$$

$$||P||_2 = \left(\int_0^{2\pi} \left|P\left(e^{i\theta}\right)\right|^2 \frac{d\theta}{2\pi}\right)^{1/2} \qquad |P|_2 = \left(\sum_{j=1}^{n} |a_j|^2\right)^{1/2}$$

$$||P||_\infty = \max_{|z|=1} |P(z)| \qquad |P|_\infty = \max_j |a_j|$$

For the double bar norms, P is considered as a function on the unit circle; for the single bar norms, P is identified with its coefficients. These norms are comparable:

$$|P|_\infty \leq ||P||_1 \leq |P|_2 = ||P||_2 \leq ||P||_\infty \leq |P|_1 \leq n|P|_\infty. \tag{2.2.6}$$

2.2.6 OTHER POLYNOMIAL PROPERTIES

1. *Jensen's inequality*: For the polynomial $P(x) = \sum_{j=0}^{n} a_j x^j$, with $a_0 \neq 0$

$$\int_0^{2\pi} \log\left|P\left(e^{i\theta}\right)\right| \frac{d\theta}{2\pi} \geq \log|a_0| \tag{2.2.7}$$

2. *Symmetric form*: The polynomial $P(x_1, \ldots, x_n) = \sum_{|\alpha|=m} a_\alpha x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_N^{\alpha_N}$, where $\alpha = (\alpha_1, \ldots, \alpha_N)$ can be written in the symmetric form

$$P(x_1, \ldots, x_n) = \sum_{i_1, \ldots, i_m = 1}^{N} c_{i_1, \ldots, i_m} x_{i_1} x_{i_2} \ldots x_{i_m} \tag{2.2.8}$$

with $c_{i_1,\ldots,i_m} = \frac{1}{m!} \frac{\partial^m P}{\partial x_{i_1} \ldots \partial x_{i_m}}$. This means that the $x_1 x_2$ term is written as $\frac{1}{2}(x_1 x_2 + x_2 x_1)$, the term $x_1 x_2^2$ becomes $\frac{1}{3}(x_1 x_2 x_2 + x_2 x_1 x_2 + x_2 x_2 x_1)$.

3. The *Mahler measure* (a valuation) of the polynomial $P(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_0 = a_n(x - z_1)(x - z_2) \ldots (x - z_n)$ is given by

$$M(P) = |a_n| \prod_{i=1}^{n} \max(1, |z_i|) = \exp\left(\int_0^1 \log\left|P\left(e^{2\pi i t}\right)\right| dt\right)$$

This valuation satisfies the properties:

(a) $M(P)\, M(Q) = M(PQ)$

(b) $M(P(x)) = M(P(-x)) = M(P(x^k))$ for $k \geq 1$

(c) $M(P(x)) = M(x^n P(x^{-1}))$

2.2.7 CYCLOTOMIC POLYNOMIALS

The d^{th} *cyclotomic polynomial*, $\Phi_d(x)$ is

$$\Phi_d(x) = \prod_{\substack{k=1,2,\ldots \\ (k,d)=1}}^{d} (x - \xi_k) = x^{\phi(d)} + \ldots \tag{2.2.9}$$

where the $\xi_k = e^{2\pi i k/d}$ are the primitive d^{th} roots of unity

$$\begin{aligned} \Phi_p(x) &= \sum_{k=0}^{p-1} x^k = \frac{x^p - 1}{x - 1} \qquad \text{if } p \text{ is prime} \\ x^n - 1 &= \prod_{d|n} \Phi_d(x) \\ x^n + 1 &= \frac{x^{2n} - 1}{x^n - 1} = \frac{\prod_{d|2n} \Phi_d(x)}{\prod_{d|n} \Phi_d(x)} = \prod_{d|m} \Phi_{2^t d}(x) \end{aligned} \tag{2.2.10}$$

where $n = 2^{t-1}m$ and m is odd.

n	cyclotomic polynomial of degree n
1	$-1 + x$
2	$1 + x$
3	$1 + x + x^2 = \frac{x^3-1}{x-1}$
4	$1 + x^2$
5	$1 + x + x^2 + x^3 + x^4 = \frac{x^5-1}{x-1}$
6	$1 - x + x^2$
7	$1 + x + x^2 + x^3 + x^4 + x^5 + x^6 = \frac{x^7-1}{x-1}$
8	$1 + x^4$
9	$1 + x^3 + x^6$
10	$1 - x + x^2 - x^3 + x^4$
11	$1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10} = \frac{x^{11}-1}{x-1}$
12	$1 - x^2 + x^4$

Some relations:

1. If p is an odd prime:

(a) $\Phi_p(x) = \dfrac{x^p - 1}{x - 1}$

(b) $\Phi_{2p}(x) = \dfrac{x^{2p} - 1}{x^p - 1}\dfrac{x - 1}{x^2 - 1}$

(c) $\Phi_{4p}(x) = \dfrac{x^{4p} - 1}{x^{2p} - 1}\dfrac{x^2 - 1}{x^4 - 1}$

(d) $x^p - 1 = \Phi_1(x)\Phi_p(x)$

(e) $x^{2p} - 1 = \Phi_1(x)\Phi_2(x)\Phi_p(x)\Phi_{2p}(x)$

2. If p is a prime and

   (a) p does not divide n then $\Phi_{np}(x) = \Phi_n(x^p)/\Phi_n(x)$

   (b) p divides n then $\Phi_{np}(x) = \Phi_n(x^p)$

2.3 VECTOR ALGEBRA

2.3.1 NOTATION FOR VECTORS AND SCALARS

A *vector* is an ordered n-tuple of values. A vector is usually represented by a lowercase, boldfaced letter, such as **v**. The individual components of a vector **v** are typically denoted by a lowercase letter along with a subscript identifying the relative position of the component in the vector, such as $\mathbf{v} = [v_1, v_2, \ldots, v_n]$. In this case, the vector is said to be n-dimensional. If the n individual components of the vector are real numbers, then $\mathbf{v} \in \mathbb{R}^n$. Similarly, if the n components of **v** are complex, then $\mathbf{v} \in \mathbb{C}^n$.

Subscripts are often used to identify individual vectors within a set of vectors all belonging to the same type. For example, a set of n velocity vectors can be denoted by $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$. In this case, a boldface type is used on the individual members of the set to signify these elements of the set are vectors and not vector components.

Two vectors, **v** and **u**, are said to be equal if all their components are equal. The negative of a vector, written as $-\mathbf{v}$, is one that acts in a direction opposite to **v**, but is of equal magnitude.

2.3.2 PHYSICAL VECTORS

Any quantity that is completely determined by its magnitude is called a *scalar*. For example, mass, density, and temperature are scalars. Any quantity that is completely determined by its magnitude and direction is called, in physics, a vector. We often use a three-dimensional vector to represent a physical vector. Examples of physical vectors include velocity, acceleration, and force. A physical vector is represented by a directed line segment, the length of which represents the magnitude of the vector. Two vectors are said to be *parallel* if they have exactly the same direction, i.e., the angle between the two vectors equals zero.

2.3.3 FUNDAMENTAL DEFINITIONS

1. A *row vector* is a vector whose components are aligned horizontally. A *column vector* has its components aligned vertically. The *transpose* operator, denoted by the superscript $^{\mathrm{T}}$, switches the orientation of a vector between horizontal and vertical.

   EXAMPLE

   $$\underbrace{\mathbf{v} = [1, 2, 3, 4]}_{\text{row vector}}, \qquad \underbrace{\mathbf{v}^{\mathrm{T}} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}}_{\text{column vector}}, \qquad \underbrace{(\mathbf{v}^{\mathrm{T}})^{\mathrm{T}} = [1, 2, 3, 4]}_{\text{row vector}}.$$

   Vectors are traditionally written either with rounded braces or with square brackets.

2. Two vectors, $\mathbf{v}$ and $\mathbf{u}$, are said to be *orthogonal* if $\mathbf{v}^{\mathrm{T}}\mathbf{u} = 0$. (This is also written $\mathbf{v} \cdot \mathbf{u} = 0$, where the "·" denotes an inner product; see page 80.)
3. A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ is said to be *orthogonal* if $\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_j = 0$ for all $i \neq j$.
4. A set of orthogonal vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_m\}$ is said to be *orthonormal* if, in addition to possessing the property of orthogonality, the set possesses the property that $\mathbf{v}_i^{\mathrm{T}}\mathbf{v}_i = 1$ for all $1 \leq i \leq m$.

2.3.4 LAWS OF VECTOR ALGEBRA

1. The vector sum of $\mathbf{v}$ and $\mathbf{u}$, represented by $\mathbf{v} + \mathbf{u}$, results in another vector of the same dimension, and is calculated by simply adding corresponding vector components, e.g., if $\mathbf{v}, \mathbf{u} \in \mathbb{R}^n$, then $\mathbf{v} + \mathbf{u} = [v_1 + u_1, \dots, v_n + u_n]$.
2. The vector subtraction of $\mathbf{u}$ from $\mathbf{v}$, represented by $\mathbf{v} - \mathbf{u}$, is equivalent to the addition of $\mathbf{v}$ and $-\mathbf{u}$.
3. If $r > 0$ is a scalar, then the scalar multiplication $r\mathbf{v}$ (equal to $\mathbf{v}r$) represents a scaling by a factor r of the vector $\mathbf{v}$ in the same direction as $\mathbf{v}$. That is, the multiplicative scalar is distributed to each component of $\mathbf{v}$.
4. If $0 \leq r < 1$, then the scalar multiplication of r and $\mathbf{v}$ shrinks the length of $\mathbf{v}$, multiplication by $r = 1$ leaves $\mathbf{v}$ unchanged, and, if $r > 1$, then $r\mathbf{v}$ stretches the length of $\mathbf{v}$. When $r < 0$, scalar multiplication of r and $\mathbf{v}$ has the same effect on the magnitude (length) of $\mathbf{v}$ as when $r > 0$, but results in a vector oriented in the direction opposite to $\mathbf{v}$.

   EXAMPLE

$$4\begin{bmatrix}1\\0\\3\end{bmatrix} = \begin{bmatrix}4\\0\\12\end{bmatrix}, \qquad -4\begin{bmatrix}1\\0\\3\end{bmatrix} = \begin{bmatrix}-4\\0\\-12\end{bmatrix}.$$

5. If r and s are scalars, and $\mathbf{v}$, $\mathbf{u}$, and $\mathbf{w}$ are vectors, the following rules of algebra are valid:

$$\begin{aligned}
\mathbf{v} + \mathbf{u} &= \mathbf{u} + \mathbf{v}, \\
(r + s)\mathbf{v} &= r\mathbf{v} + s\mathbf{v} = \mathbf{v}r + \mathbf{v}s = \mathbf{v}(r + s), \\
r(\mathbf{v} + \mathbf{u}) &= r\mathbf{v} + r\mathbf{u}, \\
\mathbf{v} + (\mathbf{u} + \mathbf{w}) &= (\mathbf{v} + \mathbf{u}) + \mathbf{w} = \mathbf{v} + \mathbf{u} + \mathbf{w}.
\end{aligned} \tag{2.3.1}$$

2.3.5 VECTOR NORMS

1. A *norm* is the vector analog to the measure of absolute value for real scalars. Norms provide a distance measure for a vector space.
2. A vector norm applied to a vector $\mathbf{v}$ is denoted by a double bar notation $\|\mathbf{v}\|$.
3. A norm on a vector space equips it with a *metric space* structure.
4. The properties of a vector norm are:
   (a) For any vector $\mathbf{v} \neq \mathbf{0}$, $\|\mathbf{v}\| > 0$,
   (b) $\|\gamma\mathbf{v}\| = |\gamma| \; \|\mathbf{v}\|$, and
   (c) $\|\mathbf{v} + \mathbf{u}\| \leq \|\mathbf{v}\| + \|\mathbf{u}\|$ (triangle inequality).

5. The most commonly used vector norms on $\mathbb{R}^n$ or $\mathbb{C}^n$ are:
   (a) The L^1 norm is defined as $\|\mathbf{v}\|_1 = |v_1| + \cdots + |v_n| = \sum_{i=1}^n |v_i|$.
   (b) The L^2 norm (Euclidean norm) is defined as

$$\|\mathbf{v}\|_2 = (|v_1|^2 + |v_2|^2 + \cdots + |v_n|^2)^{1/2} = \left(\sum_{i=1}^n v_i^2\right)^{1/2}. \tag{2.3.2}$$

   (c) The L^∞ norm is defined as $\|\mathbf{v}\|_\infty = \max_{1\le i\le n} |v_i|$.
6. When there is no subscript, the norm $\|\cdot\|$ is usually assumed to be the L^2 norm.
7. A *unit vector* with respect to a particular norm $\|\cdot\|$ is a vector that satisfies the property that $\|\mathbf{v}\| = 1$, and is sometimes denoted by $\hat{\mathbf{v}}$.

EXAMPLE The vector $\mathbf{x} = \begin{bmatrix}1 & 2 & 3\end{bmatrix}$ in different norms:
$||\mathbf{x}||_1 = 6$ $\quad ||\mathbf{x}|| = ||\mathbf{x}||_2 = \sqrt{14} \approx 3.74$ $\quad ||\mathbf{x}||_3 = 6^{2/3} \approx 3.30$ $\quad ||\mathbf{x}||_\infty = 3$

2.3.6 DOT, SCALAR, OR INNER PRODUCT

1. The *dot* (or *scalar* or *inner*) *product* of two vectors of the same dimension, represented by $\mathbf{v} \cdot \mathbf{u}$ or $\mathbf{v}^{\mathrm{T}}\mathbf{u}$, has two common definitions, depending upon the context in which this product is encountered.
   (a) The dot or scalar product is defined by $\quad \mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \, \|\mathbf{u}\| \cos\theta$, where θ represents the angle between the vectors $\mathbf{v}$ and $\mathbf{u}$.
   (b) The inner product of two vectors, $\mathbf{u}$ and $\mathbf{v}$, is equivalently defined as

$$\mathbf{u}^{\mathrm{T}}\mathbf{v} = \sum_{i=1}^n u_i v_i = u_1 v_1 + \cdots + u_n v_n. \tag{2.3.3}$$

   From the first definition, it is apparent that the inner product of two perpendicular, or orthogonal, vectors is zero, since the cosine of $90°$ is zero.
2. The inner product of two parallel vectors (with $\mathbf{u} = r\mathbf{v}$) is given by $\mathbf{v} \cdot \mathbf{u} = r\,\|\mathbf{v}\|^2$. For example, when $r > 0$,

$$\mathbf{v} \cdot \mathbf{u} = \|\mathbf{v}\| \, \|\mathbf{u}\| \cos 0 = \|\mathbf{v}\| \, \|\mathbf{u}\| = \|\mathbf{v}\| \, \|r\mathbf{v}\| = r\,\|\mathbf{v}\|^2. \tag{2.3.4}$$

3. The dot product is distributive, e.g.,

$$(\mathbf{v} + \mathbf{u}) \cdot \mathbf{w} = \mathbf{v} \cdot \mathbf{w} + \mathbf{u} \cdot \mathbf{w}. \tag{2.3.5}$$

4. For $\mathbf{v}, \mathbf{u}, \mathbf{w} \in \mathbb{R}^n$ with $n > 1$,

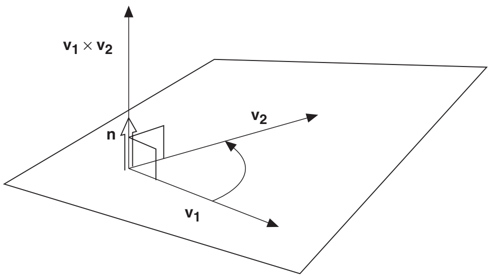
$$\mathbf{v}^{\mathrm{T}}\mathbf{u} = \mathbf{v}^{\mathrm{T}}\mathbf{w} \qquad \not\Rightarrow \qquad \mathbf{u} = \mathbf{w}. \tag{2.3.6}$$

   However, it *is* valid to conclude that

$$\mathbf{v}^{\mathrm{T}}\mathbf{u} = \mathbf{v}^{\mathrm{T}}\mathbf{w} \qquad \Rightarrow \qquad \mathbf{v}^{\mathrm{T}}(\mathbf{u} - \mathbf{w}) = 0, \tag{2.3.7}$$

   i.e., the vector $\mathbf{v}$ is orthogonal to the vector $(\mathbf{u} - \mathbf{w})$.

FIGURE 2.1
Depiction of right-hand rule.



2.3.7 VECTOR OR CROSS-PRODUCT

1. The *vector* (or *cross-*) *product*) of two non-zero 3-dimensional vectors $\mathbf{v}$ and $\mathbf{u}$ is defined as
$$\mathbf{v} \times \mathbf{u} = \hat{\mathbf{n}}\, \|\mathbf{v}\|\, \|\mathbf{u}\| \sin\theta, \tag{2.3.8}$$
where $\hat{\mathbf{n}}$ is the unit *normal* vector (i.e., vector perpendicular to both $\mathbf{v}$ and $\mathbf{u}$) in the direction adhering to the *right-hand rule* (see Figure 2.1) and θ is the angle between $\mathbf{v}$ and $\mathbf{u}$.
2. If $\mathbf{v}$ and $\mathbf{u}$ are parallel, then $\mathbf{v} \times \mathbf{u} = \mathbf{0}$.
3. The quantity $\|\mathbf{v}\|\, \|\mathbf{u}\|\, |\sin\theta|$ represents the area of the parallelogram determined by $\mathbf{v}$ and $\mathbf{u}$.
4. The following rules apply for vector products:
$$\begin{aligned}
(\gamma\mathbf{v}) \times (\alpha\mathbf{u}) &= (\gamma\alpha)\mathbf{v} \times \mathbf{u},\\
\mathbf{v} \times \mathbf{u} &= -\mathbf{u} \times \mathbf{v},\\
\mathbf{v} \times (\mathbf{u} + \mathbf{w}) &= \mathbf{v} \times \mathbf{u} + \mathbf{v} \times \mathbf{w},\\
(\mathbf{v} + \mathbf{u}) \times \mathbf{w} &= \mathbf{v} \times \mathbf{w} + \mathbf{u} \times \mathbf{w},\\
\mathbf{v} \times (\mathbf{u} \times \mathbf{w}) &= \mathbf{u}(\mathbf{w} \cdot \mathbf{v}) - \mathbf{w}(\mathbf{v} \cdot \mathbf{u}),\\
(\mathbf{v} \times \mathbf{u}) \cdot (\mathbf{w} \times \mathbf{z}) &= (\mathbf{v} \cdot \mathbf{w})(\mathbf{u} \cdot \mathbf{z}) - (\mathbf{v} \cdot \mathbf{z})(\mathbf{u} \cdot \mathbf{w}),\\
(\mathbf{v} \times \mathbf{u}) \times (\mathbf{w} \times \mathbf{z}) &= [\mathbf{v} \cdot (\mathbf{u} \times \mathbf{z})]\mathbf{w} - [\mathbf{v} \cdot (\mathbf{u} \times \mathbf{w})]\mathbf{z}\\
&= [\mathbf{v} \cdot (\mathbf{w} \times \mathbf{z})]\mathbf{u} - [\mathbf{u} \cdot (\mathbf{w} \times \mathbf{z})]\mathbf{v}.
\end{aligned} \tag{2.3.9}$$
5. The pairwise cross-products of the unit vectors $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, and $\hat{\mathbf{k}}$, corresponding to the directions of $\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}} + v_3\hat{\mathbf{k}}$, are given by
$$\begin{aligned}
\hat{\mathbf{i}} \times \hat{\mathbf{j}} &= -(\hat{\mathbf{j}} \times \hat{\mathbf{i}}) = \hat{\mathbf{k}},\\
\hat{\mathbf{j}} \times \hat{\mathbf{k}} &= -(\hat{\mathbf{k}} \times \hat{\mathbf{j}}) = \hat{\mathbf{i}},\\
\hat{\mathbf{k}} \times \hat{\mathbf{i}} &= -(\hat{\mathbf{i}} \times \hat{\mathbf{k}}) = \hat{\mathbf{j}}, \qquad \text{and}\\
\hat{\mathbf{i}} \times \hat{\mathbf{i}} &= \hat{\mathbf{j}} \times \hat{\mathbf{j}} = \hat{\mathbf{k}} \times \hat{\mathbf{k}} = \mathbf{0}.
\end{aligned} \tag{2.3.10}$$

6. If $\mathbf{v} = v_1\hat{\mathbf{i}} + v_2\hat{\mathbf{j}} + v_3\hat{\mathbf{k}}$ and $\mathbf{u} = u_1\hat{\mathbf{i}} + u_2\hat{\mathbf{j}} + u_3\hat{\mathbf{k}}$, then

$$\mathbf{v}\times\mathbf{u} = \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \end{vmatrix} = (v_2u_3 - u_2v_3)\hat{\mathbf{i}} + (v_3u_1 - u_3v_1)\hat{\mathbf{j}} + (v_1u_2 - u_1v_2)\hat{\mathbf{k}}. \tag{2.3.11}$$

2.3.8 SCALAR AND VECTOR TRIPLE PRODUCTS

1. The *scalar triple product* involving three 3-dimensional vectors **v**, **u**, and **w**, sometimes denoted by [**vuw**] (not to be confused with a matrix containing three columns $\begin{bmatrix}\mathbf{v} & \mathbf{u} & \mathbf{w}\end{bmatrix}$), is defined to be

$$\begin{aligned} [\mathbf{vuw}] = \mathbf{v}\cdot(\mathbf{u}\times\mathbf{w}) &= \mathbf{v}\cdot\left[\begin{vmatrix} u_2 & u_3 \\ w_2 & w_3 \end{vmatrix}\hat{\mathbf{i}} - \begin{vmatrix} u_1 & u_3 \\ w_1 & w_3 \end{vmatrix}\hat{\mathbf{j}} + \begin{vmatrix} u_1 & u_2 \\ w_1 & w_2 \end{vmatrix}\hat{\mathbf{k}}\right] \\ &= v_1\begin{vmatrix} u_2 & u_3 \\ w_2 & w_3 \end{vmatrix} - v_2\begin{vmatrix} u_1 & u_3 \\ w_1 & w_3 \end{vmatrix} + v_3\begin{vmatrix} u_1 & u_2 \\ w_1 & w_2 \end{vmatrix} \\ &= \begin{vmatrix} v_1 & v_2 & v_3 \\ u_1 & u_2 & u_3 \\ w_1 & w_2 & w_3 \end{vmatrix} \qquad (2.3.12) \\ &= \|v\|\ \|u\|\ \|w\| \cos\phi\ \sin\theta, \end{aligned}$$

where θ is the angle between **u** and **w**, and ϕ is the angle between **v** and the normal to the plane defined by **u** and **w**.
2. The absolute value of a triple scalar product calculates the volume of the parallelepiped determined by the three vectors. The result is independent of the order in which the triple product is taken.
3. $(\mathbf{v}\times\mathbf{u})\times(\mathbf{w}\times\mathbf{z}) = [\mathbf{vwz}]\mathbf{u} - [\mathbf{uwz}]\mathbf{v} = [\mathbf{vuz}]\mathbf{w} - [\mathbf{vuw}]\mathbf{z}$
4. Given three non-coplanar reference vectors **u**, **v**, and **w**, the *reciprocal system* is given by $\mathbf{u}^*$, $\mathbf{v}^*$, and $\mathbf{w}^*$, where

$$\mathbf{u}^* = \frac{\mathbf{w}\times\mathbf{v}}{[\mathbf{vuw}]}, \qquad \mathbf{v}^* = \frac{\mathbf{u}\times\mathbf{w}}{[\mathbf{vuw}]}, \qquad \mathbf{w}^* = \frac{\mathbf{v}\times\mathbf{u}}{[\mathbf{vuw}]}. \tag{2.3.13}$$

If the vectors **v**, **u**, and **w** are mutually perpendicular, then

$$\begin{aligned} 1 &= \mathbf{v}\cdot\mathbf{v}^* = \mathbf{u}\cdot\mathbf{u}^* = \mathbf{w}\cdot\mathbf{w}^* \\ \text{and}\ \ 0 &= \mathbf{v}\cdot\mathbf{u}^* = \mathbf{v}\cdot\mathbf{w}^* = \mathbf{u}\cdot\mathbf{v}^*, \ \ \text{etc.} \end{aligned} \tag{2.3.14}$$

The system $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ is its own reciprocal.
5. The *vector triple product* involving three 3-dimensional vectors **v**, **u**, and **w**, given by $\mathbf{v}\times(\mathbf{u}\times\mathbf{w})$, results in a vector, perpendicular to **v**, lying in the plane of **u** and **w**, and is defined as

$$\begin{aligned} \mathbf{v}\times(\mathbf{u}\times\mathbf{w}) &= (\mathbf{v}\cdot\mathbf{w})\mathbf{u} - (\mathbf{v}\cdot\mathbf{u})\mathbf{w}, \\ &= \begin{vmatrix} \hat{\mathbf{i}} & \hat{\mathbf{j}} & \hat{\mathbf{k}} \\ v_1 & v_2 & v_3 \\ \begin{vmatrix} u_2 & u_3 \\ w_2 & w_3 \end{vmatrix} & \begin{vmatrix} u_3 & u_1 \\ w_3 & w_1 \end{vmatrix} & \begin{vmatrix} u_1 & u_2 \\ w_1 & w_2 \end{vmatrix} \end{vmatrix}. \end{aligned} \tag{2.3.15}$$

2.4 LINEAR AND MATRIX ALGEBRA

2.4.1 DEFINITIONS

1. An $m \times n$ *matrix* is a 2-dimensional array of numbers consisting of *m* rows and *n* columns. By convention, a matrix is denoted by a capital letter emphasized with italics, as in A, B, D, or boldface, **A**, **B**, **D**. Sometimes a matrix has a subscript denoting the dimensions of the matrix, e.g., $A_{2\times 3}$. If A is a real $n \times m$ matrix, then we write $A \in \mathbb{R}^{n\times m}$. Higher-dimensional matrices, although less frequently encountered, are accommodated in a similar fashion, e.g., a 3-dimensional matrix $A_{m\times n\times p}$, and so on.
2. $A_{m\times n}$ is called *rectangular* if $m \neq n$.
3. $A_{m\times n}$ is called *square* if $m = n$.
4. A particular component (or element) of a matrix is denoted by the lowercase letter corresponding to the matrix name, along with two subscripts for the row i and column j location of the component in the matrix, e.g.,

$$\begin{aligned} A_{m\times n} &\quad \text{has components } a_{ij}; \\ B_{m\times n} &\quad \text{has components } b_{ij}. \end{aligned}$$

   For example, a_{23} is the component in the second row and third column of matrix A.
5. Any component a_{ij} with $i = j$ is called a *diagonal* component.
6. The diagonal alignment of components in a matrix extending from the upper left to the lower right is called the *principal* or *main* diagonal.
7. Any component a_{ij} with $i \neq j$ is called an *off-diagonal* component.
8. Two matrices A and B are said to be equal if they have the same number of rows (m) and columns (n), and $a_{ij} = b_{ij}$ for all $1 \leq i \leq m$, $1 \leq j \leq n$.
9. An $m \times 1$ dimensional matrix is called a *column vector*. Similarly, a $1 \times n$ dimensional matrix is called a *row vector*.
10. A vector with all components equal to zero is called a *null* vector and is usually denoted by **0**.
11. A *zero*, or *null matrix* is one whose elements are all zero (notation is "0").
12. A column vector with all components equal to one is often denoted by **e**. The analogous row vector is denoted by $\mathbf{e}^{\mathrm{T}}$.
13. The standard basis consists of the vectors $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ where $\mathbf{e}_i$ is an $n \times 1$ vector of all zeros, except for the i^{th} component, which is one.
14. The scalar $\mathbf{x}^{\mathrm{T}}\mathbf{x} = \sum_{i=1}^{n} x_i^2$ is the sum of squares of all components of the vector **x**.
15. The *weighted* sum of squares is defined by $\mathbf{x}^{\mathrm{T}} D_w \mathbf{x} = \sum_{i=1}^{n} w_i x_i^2$, when **x** has n components and the diagonal matrix D_w is of dimension $(n \times n)$.
16. If Q is a square matrix, then $\mathbf{x}^{\mathrm{T}} Q \mathbf{x}$ is called a *quadratic* form (see page 101).
17. An $n \times n$ matrix A is called *non-singular*, or *invertible*, or *regular*, if there exists an $n \times n$ matrix B such that $AB = BA = I$. The unique matrix B satisfying this condition is called the *inverse* of A, and is denoted by A^{-1}.

18. The scalar $\mathbf{x}^{\mathrm{T}}\mathbf{y} = \sum_{i=1}^{n} x_i y_i$, the inner product of $\mathbf{x}$ and $\mathbf{y}$, is the sum of products of the components of $\mathbf{x}$ by those of $\mathbf{y}$.
19. The *weighted* sum of products is $\mathbf{x}^{\mathrm{T}} D_w \mathbf{y} = \sum_{i=1}^{n} w_i x_i y_i$, when $\mathbf{x}$ and $\mathbf{y}$ have n components, and the diagonal matrix D_w is $(n \times n)$.
20. The map $\mathbf{x}, \mathbf{y} \mapsto \mathbf{x}^{\mathrm{T}} Q \mathbf{y}$ is called a *bilinear* form, where Q is a matrix of appropriate dimension.
21. The *transpose* of an $m \times n$ matrix A, denoted by A^{T}, is an $n \times m$ matrix with rows and columns interchanged, so that the (i, j) component of A is the (j, i) component of A^{T}, and $(A^{\mathrm{T}})_{ji} = (A)_{ij} = a_{ij}$.
22. The *Hermitian conjugate* of a matrix A, denoted by A^{H}, is obtained by transposing A and replacing each element by its complex conjugate. Hence, if $a_{kl} = u_{kl} + iv_{kl}$, then $(A^{\mathrm{H}})_{kl} = u_{lk} - iv_{ik}$, with $i = \sqrt{-1}$.
23. If Q is a square matrix, then the map $\mathbf{x} \mapsto \mathbf{x}^{\mathrm{H}} Q \mathbf{x}$ is called a *Hermitian* form.

2.4.2 TYPES OF MATRICES

1. A square matrix with all components not on the principal diagonal equal to zero is called a *diagonal matrix*, typically denoted by the letter D with a subscript indicating the typical element in the principal diagonal. For example:

$$D_a = \begin{bmatrix} a_{11} & 0 & 0 \\ 0 & a_{22} & 0 \\ 0 & 0 & a_{33} \end{bmatrix}, \qquad D_\lambda = \begin{bmatrix} \lambda_{11} & 0 & 0 \\ 0 & \lambda_{22} & 0 \\ 0 & 0 & \lambda_{33} \end{bmatrix}.$$

2. A matrix whose components are arranged in m rows and a single column is called a *column matrix*, or *column vector*, and is typically denoted using boldface, lowercase letters, e.g., $\mathbf{a}$ and $\mathbf{b}$.
3. A matrix whose components are arranged in n columns and a single row is called a *row matrix*, or *row vector*, and is typically denoted as a transposed column vector, e.g., $\mathbf{a}^{\mathrm{T}}$ and $\mathbf{b}^{\mathrm{T}}$.
4. The *identity matrix*, denoted by I, is the diagonal matrix with $a_{ij} = 1$ for all $i = j$, and $a_{ij} = 0$ for $i \neq j$. The $n \times n$ identity matrix is denoted I_n.
5. The *elementary matrix*, E_{ij}, is defined differently in different contexts:
   (a) Elementary matrices have the form $E_{ij} = \mathbf{e}_i \mathbf{e}_j^{\mathrm{T}}$. Hence, the matrix $A = (a_{ij})$ can be written as $A = \sum_i \sum_j a_{ij} E_{ij}$.
   (b) In Gaussian elimination, the matrix that subtracts a multiple l of row j from row i is called E_{ij}, with 1's on the diagonal and the number $-l$ in row i column j. For example:

$$E_{31} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -5 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

   (c) Elementary matrices are also written as $E = I - \alpha u v^{\mathrm{T}}$, where I is the identity matrix, α is a scalar, and u and v are vectors of the same dimension. In this context, the elementary matrix is referred to as a *rank one modification* of an identity matrix.

6. A matrix with all components above the principal diagonal equal to zero is called a *lower triangular matrix*. For example:

$$A = \begin{bmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \qquad \text{is lower triangular.}$$

7. A matrix with all components below the principal diagonal equal to zero is called an *upper triangular matrix*. (The transpose of a lower triangular matrix is an upper triangular matrix.)
8. A matrix $A = (a_{ij})$ has *lower bandwidth* p if $a_{ij} = 0$ whenever $i > j + p$ and *upper bandwidth* q if $a_{ij} = 0$ whenever $j > i + q$. When they are equal, they are the *bandwidth* of A. A diagonal matrix has bandwidth 0. A *tridiagonal matrix* has bandwidth 1. An upper (resp. lower) triangular matrix has upper (resp. lower) bandwidth of $n - 1$ (resp. $m - 1$).
9. An $m \times n$ matrix A with orthonormal columns has the property $A^{\mathrm{T}}A = I$.
10. A square matrix is called *symmetric* if $A = A^{\mathrm{T}}$.
11. A square matrix is called *skew symmetric* if $A^{\mathrm{T}} = -A$.
12. A square matrix is called *idempotent* if $AA = A^2 = A$.
13. A square matrix is called *Hermitian* if $A = A^{\mathrm{H}}$. A square matrix A is called *skew-Hermitian* if $A^{\mathrm{H}} = -A$. All real symmetric matrices are Hermitian.
14. A square matrix is called *unitary* if $A^{\mathrm{H}}A = I$. A real unitary matrix is orthogonal. All eigenvalues of a unitary matrix have an absolute value of one.
15. A square matrix is called a *permutation matrix* if its columns are a permutation of the columns of I. A permutation matrix is orthogonal.
16. A square matrix is called a *projection matrix* if it is both Hermitian and idempotent: $A^{\mathrm{H}} = A^2 = A$.
17. A square matrix is called *normal* if $A^{\mathrm{H}}A = AA^{\mathrm{H}}$. The following matrices are normal: diagonal, Hermitian, unitary, skew-Hermitian. Pseudospectra are useful for understanding non-normal operators, see page 105.
18. A square matrix is called *nilpotent* to index k if $A^k = 0$ but $A^{k-1} \neq 0$. The eigenvalues of a nilpotent matrix are all zero.
19. A square matrix, whose elements are constant along each diagonal, is called a *Toeplitz* matrix. Toeplitz matrices are symmetric about a diagonal extending from the upper right-hand corner element to the lower left-hand corner element. This type of symmetry is called *persymmetry*. Example Toeplitz matrices:

$$A = \begin{bmatrix} a & d & e \\ b & a & d \\ c & b & a \end{bmatrix} \qquad \text{and} \qquad M = \begin{bmatrix} 4 & 0 & 1 \\ -11 & 4 & 0 \\ 3 & -11 & 4 \end{bmatrix}$$

20. A square matrix Q with orthonormal columns is said to be *orthogonal*. It follows directly that the rows of Q must also be orthonormal, so that $QQ^{\mathrm{T}} = Q^{\mathrm{T}}Q = I$, or $Q^{\mathrm{T}} = Q^{-1}$. The determinant of an orthogonal matrix is ± 1. A *rotation matrix* is an orthogonal matrix whose determinant is equal to $+1$.

21. A *Vandermonde* matrix is a square matrix $V \in \mathbb{R}^{(n+1)\times(n+1)}$ in which each column contains unit increasing powers of a single matrix value:

$$V = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ v_1 & v_2 & \cdots & v_{(n+1)} \\ v_1^2 & v_2^2 & \cdots & v_{(n+1)}^2 \\ \vdots & \vdots & & \vdots \\ v_1^n & v_2^n & \cdots & v_{(n+1)}^n \end{bmatrix}. \tag{2.4.1}$$

22. A square matrix U is said to be in *upper Hessenberg* form if $u_{ij} = 0$ whenever $i > j+1$. An upper Hessenberg matrix is essentially an upper triangular matrix with an extra non-zero element immediately below the main diagonal entry in each column of U. For example,

$$U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ b_{21} & u_{22} & u_{23} & u_{24} \\ 0 & b_{32} & u_{33} & u_{34} \\ 0 & 0 & b_{43} & u_{44} \end{bmatrix} \qquad \text{is upper Hessenberg.}$$

23. A *circulant* matrix is an $n \times n$ matrix of the form

$$C = \begin{bmatrix} c_0 & c_1 & c_2 & \cdots & c_{n-2} & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \cdots & c_{n-3} & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \cdots & c_{n-4} & c_{n-3} \\ \vdots & & & & & \vdots \\ c_1 & c_2 & c_3 & \cdots & c_{n-1} & c_0 \end{bmatrix}, \tag{2.4.2}$$

where the components c_{ij} are such that $(j - i) = k \mod n$ have the same value c_k. These components comprise the k^{th} *stripe* of C.

24. If a rotation is defined by the matrix $A = \left[\begin{smallmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{smallmatrix}\right]$ then its fixed axis of rotation is given by $\mathbf{v} = \mathbf{i}(a_{23} - a_{32}) + \mathbf{j}(a_{31} - a_{13}) + \mathbf{k}(a_{12} - a_{21})$.

25. A *Givens rotation* is defined as rank two correction to the identity matrix given by

$$G(i,k,\theta) = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c & \cdots & s & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s & \cdots & c & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \begin{matrix} \\ \\ i \\ \\ k \\ \\ \\ \end{matrix} \tag{2.4.3}$$

$$\qquad\qquad\qquad i \qquad\quad k$$

where $c = \cos\theta$ and $s = \sin\theta$ for some angle θ. Pre-multiplication by $G(i,k,\theta)^{\mathrm{T}}$ induces a counterclockwise rotation of θ radians in the (i,k) plane. For $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} = G(i,k,\theta)^{\mathrm{T}}\mathbf{x}$, the components of $\mathbf{y}$ are given by

$$y_j = \begin{cases} cx_i - sx_k, & \text{for } j = i \\ sx_i + cx_k, & \text{for } j = k \\ x_j, & \text{for } j \neq i, k. \end{cases} \tag{2.4.4}$$

26. If the sum of the components of each column of a matrix $A \in \mathbb{R}^{n \times n}$ equals one, then A is called a *Markov* matrix.
27. A *Householder transformation*, or *Householder reflection*, is an $n \times n$ matrix H of the form $H = I - (2\mathbf{u}\mathbf{u}^{\mathrm{T}})/(\mathbf{u}^{\mathrm{T}}\mathbf{u})$, where the *Householder vector* $\mathbf{u} \in \mathbb{R}^n$ is non-zero.
28. A *principal sub-matrix* of a symmetric matrix A is formed by deleting rows and columns of A simultaneously, e.g., row 1 and column 1; row 9 and column 9, etc.

2.4.3 HADAMARD MATRICES

A *Hadamard matrix* of order n is an $n \times n$ matrix H with entries ± 1 such that $HH^{\mathrm{T}} = nI_n$. In order for a Hadamard matrix to exist, n must be 1, 2, or a multiple of 4. It is conjectured that this condition is also sufficient. If H_1 and H_2 are Hadamard matrices, then so is the Kronecker product $H_1 \otimes H_2$.

Without loss of generality, a Hadamard matrix can be assumed to have a first row and column consisting of all $+1$s.

For the constructs below, we use either "$-$" to denote -1 or ■ to denote $+1$.

$n = 2$

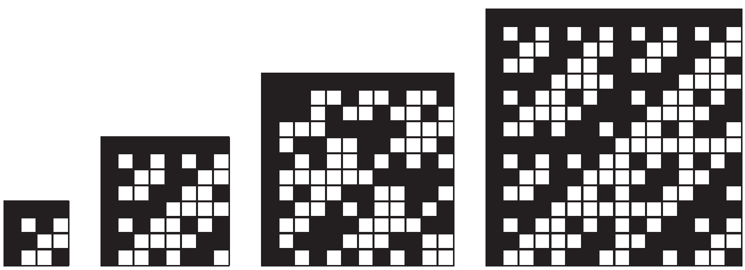
$$\begin{bmatrix} 1 & 1 \\ 1 & - \end{bmatrix}$$

$n = 4$

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & - & 1 & - \\ 1 & 1 & - & - \\ 1 & - & - & 1 \end{bmatrix}$$

$n = 8$

$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & - & 1 & - & 1 & - & 1 & - \\ 1 & 1 & - & - & 1 & 1 & - & - \\ 1 & - & - & 1 & 1 & - & - & 1 \\ 1 & 1 & 1 & 1 & - & - & - & - \\ 1 & - & 1 & - & - & 1 & - & 1 \\ 1 & 1 & - & - & - & - & 1 & 1 \\ 1 & - & - & 1 & - & 1 & 1 & - \end{bmatrix}$$



$n = 4$ $n = 8$ $n = 12$ $n = 16$

2.4.4 MATRIX ADDITION AND MULTIPLICATION

1. Two matrices A and B can be added (subtracted) if they are of the same dimension. The result is a matrix of the same dimension.

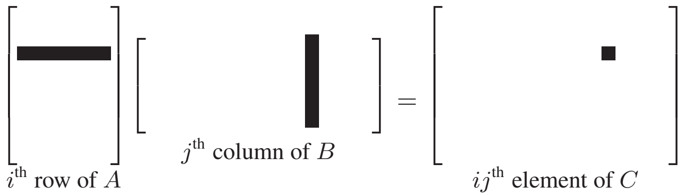
   EXAMPLE

   $$A_{2\times 3} + B_{2\times 3} = \begin{bmatrix} 3 & 2 & -1 \\ 4 & 0 & 9 \end{bmatrix} + \begin{bmatrix} 11 & -2 & 3 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 14 & 0 & 2 \\ 4 & 1 & 10 \end{bmatrix}.$$

2. Multiplication of a matrix or a vector by a scalar is achieved by multiplying each component by that scalar. If $B = \alpha A$, then $b_{ij} = \alpha a_{ij}$.
3. The matrix multiplication AB is only defined if the number of columns of A is equal to the number of rows of B.
4. The multiplication of two matrices $A_{m\times n}$ and $B_{n\times q}$ results in a matrix $C_{m\times q}$ whose components are defined as

   $$c_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} \tag{2.4.5}$$

   for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, q$. Each c_{ij} is the result of the inner (dot) product of the i^{th} row of A with the j^{th} column of B.

   This rule applies similarly for matrix multiplication involving more than two matrices. If $ABCD = E$ then

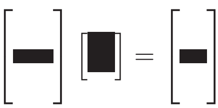
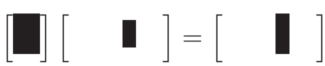
   $$e_{ij} = \sum_k \sum_l \sum_m a_{ik} b_{kl} c_{lm} d_{mj}. \tag{2.4.6}$$

   The second subscript for each matrix component must coincide with the first subscript of the next one.

   EXAMPLE

   $$\begin{bmatrix} 2 & -1 & 3 \\ -4 & 1 & 4 \end{bmatrix} \begin{bmatrix} 5 & -3 & -3 \\ 2 & 2 & -1 \\ -7 & 1 & 5 \end{bmatrix} = \begin{bmatrix} -13 & -5 & 10 \\ -46 & 18 & 31 \end{bmatrix}.$$

5. Multiplication of rows times matrices and matrices times columns can be illustrated as follows:

   
   and
   

6. In general, matrix multiplication is not commutative: $AB \neq BA$.
7. Matrix multiplication is associative: $A(BC) = (AB)C$.
8. The distributive law of multiplication and addition holds: $C(A+B) = CA + CB$ and $(A+B)C = AC + BC$.
9. Both the transpose operator and the Hermitian operator reverse the order of matrix multiplication: $(ABC)^{\mathrm{T}} = C^{\mathrm{T}}B^{\mathrm{T}}A^{\mathrm{T}}$ and $(ABC)^{\mathrm{H}} = C^{\mathrm{H}}B^{\mathrm{H}}A^{\mathrm{H}}$.
10. *Strassen algorithm*: The matrix product

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$$

can be computed in the following way:

$$\begin{aligned} m_1 &= (a_{12} - a_{22})(b_{21} + b_{22}), & c_{11} &= m_1 + m_2 - m_4 + m_6, \\ m_2 &= (a_{11} + a_{22})(b_{11} + b_{22}), & c_{12} &= m_4 + m_5, \\ m_3 &= (a_{11} - a_{21})(b_{11} + b_{12}), & c_{21} &= m_6 + m_7, \text{ and} \\ m_4 &= (a_{11} + a_{12})b_{22}, & c_{22} &= m_2 - m_3 + m_5 - m_7. \\ m_5 &= a_{11}(b_{12} - b_{22}), \\ m_6 &= a_{22}(b_{21} - b_{11}), \\ m_7 &= (a_{21} + a_{22})b_{11}, \end{aligned}$$

This computation uses 7 multiplications and 18 additions and subtractions. Using this formula recursively allows multiplication of two $n \times n$ matrices using $O(n^{\log_2 7}) = O(n^{2.807\ldots})$ scalar multiplications. Improved algorithms can achieve $O(n^{2.376\ldots})$.
11. The order in which matrices are grouped for multiplication changes the number of scalar multiplications required. The number of scalar multiplications needed to multiply matrix $X_{a\times b}$ by matrix $Y_{b\times c}$ is abc, without using clever algorithms such as Strassen's. Consider the product $P = A_{10\times 100}B_{100\times 5}C_{5\times 50}$. The grouping $P = ((AB)C)$ requires $(10\times 100\times 5) + (10\times 5\times 50) = 7,500$ scalar multiplications. The grouping $P = (A(BC))$ requires $(10\times 100\times 50) + (100\times 5\times 50) = 75{,}000$ scalar multiplications.
12. Pre-multiplication by a diagonal matrix scales the rows

$$\begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & & 0 \\ \vdots & & & \\ 0 & 0 & & d_{nn} \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} = \begin{bmatrix} d_{11}a_{11} & \cdots & d_{11}a_{1m} \\ d_{22}a_{21} & \cdots & d_{22}a_{2m} \\ \vdots & \ddots & \vdots \\ d_{nn}a_{n1} & \cdots & d_{nn}a_{nm} \end{bmatrix}$$

13. Post-multiplication by a diagonal matrix scales the columns

$$\begin{bmatrix} a_{11} & \cdots & a_{1m} \\ a_{21} & \cdots & a_{2m} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} \begin{bmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & & 0 \\ \vdots & & & \\ 0 & 0 & & d_{mm} \end{bmatrix} = \begin{bmatrix} d_{11}a_{11} & \cdots & d_{mm}a_{1m} \\ d_{11}a_{21} & \cdots & d_{mm}a_{2m} \\ \vdots & \ddots & \vdots \\ d_{11}a_{n1} & \cdots & d_{mm}a_{nm} \end{bmatrix}$$

2.4.5 DETERMINANTS

1. The *determinant* of a square matrix A, denoted by $|A|$ or $\det(A)$ or $\det A$, is a scalar function of A defined as

$$\det(A) = \det \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \tag{2.4.7}$$
$$= \sum_{\sigma} \operatorname{sgn}(\sigma) a_{1,\sigma(1)}\ a_{2,\sigma(2)}\ \cdots a_{n,\sigma(n)}$$
$$= \sum (-1)^{\delta} a_{1i_1} a_{2i_2} \cdots a_{ni_n}$$

where the sum is taken either

(a) over all permutations σ of $\{1, 2, \ldots, n\}$ and the signum function, $\operatorname{sgn}(\sigma)$, is (-1) raised to the power of the number of successive transpositions required to change the permutation σ to the identity permutation; or

(b) over all permutations $i_1 \neq i_2 \neq \cdots \neq i_n$, and δ denotes the number of transpositions necessary to bring the sequence $(i_1, i_2, \ldots, i_n)$ back into the natural order $(1, 2, \ldots, n)$.

2. For a 2×2 matrix, $\left|\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right| = a_{11}a_{22} - a_{12}a_{21}$.
For a 3×3 matrix,

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \tag{2.4.8}$$
$$- a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33}.$$

3. The determinant of the identity matrix is one.
4. Note that $|A|\,|B| = |AB|$ and $|A| = |A^{\mathrm{T}}|$.
5. Interchanging two rows (or columns) of a matrix changes the sign of its determinant.
6. A determinant does not change its value if a linear combination of other rows (or columns) is added to or subtracted from any given row (or column).
7. Multiplying an entire row (or column) of A by a scalar γ causes the determinant to be multiplied by the same scalar γ.
8. For an $n \times n$ matrix A, $|\gamma A| = \gamma^n |A|$.
9. If $\det(A) = 0$, then A is *singular*; if $\det(A) \neq 0$, then A is *non-singular* or *invertible*.
10. $\det(A^{-1}) = 1/\det(A)$.
11. When the edges of a parallelepiped P are defined by the rows (or columns) of A, the absolute value of the determinant of A measures the volume of P. Thus, if any row (or column) of A is dependent upon another row (or column) of A, the determinant of A equals zero.

12. The *cofactor* of a square matrix A, $\operatorname{cof}_{ij}(A)$, is the determinant of a sub-matrix obtained by striking the i^{th} row and the j^{th} column of A and choosing a positive (negative) sign if $(i+j)$ is even (odd).

    EXAMPLE

$$\operatorname{cof}_{23}\begin{bmatrix} 2 & 4 & 3 \\ 6 & 1 & 5 \\ -2 & 1 & 3 \end{bmatrix} = (-1)^{2+3}\begin{vmatrix} 2 & 4 \\ -2 & 1 \end{vmatrix} = -(2+8) = -10. \tag{2.4.9}$$

13. Let a_{ij} denote the components of A and a^{ij} those of A^{-1}. Then,

$$a^{ij} = \operatorname{cof}_{ji}(A)/|A|. \tag{2.4.10}$$

14. *Partitioning of determinants*: Let $A = \left[\begin{smallmatrix} B & C \\ D & E \end{smallmatrix}\right]$. Assuming all inverses exist, then

$$|A| = |E|\;\left|(B - CE^{-1}D)\right| = |B|\;\left|(E - DB^{-1}C)\right|. \tag{2.4.11}$$

15. *Laplace development*: The determinant of A is a combination of row i (column j) and the cofactors of row i (column j), i.e.,

$$\begin{aligned} |A| &= a_{i1}\operatorname{cof}_{i1}(A) + a_{i2}\operatorname{cof}_{i2}(A) + \cdots + a_{in}\operatorname{cof}_{in}(A), \\ &= a_{1j}\operatorname{cof}_{1j}(A) + a_{2j}\operatorname{cof}_{2j}(A) + \cdots + a_{nj}\operatorname{cof}_{nj}(A), \end{aligned} \tag{2.4.12}$$

    for any row i or any column j.

16. Omitting the signum function in Equation (2.4.7) yields the definition of *permanent* of A, given by $\operatorname{per} A = \sum_\sigma a_{1,\sigma(1)} \cdots a_{n,\sigma(n)}$. Properties of the permanent include:

    (a) If A is an $m \times n$ matrix and B is an $n \times m$ matrix, then

$$|\operatorname{per}(AB)|^2 \le \operatorname{per}(AA^{\mathrm{H}})\operatorname{per}(B^{\mathrm{H}}B). \tag{2.4.13}$$

    (b) If P and Q are permutation matrices, then $\operatorname{per} PAQ = \operatorname{per} A$.
    (c) If D and G are diagonal matrices, then $\operatorname{per} DAG = \operatorname{per} D \operatorname{per} A \operatorname{per} G$.
    (d) Computation of the permanent is #P-complete.

2.4.6 TRACES

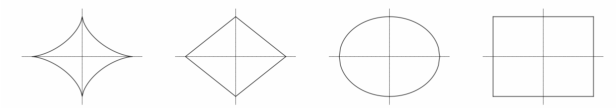
1. The *trace* of an $n \times n$ matrix A, usually denoted as $\operatorname{tr}(A)$, is defined as the sum of the n diagonal components of A.
2. The trace of an $n \times n$ matrix A equals the sum of the n eigenvalues of A, i.e., $\operatorname{tr} A = a_{11} + a_{22} + \cdots + a_{nn} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.
3. The trace of a 1×1 matrix, a scalar, is itself.
4. If $A \in \mathbb{R}^{m\times k}$ and $B \in \mathbb{R}^{k\times m}$, then $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.
5. $\operatorname{tr}(A + \gamma B) = \operatorname{tr}(A) + \gamma\operatorname{tr}(B)$, where γ is a scalar.
6. $\operatorname{tr}(AB) = (\operatorname{Vec} A^{\mathrm{T}})^{\mathrm{T}} \operatorname{Vec} B$ (see Section 2.4.18 for the Vec operation)

2.4.7 MATRIX NORMS

1. The mapping $g : \mathbb{R}^{m\times n} \Rightarrow \mathbb{R}$ is a *matrix norm* if g satisfies the same three properties as a vector norm:
   (a) $g(A) \geq 0$ for all A and $g(A) = 0$ if and only if $A \equiv 0$, so that (in norm notation) $\|A\| > 0$ for all non-zero A.
   (b) For two matrices $A, B \in \mathbb{R}^{m\times n}$, $g(A + B) \leq g(A) + g(B)$, so that $\|A + B\| \leq \|A\| + \|B\|$.
   (c) $g(rA) = |r|g(A)$, where $r \in \mathbb{R}$, so that $\|\gamma A\| = |\gamma|\ \|A\|$.
2. The L^p *norm* of a matrix A is the number defined by

$$\|A\|_p = \sup_{\mathbf{x}\neq \mathbf{0}} \frac{\|A\mathbf{x}\|_p}{\|\mathbf{x}\|_p} \tag{2.4.14}$$

   where $\|\cdot\|_p$ represents one of the L^p (vector) norms with $p = 1, 2$, or ∞.
   (a) The L^1 norm of $A_{m\times n}$ is defined as $\|A\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |a_{ij}|$.
   (b) The L^2 norm of A is the square root of the largest eigenvalue of $A^{\mathrm{T}}A$, (i.e., $\|A\|_2^2 = \lambda_{\max}(A^{\mathrm{T}}A)$), which is the same as the largest singular value of A, $\|A\|_2 = \sigma_1(A)$.
   (c) The L^∞ norm is defined as $\|A\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |a_{ij}|$.
   (d) The unit circle, $||\mathbf{x}|| = 1$ in different norms ($p = 2/3$, 1, 2, ∞):



3. The *Frobenius* or *Hilbert–Schmidt norm* of a matrix A is the number $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$ which satisfies $\|A\|_F^2 = \mathrm{tr}(A^{\mathrm{T}}A)$.
   Since $\mathbb{R}^{m\times n}$ is isomorphic to $\mathbb{R}^{mn}$, the Frobenius norm can be interpreted as the L^2 norm of an $nm \times 1$ column vector in which each column of A is appended to the next in succession.
4. If A is symmetric, then $\|A\|_2 = \max_j |\lambda_j|$, where $\{\lambda_j\}$ are A's eigenvalues.
5. The following properties hold:

$$\begin{aligned} &\frac{1}{\sqrt{m}} \|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\,\|A\|_1\,, \\ &\max_{i,j} |a_{ij}| \leq \|A\|_2 \leq \sqrt{mn}\ \max_{i,j} |a_{ij}|, \ \text{ and} \\ &\frac{1}{\sqrt{n}} \|A\|_\infty \leq \|A\|_2 \leq \sqrt{m}\,\|A\|_\infty\,. \end{aligned} \tag{2.4.15}$$

6. The matrix p norms satisfy the additional property of *consistency*, defined as $\|AB\|_p \leq \|A\|_p \|B\|_p$.
7. The Frobenius norm is *compatible* with the vector 2 norm, that is $\|A\mathbf{x}\|_F \leq \|A\|_F\ \|\mathbf{x}\|_2$. Additionally, the Frobenius norm satisfies the condition $\|A\|_2 \leq \|A\|_F \leq \sqrt{n}\,\|A\|_2$.

2.4.8 SINGULARITY, RANK, AND INVERSES

1. An $n \times n$ matrix A is called *singular* if there exists a vector $\mathbf{x} \neq \mathbf{0}$ such that $A\mathbf{x} = \mathbf{0}$ or $A^{\mathrm{T}}\mathbf{x} = \mathbf{0}$. (Note that $\mathbf{x} = \mathbf{0}$ means all components of $\mathbf{x}$ are zero). If a matrix is not singular, it is *non-singular*.
2. $(AB)^{-1} = B^{-1}A^{-1}$, provided all inverses exist.
3. $(A^{-1})^{\mathrm{T}} = (A^{\mathrm{T}})^{-1}$.
4. $(\gamma A)^{-1} = (1/\gamma)A^{-1}$.
5. If D_w is a diagonal matrix, then $D_w^{-1} = D_{1/w}$.

   EXAMPLE If $D_w = \begin{bmatrix} w_{11} & 0 \\ 0 & w_{22} \end{bmatrix}$ then $D_w^{-1} = D_{1/w} = \begin{bmatrix} \frac{1}{w_{11}} & 0 \\ 0 & \frac{1}{w_{22}} \end{bmatrix}$.
6. Partitioning: Let $A = \left[\begin{smallmatrix} B & C \\ D & E \end{smallmatrix}\right]$. Assuming that all inverses exist, then $A^{-1} = \left[\begin{smallmatrix} X & Y \\ Z & U \end{smallmatrix}\right]$, where
$$X = (B - CE^{-1}D)^{-1}, \qquad U = (E - DB^{-1}C)^{-1},$$
$$Y = -B^{-1}CU, \qquad Z = -E^{-1}DX.$$
7. The inverse of a 2×2 matrix is as follows (defined when $ad \neq bc$):
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$
8. If A and B are compatible square matrices and are both invertible, then
$$(A + B)^{-1} = B^{-1}\left(A^{-1} + B^{-1}\right)^{-1}A^{-1} = A^{-1}\left(A^{-1} + B^{-1}\right)^{-1}B^{-1}.$$
9. If A and B are compatible square matrices and B is of rank one, then
$$(A + B)^{-1} = A^{-1} - \frac{1}{1 + \alpha}A^{-1}BA^{-1}$$
where α is the trace of $\left(BA^{-1}\right)$.

   EXAMPLE If $C = \left[\begin{smallmatrix} c & 1 & 1 \\ 1 & c & 1 \\ 1 & 1 & c \end{smallmatrix}\right]$ with $c \neq 1, -2$ then $C = A + B$ where $A = (c-1)I = \left[\begin{smallmatrix} c-1 & 0 & 0 \\ 0 & c-1 & 0 \\ 0 & 0 & c-1 \end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{smallmatrix}\right]$. Since $\alpha = \text{trace}\left(BA^{-1}\right) = \frac{3}{c-1}$
$$C^{-1} = (A + B)^{-1} = \frac{1}{c-1}\left[I - \frac{1}{c+2}B\right] = \frac{1}{(c-1)(c+2)}\left[\begin{smallmatrix} c+1 & -1 & -1 \\ -1 & c+1 & -1 \\ -1 & -1 & c+1 \end{smallmatrix}\right]$$
10. Let $B = (b_{ij})$ have inverse $B^{-1} = \left(b^{ij}\right)$. Let $A = B$ except for one element $a_{rs} = b_{rs} + k$. Then the elements of A^{-1} are: $a^{ij} = b^{ij} - \dfrac{kb^{ir}b^{sj}}{1 + kb^{sr}}$.
11. The *row rank* of a matrix A is defined as the number of linearly independent rows of A. Likewise, the *column rank* equals the number of linearly independent columns of A. For any matrix, the row rank equals the column rank.
12. If $A \in \mathbb{R}^{n \times n}$ has rank of n, then A is said to have *full rank*.
13. A square matrix is invertible if and only if it has full rank.
14. $\text{Rank}(AB) \leq \min\left[\text{rank}(A),\ \text{rank}(B)\right]$.
15. $\text{Rank}(A^{\mathrm{T}}A) = \text{rank}(AA^{\mathrm{T}}) = \text{rank}(A)$.

2.4.9 SYSTEMS OF LINEAR EQUATIONS

1. If A is a matrix then $A\mathbf{x} = \mathbf{b}$ is a *system of linear equations*. If A is square and non-singular, there exists a unique solution $\mathbf{x} = A^{-1}\mathbf{b}$.
2. For the linear system $A\mathbf{x} = \mathbf{b}$ involving n variables and m equations

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1m}x_m &= b_1, \\ &\vdots \\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nm}x_m &= b_n, \end{aligned} \tag{2.4.16}$$

the possible outcomes when searching for a solution are:
   (a) A unique solution exists, and the system is called *consistent*.
   (b) No such solution exists and the system is called *inconsistent*.
   (c) Multiple solutions exist, the system has an infinite number of solutions, and the system is called *undetermined*.

3. The solvability cases of the linear systems $A\mathbf{x} = \mathbf{b}$ (when A is $m \times n$) are:
   (a) If $\text{rank}(A) = m = n$, then there is a unique solution.
   (b) If $\text{rank}(A) = m < n$, then there is an exact solution with free parameters.
   (c) If $\text{rank}(A) = n < m$, then either there is a unique solution, or there is a unique least squares solution.
   (d) If $\text{rank}(A) < m < n$, or $\text{rank}(A) < n < m$, or $\text{rank}(A) < n = m$: then either there is an exact solution with free parameters, or there are non-unique least squares solutions.

4. *Fredholm's alternative*
   Either $A\mathbf{x} = \mathbf{b}$ has a solution or $\mathbf{y}^{\mathrm{T}}A = 0$ has a solution with $\mathbf{y}^{\mathrm{T}}\mathbf{b} \neq 0$.
5. If the system of equations $A\mathbf{x} = \mathbf{b}$ is underdetermined, then we may find the $\mathbf{x}$ that minimizes $\|A\mathbf{x} - \mathbf{b}\|_p$ for some p.

EXAMPLE If $A = \begin{bmatrix}1 & 1 & 1\end{bmatrix}^{\mathrm{T}}$ and $\mathbf{b} = \begin{bmatrix}b_1 & b_2 & b_3\end{bmatrix}^{\mathrm{T}}$ with $b_1 \geq b_2 \geq b_3 \geq 0$ then the minimum value of $\|A\mathbf{x} - \mathbf{b}\|_p$ in different norms is as follows:

$$\begin{aligned} p = 1 &\Rightarrow \mathbf{x}_{\text{optimal}} = b_2 \\ p = 2 &\Rightarrow \mathbf{x}_{\text{optimal}} = (b_1 + b_2 + b_3)/3 \\ p = \infty &\Rightarrow \mathbf{x}_{\text{optimal}} = (b_1 + b_3)/2 \end{aligned} \tag{2.4.17}$$

6. The *condition number* of the square matrix A is

$$\text{cond}(A) = \left\|A^{+}\right\| \; \|A\| \tag{2.4.18}$$

where A^{+} is the pseudo-inverse of A and $\|\cdot\|$ is any of the L^p norms. When A is non-singular, this is equivalent to

$$\text{cond}(A) = \left\|A^{-1}\right\| \; \|A\|. \tag{2.4.19}$$

In all cases, $\text{cond}(A) \geq 1$. When $\text{cond}(A)$ is equal to one, A is *perfectly conditioned*. Matrices with small condition numbers are *well-conditioned*. If $\text{cond}(A)$ is large, then A is *ill-conditioned*.

7. For the system of equations $(A+\epsilon F)\mathbf{x}(\epsilon) = (\mathbf{b}+\epsilon\mathbf{f})$, the solution satisfies

$$\frac{\|\mathbf{x}(\epsilon)-\mathbf{x}(0)\|}{\|\mathbf{x}(0)\|} \leq \text{cond}(A)\left(\epsilon\frac{\|F\|}{\|A\|} + \epsilon\frac{\|\mathbf{f}\|}{\|\mathbf{b}\|}\right) + O(\epsilon^2). \qquad (2.4.20)$$

8. For a square matrix, the size of its determinant is *not* related to its condition number. For example, the $n \times n$ matrices below have $\kappa_\infty(B_n) = n2^{n-1}$ and $\det(B_n) = 1$; $\kappa_p(D_n) = 1$ and $\det(D_n) = 10^{-n}$.

$$B_n = \begin{bmatrix} 1 & -1 & \dots & -1 \\ 0 & 1 & & -1 \\ \vdots & & \ddots & \\ 0 & 0 & & 1 \end{bmatrix}, \qquad D_n = \text{diag}(10^{-1}, \dots, 10^{-1}).$$

9. Let $A = (a_{ij})$ be an $n \times n$ matrix. Using the L^2 condition number we find $\text{cond}\, A = \max_j |\lambda_j(A)| \,/\, \min_i |\lambda_i(A)|$:

Matrix $A_{n\times n} = (a_{ij})$	Condition number
A is orthogonal	$\text{cond}\,(A) = 1$
$a_{ij} = n\delta_{ij} + 1$	$\text{cond}\,(A) = 2$
$a_{ij} = (i+j)/p, \quad n = p-1$, p a prime	$\text{cond}\,(A) = \sqrt{n+1}$
The circulant whose first row is $(1, 2, \dots, n)$	$\text{cond}\,(A) \sim n$
$a_{ij} = \begin{cases} i/j & \text{if } i \leq j \\ j/i & \text{if } i > j \end{cases}$	$\text{cond}\,(A) \sim cn^{1+\epsilon}$, $0 \leq \epsilon \leq 1$
$a_{ij} = \begin{cases} -2 & \text{if } i = j \\ 1 & \text{if } \|i-j\| = 1 \\ 0 & \text{if } \|i-j\| \geq 2 \end{cases}$	$\text{cond}\,(A) \sim \dfrac{4n^2}{\pi^2}$
$a_{ij} = 2\min(i,j) - 1$	$\text{cond}\,(A) \sim 16n^2/\pi^2$
$a_{ij} = (i+j-1)^{-1}$ (Hilbert matrix)	$\log \text{cond}\,(A) \sim Kn$, $K \approx 3.5$

2.4.10 MATRIX EXPONENTIALS

1. *Matrix exponentiation* is defined as (the series always converges):

$$e^{At} = I + At + \frac{(At)^2}{2!} + \frac{(At)^3}{3!} + \cdots. \qquad (2.4.21)$$

2. Common properties of matrix exponentials are:
   (a) $\left(e^{As}\right)\left(e^{At}\right) = e^{A(s+t)}$,
   (b) $\left(e^{At}\right)\left(e^{-At}\right) = I$,
   (c) $\frac{d}{dt}e^{At} = Ae^{At}$,
3. When A and B are square matrices $C = [B, A] = BA - AB$. is the *commutator* of A and B. Note that $e^{(A+B)} = e^A e^B e^{C/2}$ provided that $[C, A] = [C, B] = \mathbf{0}$ (i.e., each of A and B commute with their commutator). In particular, if A and B commute then $e^{A+B} = e^A e^B$.
4. For a matrix $A \in \mathbb{R}^{n\times n}$ $\qquad \det\left(e^{At}\right) = e^{\lambda_1 t} e^{\lambda_2 t} \cdots e^{\lambda_n t} = e^{\text{tr}(A)t}$.
5. The diagonalization of e^{At}, when A is diagonalizable, is $e^{At} = Se^{Dt}S^{-1}$ where the columns of S are the eigenvectors of A, and the entries of the diagonal matrix D are the corresponding eigenvalues of A, that is, $A = SDS^{-1}$.

2.4.11 LINEAR SPACES AND LINEAR MAPPINGS

1. A *subspace* is the space generated by linear combinations of any set of rows or set of columns of a real matrix.
2. Let $R(A)$ and $N(A)$ denote, respectively, the range space and null space of an $m \times n$ matrix A. They are defined by:

$$\begin{aligned} R(A) &= \{\mathbf{y} \mid \mathbf{y} = A\mathbf{x};\ \text{for some}\ \ \mathbf{x} \in \mathbb{R}^n\}, \\ N(A) &= \{\mathbf{x} \in \mathbb{R}^n \mid A\mathbf{x} = \mathbf{0}\}. \end{aligned} \tag{2.4.22}$$

3. The *projection matrix*, onto a subspace S, denoted P_S, is the unique square matrix possessing the three properties:
   (a) $P_S = P_S^{\mathrm{T}}$;
   (b) $P_S^2 = P_S$ (the projection matrix is *idempotent*);
   (c) The vector $\mathbf{b}_S$ lies in the subspace S if and only if $\mathbf{b}_S = P_S\mathbf{v}$ for some vector $\mathbf{v}$. In other words, $\mathbf{b}_S$ can be written as a linear combination of the columns of P_S.
4. When the $m \times n$ matrix A (with $n \leq m$) has rank n, the projection of A onto the subspaces of A is given by:

$$\begin{aligned} P_{R(A)} &= A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}, \\ P_{R(A^{\mathrm{T}})} &= I, \\ P_{N(A^{\mathrm{T}})} &= I - A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}. \end{aligned} \tag{2.4.23}$$

   When A is of rank m, the projection of A onto the subspaces of A is given by:

$$\begin{aligned} P_{R(A)} &= I, \\ P_{R(A^{\mathrm{T}})} &= A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A, \\ P_{N(A)} &= I - A^{\mathrm{T}}(AA^{\mathrm{T}})^{-1}A. \end{aligned} \tag{2.4.24}$$

5. When A is not of full rank, the matrix $A\tilde{A}$ satisfies the requirements for a projection matrix, where $\tilde{A}$ is the coefficient matrix of the system of equations $\mathbf{x}_+ = \tilde{A}\mathbf{b}$, generated by the *least squares* problem $\min \|\mathbf{b} - A\mathbf{x}\|_2^2$. Thus,

$$\begin{aligned} P_{R(A)} &= A\tilde{A}, \\ P_{R(A^{\mathrm{T}})} &= \tilde{A}A, \\ P_{N(A)} &= I - \tilde{A}A, \\ P_{N(A^{\mathrm{T}})} &= I - A\tilde{A}. \end{aligned} \tag{2.4.25}$$

6. A matrix $B \in \mathbb{R}^{n \times n}$ is called *similar* to a matrix $A \in \mathbb{R}^{n \times n}$ if $B = T^{-1}AT$ for some non-singular matrix T.
7. If B is similar to A, then B has the same eigenvalues as A.
8. If B is similar to A and if $\mathbf{x}$ is an eigenvector of A, then $\mathbf{y} = T^{-1}\mathbf{x}$ is an eigenvector of B corresponding to the same eigenvalue.

2.4.12 GENERALIZED INVERSES

1. Every matrix A (singular or non-singular, rectangular or square) has a *generalized inverse*, or *pseudo-inverse,* A^+ defined by the *Moore–Penrose* conditions:

$$\begin{aligned} AA^+A &= A, \\ A^+AA^+ &= A^+, \\ (AA^+)^{\mathrm{T}} &= AA^+, \\ (A^+A)^{\mathrm{T}} &= A^+A. \end{aligned} \tag{2.4.26}$$

2. There is a unique pseudo-inverse satisfying the conditions in (2.4.26). If (and only if) A is square and non-singular, then $A^+ = A^{-1}$.
3. If A is a rectangular $m \times n$ matrix of rank n, with $m > n$, then A^+ is of order $n \times m$ and $A^+A = I \in \mathbb{R}^{n \times n}$. In this case A^+ is called a *left inverse*, and $AA^+ \neq I$.
4. If A is a rectangular $m \times n$ matrix of rank m, with $m < n$, then A^+ is of order $n \times m$ and $AA^+ = I \in \mathbb{R}^{m \times m}$. In this case A^+ is called a *right inverse*, and $A^+A \neq I$.
5. For a square singular matrix A, $AA^+ \neq I$, and $A^+A \neq I$.
6. The matrices AA^+ and A^+A are idempotent.
7. The least squares problem is to find the **x** that minimizes $\|\mathbf{y} - A\mathbf{x}\|_2$. The **x** of least norm is $\mathbf{x} = A^+\mathbf{y}$.
8. The pseudo-inverse is ill-conditioned with respect to rank changing perturbations. For example

$$\left(\begin{bmatrix} 1 & -1 \\ 2 & -2 \end{bmatrix} + \epsilon \begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}\right)^+ = \frac{1}{\epsilon^2}\begin{bmatrix} -1 & \frac{1}{2} \\ -1 & \frac{1}{2} \end{bmatrix} + \frac{1}{\epsilon}\begin{bmatrix} 1 & 0 \\ 0 & \frac{1}{2} \end{bmatrix}.$$

9. *Computing the pseudo-inverse*: The pseudo-inverse of $A_{m \times n}$ can be determined by the singular value decomposition $A = U\Sigma V^{\mathrm{T}}$. If A has rank $r > 0$, then $\Sigma_{m \times n}$ has r positive singular values (σ_i) along the main diagonal extending from the upper left-hand corner and the remaining components of Σ are zero. Then $A^+ = (U\Sigma V^{\mathrm{T}})^+ = (V^{\mathrm{T}})^+\Sigma^+U^+ = V\Sigma^+U^{\mathrm{T}}$ since $(V^{\mathrm{T}})^+ = V$ and $U^+ = U^{\mathrm{T}}$ because of their orthogonality. The components σ_i^+ in Σ^+ are

$$\sigma_i^+ = \begin{cases} 1/\sigma_i, & \text{if } \sigma_i \neq 0\,; \\ 0, & \text{if } \sigma_i = 0. \end{cases} \tag{2.4.27}$$

10. *Computing the pseudo-inverse*: Let $A_{m \times n}$ be of rank r. Select r rows and r columns which form a basis of A. Then compute the pseudo-inverse of A as follows: invert the regular $r \times r$ matrix and place the inverse (without transposing) into the r rows corresponding to the column numbers and the r columns corresponding to the row numbers of the basis, and place zero into the remaining component positions. For example, if A is of order 5×4 and rank 3, and if rows 1, 2, 4 and columns 2, 3, 4 are selected as a basis of A, then A^+ will be of order 4×5 and will contain the inverse components of the basis in rows 2, 3, 4 and column 1, 2, 4, and zeros elsewhere.

2.4.13 EIGENSTRUCTURE

1. If A is a square $n \times n$ matrix, then the n^{th} degree polynomial defined by $\det(A - \lambda I) = 0$ is called the *characteristic polynomial*, or *characteristic equation* of A.
2. The n roots (not necessarily distinct) of the characteristic polynomial are called the *eigenvalues* (or *characteristic roots*) of A. Therefore, the values, λ_i, $i = 1, \ldots, n$, are eigenvalues if and only if $|A - \lambda_i I| = 0$.
3. The characteristic polynomial $\det(A - \lambda I) = \sum_{i=0}^{n} r_i \lambda^i$ has the properties

$$\begin{aligned} r_n &= (-1)^n, \\ r_{n-1} &= -r_n \operatorname{tr}(A), \\ r_{n-2} &= -\frac{1}{2}\left[r_{n-1} \operatorname{tr}(A) + r_n \operatorname{tr}(A^2)\right], \\ r_{n-3} &= -\frac{1}{3}\left[r_{n-2} \operatorname{tr}(A) + r_{n-1} \operatorname{tr}(A^2) + r_n \operatorname{tr}(A^3)\right], \\ &\vdots \\ r_0 &= -\frac{1}{n}\left[\sum_{p=1}^{n} r_p \operatorname{tr}(A^p)\right]. \end{aligned}$$

4. Each eigenvalue λ has a corresponding *eigenvector* $\mathbf{x}$ (different from $\mathbf{0}$) that solves the system $A\mathbf{x} = \lambda\mathbf{x}$, or $(A - \lambda I)\mathbf{x} = \mathbf{0}$.
5. If $\mathbf{x}$ solves $A\mathbf{x} = \lambda\mathbf{x}$, then so does $\gamma\mathbf{x}$, where γ is an arbitrary scalar.
6. *Cayley–Hamilton theorem*: Any matrix A satisfies its own characteristic equation. That is $\sum_{i=0}^{n} r_i A^i = 0$.
7. If A is a real matrix with positive eigenvalues, then

$$\lambda_{\text{min}}(AA^{\mathrm{T}}) \le [\lambda_{\text{min}}(A)]^2 \le [\lambda_{\text{max}}(A)]^2 \le \lambda_{\text{max}}(AA^{\mathrm{T}}), \tag{2.4.28}$$

   where λ_{min} denotes the smallest and λ_{max} the largest eigenvalue.
8. If all the eigenvalues of a real symmetric matrix are distinct, then their associated eigenvectors are also distinct (linearly independent).
9. The determinant of a matrix is equal to the product of the eigenvalues. That is, if A has the eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, then $\det(A) = \lambda_1 \lambda_2 \cdots \lambda_n$.
10. The following table shows the eigenvalues of specific matrices

matrix	eigenvalues
diagonal matrix	diagonal elements
upper or lower triangular	diagonal elements
A is $n \times n$ and nilpotent	0 (n times)
A is $n \times n$ and idempotent of rank r	1 (r times); and 0 ($n - r$ times)
$(a - b)I_n + bJ_n$, where J_n is the $n \times n$ matrix of all 1's	$a + (n-1)b$; and $a - b$ ($n - 1$ times)

11. The eigenvalues of a triangular (or diagonal) matrix are the diagonal components of the matrix.
12. The eigenvalues of idempotent matrices are zeros and ones.
13. Real symmetric and Hermitian matrices have real eigenvalues.
14. Let A have the eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$. The eigenvalues of some functions of A are shown below:

matrix	eigenvalues
A^{T}	eigenvalues of A
A^{H}	complex conjugates of $\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\}$
A^k, k an integer	$\{\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k\}$
A^{-k}, k an integer, A non-singular	$\{\lambda_1^{-k}, \lambda_2^{-k}, \ldots, \lambda_n^{-k}\}$
$q(A)$, q is a polynomial	$\{q(\lambda_1^k), q(\lambda_2^k), \ldots, q(\lambda_n^k)\}$
SAS^{-1}, S non-singular	eigenvalues of A
AB, where A is $m \times n$, B is $n \times m$, and $m \geq n$	eigenvalues of BA; and 0 ($m - n$ times)

2.4.14 MATRIX FACTORIZATIONS

1. *Singular value decomposition* (SVD): Any $m \times n$ matrix A can be written as the product $A = U\Sigma V^{\mathrm{T}}$, where U is an $m \times m$ orthogonal matrix, V is an $n \times n$ orthogonal matrix, and $\Sigma = \operatorname{diag}(\sigma_1, \sigma_2, \ldots, \sigma_p)$, with $p = \min(m, n)$ and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$. The values σ_i, $i = 1, \ldots, p$, are called the *singular values* of A.
   (a) When $\operatorname{rank}(A) = r > 0$, A has exactly r positive singular values, and $\sigma_{r+1} = \cdots = \sigma_p = 0$.
   (b) When A is a symmetric $n \times n$ matrix, then $\sigma_1 = |\lambda_1|, \ \ldots, \ \sigma_n = |\lambda_n|$, where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A.
   (c) When A is an $m \times n$ matrix, if $m \geq n$ then the singular values of A are the square roots of the eigenvalues of $A^{\mathrm{T}}A$. Otherwise, they are the square roots of the eigenvalues of AA^{T}.
2. *Schur decomposition*: If $A \in \mathbb{C}^{n\times n}$ then $A = UTU^{\mathrm{H}}$, where U is a unitary matrix and T is an upper triangular matrix which has the eigenvalues of A along its diagonal.
3. *QR factorization*: If all the columns of $A \in \mathbb{R}^{m\times n}$ are linearly independent, then A can be factored as $A = QR$, where $Q \in \mathbb{R}^{m\times n}$ has orthonormal columns and $R \in \mathbb{R}^{n\times n}$ is upper triangular and non-singular.
4. If $A \in \mathbb{R}^{n\times n}$ is symmetric positive definite, then

$$A = LDL^{\mathrm{T}} = LD^{1/2}D^{1/2}L^{\mathrm{T}} = \left(LD^{1/2}\right)\left(LD^{1/2}\right)^{\mathrm{T}} = GG^{\mathrm{T}}$$

   where L is a lower triangular matrix and D is a diagonal matrix. The *Cholesky factorization* is $A = GG^{\mathrm{T}}$ where the matrix G is the *Cholesky triangle*.
5. Any $m \times n$ matrix A can be factored as $PA = LU$, where P is a permutation matrix, L is lower triangular, and U is upper triangular.

2.4.15 MATRIX DIAGONALIZATION

1. If $A \in \mathbb{R}^{n\times n}$ possesses n linearly independent eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$, then A can be diagonalized as $S^{-1}AS = \Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, where the eigenvectors of A are chosen to comprise the columns of S.
2. If $A \in \mathbb{R}^{n\times n}$ can be diagonalized into $S^{-1}AS = \Lambda$, then $A^k = S\Lambda^k S^{-1}$, or $\Lambda^k = S^{-1}A^k S$.
3. *Spectral decomposition*: Any real symmetric matrix $A \in \mathbb{R}^{n\times n}$ can be diagonalized into the form $A = U\Lambda U^{\mathrm{T}}$, where Λ is the diagonal matrix of ordered eigenvalues of A such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, and the columns of U are the corresponding n *orthonormal* eigenvectors of A.
   That is, if $A \in \mathbb{R}^{n\times n}$ is symmetric, then a real orthogonal matrix Q exists such that $Q^{\mathrm{T}}AQ = \operatorname{diag}(\lambda_1, \ldots, \lambda_n) =$ diagonal matrix with $\{\lambda_i\}$ on the diagonal.
4. The *spectral radius* of a real symmetric matrix A, commonly denoted by $\rho(A)$, is defined as $\rho(A) = \max_{1\leq i\leq n} |\lambda_i(A)|$.
5. If $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times n}$ are diagonalizable, then they share a common eigenvector matrix S if and only if $AB = BA$. (Not every eigenvector of A need be an eigenvector for B, e.g., the above equation is always true if $A = I$.)
6. *Schur decomposition*: If $A \in \mathbb{C}^{n\times n}$, then a unitary matrix $Q \in \mathbb{C}^{n\times n}$ exists such that $Q^{\mathrm{H}}AQ = D + N$, where $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ and $N \in \mathbb{C}^{n\times n}$ is strictly upper triangular (i.e., there are zeros on the diagonal). The matrix Q can be chosen so that the eigenvalues λ_i appear in any order along the diagonal.
7. If $A \in \mathbb{R}^{n\times n}$ possesses $s \leq n$ linearly independent eigenvectors, it is similar to a matrix with s *Jordan blocks* (for some matrix M)

$$J = M^{-1}AM = \begin{bmatrix} J_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \ddots & \\ & & & J_s \end{bmatrix},$$

   where each Jordan block J_i is an upper triangular matrix with (a) the single eigenvalue λ_i repeated n_i times along the main diagonal; (b) $(n_i - 1)$ 1's appearing above the diagonal entries; and (c) all other components zero:

$$J_i = \begin{bmatrix} \lambda_i & 1 & & \mathbf{0} \\ & \ddots & 1 & \\ & & \ddots & 1 \\ \mathbf{0} & & & \lambda_i \end{bmatrix}. \tag{2.4.29}$$

2.4.16 QUADRATIC FORMS

1. For a symmetric matrix A, the map $\mathbf{x} \mapsto \mathbf{x}^{\mathrm{T}} A\mathbf{x}$ is called a *pure quadratic form*. It has the form

$$\mathbf{x}^{\mathrm{T}} A\mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} x_i x_j = a_{11} x_1^2 + a_{12} x_1 x_2 + a_{21} x_2 x_1 + \cdots + a_{nn} x_n^2. \tag{2.4.30}$$

2. For A symmetric, the gradient of $\mathbf{x}^{\mathrm{T}} A\mathbf{x}/\mathbf{x}^{\mathrm{T}}\mathbf{x}$ equals zero if and only if $\mathbf{x}$ is an eigenvector of A. Thus, the stationary values of this function (where the gradient vanishes) are the eigenvalues of A.
3. The ratio of two quadratic forms (B non-singular) $u(\mathbf{x}) = (\mathbf{x}^{\mathrm{T}} A\mathbf{x})/(\mathbf{x}^{\mathrm{T}} B\mathbf{x})$ attains stationary values at the eigenvalues of $B^{-1}A$. In particular,

$$u_{\text{max}} = \lambda_{\text{max}}(B^{-1}A), \qquad \text{and} \qquad u_{\text{min}} = \lambda_{\text{min}}(B^{-1}A).$$

4. A matrix A is *positive definite* if $\mathbf{x}^{\mathrm{T}} A\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$.
5. A matrix A is *positive semi-definite* if $\mathbf{x}^{\mathrm{T}} A\mathbf{x} \geq 0$ for all $\mathbf{x}$.
6. For a real, symmetric matrix $A \in \mathbb{R}^{n\times n}$, the following are necessary and sufficient conditions to establish the positive definiteness of A:

   (a) All eigenvalues of A have $\lambda_i > 0$, for $i = 1, \ldots, n$, and

   (b) The upper-left sub-matrices of A, called the *principal sub-matrices*, defined by $A_1 = \begin{bmatrix} a_{11} \end{bmatrix}$,

$$A_2 = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \qquad \ldots, \qquad A_n = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix},$$

   have $\det(A_k) > 0$, for all $k = 1, \ldots, n$.

7. If A is positive definite, then all of the principal sub-matrices of A are also positive definite. Additionally, all diagonal entries of A are positive.
8. For a real, symmetric matrix $A \in \mathbb{R}^{n\times n}$, the following are necessary and sufficient conditions to establish the positive semi-definiteness of A:

   (a) All eigenvalues of A have $\lambda_i \geq 0$, for $i = 1, \ldots, n$,

   (b) The principal sub-matrices of A have $\det A_k \geq 0$, for all $k = 1, \ldots, n$.

9. If A is positive semi-definite, then all of the principal sub-matrices of A are also positive semi-definite. Additionally, all diagonal entries of A are non-negative.
10. If the matrix Q is positive definite, then $(\mathbf{x}^{\mathrm{T}} - \mathbf{x}_0^{\mathrm{T}})Q^{-1}(\mathbf{x} - \mathbf{x}_0) = 1$ is the equation of an ellipsoid with its center at $\mathbf{x}_0^{\mathrm{T}}$. The lengths of the semi-axes are equal to the square roots of the eigenvalues of Q; see page 234.

2.4.17 THEOREMS

1. *Frobenius–Perron theorem:* If $A > 0$ (that is, $a_{ij} > 0$) then there exists $\lambda_0 > 0$ and $\mathbf{x}_0 > \mathbf{0}$ such that
   (a) $A\mathbf{x}_0 = \lambda_0\mathbf{x}_0$,
   (b) if λ is any other eigenvalue of A, $\lambda \neq \lambda_0$, then $|\lambda| < \lambda_0$, and
   (c) λ_0 is an eigenvalue with geometric and algebraic multiplicity equal to one. (That is, there is no $\mathbf{x}_1 \neq \mathbf{x}_0$ with $A\mathbf{x}_1 = \lambda_0\mathbf{x}_1$.)
2. If $A \geq 0$ (that is, $a_{ij} \geq 0$), and $A^k > 0$ for some positive integer k, then the results of the Frobenius–Perron theorem apply to A.
3. *Courant–Fischer minimax theorem:* If $\lambda_i(A)$ denotes the i^{th} largest eigenvalue of a matrix $A = A^{\mathrm{T}} \in \mathbb{R}^{n\times n}$, then

$$\lambda_j(A) = \max_{S_j} \min_{\mathbf{0}\neq\mathbf{x}\in S_j} \frac{\mathbf{x}^{\mathrm{T}}A\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}} \qquad j = 1, \ldots, n \tag{2.4.31}$$

   where $\mathbf{x} \in \mathbb{R}^n$ and S_j is a j-dimensional subspace.
   From this follows *Raleigh's principle*: The quotient $R(\mathbf{x}) = \mathbf{x}^{\mathrm{T}}A\mathbf{x}/\mathbf{x}^{\mathrm{T}}\mathbf{x}$ is minimized by the eigenvector $\mathbf{x} = \mathbf{x}_1$ corresponding to the smallest eigenvalue λ_1 of A. The minimum of $R(\mathbf{x})$ is λ_1, that is,

$$\min R(\mathbf{x}) = \min \frac{\mathbf{x}^{\mathrm{T}}A\mathbf{x}}{\mathbf{x}^{\mathrm{T}}\mathbf{x}} = R(\mathbf{x}_1) = \frac{\mathbf{x}_1^{\mathrm{T}}A\mathbf{x}_1}{\mathbf{x}_1^{\mathrm{T}}\mathbf{x}} = \frac{\mathbf{x}_1^{\mathrm{T}}\lambda_1\mathbf{x}_1}{\mathbf{x}_1^{\mathrm{T}}\mathbf{x}} = \lambda_1. \tag{2.4.32}$$

4. *Cramer's rule:* The j^{th} component of $\mathbf{x} = A^{-1}\mathbf{b}$ is given by

$$x_j = \frac{\det B_j}{\det A}, \qquad \text{where}$$
$$B_j = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1,j-1} & b_1 & a_{1,j+1} & \ldots & a_{1n} \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \ldots & a_{n,j-1} & b_n & a_{n,j+1} & \ldots & a_{nn} \end{bmatrix} \tag{2.4.33}$$

   The vector $\mathbf{b} = \begin{bmatrix} b_1 & \ldots & b_n \end{bmatrix}^{\mathrm{T}}$ replaces the j^{th} column of the matrix A to form the matrix B_j.
5. *Sylvester's law of inertia:* For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, the matrices A and $C^{\mathrm{T}}AC$, for C non-singular, have the same number of positive, negative, and zero eigenvalues.
6. *Gerschgorin circle theorem*: Each eigenvalue of an arbitrary $n \times n$ matrix $A = (a_{ij})$ lies in at least one of the circles $\{C_1, C_2, \ldots, C_n\}$ in the complex plane, where circle C_i has center a_{ii} and radius ρ_i given by $\rho_i = \displaystyle\sum_{\substack{j=1\\j\neq i}}^{n} |a_{ij}|$.

2.4.18 THE VECTOR OPERATION

The matrix $A_{m\times n}$ can be represented as a collection of $m \times 1$ column vectors: $A = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n\end{bmatrix}$. Define Vec A as the matrix of size $nm \times 1$ (i.e., a vector) by

$$\operatorname{Vec} A = \begin{bmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_n\end{bmatrix}. \tag{2.4.34}$$

This operator has the following properties:

1. $\operatorname{tr} AB = \left(\operatorname{Vec} A^{\mathrm{T}}\right)^{\mathrm{T}} \operatorname{Vec} B$.
2. The permutation matrix U that associates $\operatorname{Vec} X$ and $\operatorname{Vec} X^{\mathrm{T}}$ (that is, $\operatorname{Vec} X^{\mathrm{T}} = U \operatorname{Vec} X$) is given by:

$$U = \begin{bmatrix}\operatorname{Vec} E_{11}^{\mathrm{T}} & \operatorname{Vec} E_{21}^{\mathrm{T}} & \dots & \operatorname{Vec} E_{n1}^{\mathrm{T}}\end{bmatrix} = \sum_{r,s} E_{rs} \otimes E_{rs}^{\mathrm{T}}. \tag{2.4.35}$$

3. $\operatorname{Vec}(AYB) = (B^{\mathrm{T}} \otimes A) \operatorname{Vec} Y$.
4. If A and B are both of size $n \times n$, then
   (a) $\operatorname{Vec} AB = (I_n \otimes A) \operatorname{Vec} B$.
   (b) $\operatorname{Vec} AB = (B^{\mathrm{T}} \otimes A) \operatorname{Vec} I_n$.

2.4.19 KRONECKER SUMS

If the matrix $A = (a_{ij})$ has size $n \times n$ and matrix $B = (b_{ij})$ has size $m \times m$, then the *Kronecker sum* of these matrices, denoted $A \oplus B$, is defined[1] as

$$A \oplus B = A \otimes I_m + I_n \otimes B. \tag{2.4.36}$$

The Kronecker sum has the following properties:

1. If A has eigenvalues $\{\lambda_i\}$ and B has eigenvalues $\{\mu_j\}$, then $A \oplus B$ has eigenvalues $\{\lambda_i + \mu_j\}$.
2. The matrix equation $AX + XB = C$ may be equivalently written as $(B^{\mathrm{T}} \oplus A) \operatorname{Vec} X = \operatorname{Vec} C$, where Vec is defined in Section 2.4.18.
3. $e^{A\oplus B} = e^A \otimes e^B$.

[1] Note that $A \oplus B$ is also used to denote the $(m+n) \times (m+n)$ matrix $\begin{bmatrix}A & 0 \\ 0 & B\end{bmatrix}$.

2.4.20 KRONECKER PRODUCTS

If the matrix $A = (a_{ij})$ has size $m \times n$, and the matrix $B = (b_{ij})$ has size $r \times s$, then the *Kronecker product* (sometimes called the *tensor product*) of these matrices, denoted $A \otimes B$, is defined as the partitioned matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \vdots & \vdots & & \vdots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{bmatrix}. \tag{2.4.37}$$

Hence, the matrix $A \otimes B$ has size $mr \times ns$.

EXAMPLE If $A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ and $B = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, then

$$A \otimes B = \begin{bmatrix} a_{11}B & a_{12}B \\ a_{21}B & a_{22}B \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{12}b_{11} & a_{12}b_{12} \\ a_{11}b_{21} & a_{11}b_{22} & a_{12}b_{21} & a_{12}b_{22} \\ a_{21}b_{11} & a_{21}b_{12} & a_{22}b_{11} & a_{22}b_{12} \\ a_{21}b_{21} & a_{21}b_{22} & a_{22}b_{21} & a_{22}b_{22} \end{bmatrix}. \tag{2.4.38}$$

The Kronecker product has the following properties:

1. If **z** and **w** are vectors of appropriate dimensions, then $A\mathbf{z} \otimes B\mathbf{w} = (A \otimes B)(\mathbf{z} \otimes \mathbf{w})$.
2. If α is a scalar, then $(\alpha A) \otimes B = A \otimes (\alpha B) = \alpha(A \otimes B)$
3. The Kronecker product is distributive with respect to addition:
   (a) $(A + B) \otimes C = A \otimes C + B \otimes C$, and
   (b) $A \otimes (B + C) = A \otimes B + A \otimes C$.
4. The Kronecker product is associative: $A \otimes (B \otimes C) = (A \otimes B) \otimes C$.
5. $(A \otimes B)^{\mathrm{T}} = A^{\mathrm{T}} \otimes B^{\mathrm{T}}$.
6. The *mixed product rule*: If the dimensions of the matrices are such that the following expressions exist, then $(A \otimes B)(C \otimes D) = AC \otimes BD$.
7. If the inverses exist, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$.
8. If $\{\lambda_i\}$ and $\{\mathbf{x}_i\}$ are the eigenvalues and the corresponding eigenvectors for A, and $\{\mu_j\}$ and $\{\mathbf{y}_j\}$ are the eigenvalues and the corresponding eigenvectors for B, then $A \otimes B$ has eigenvalues $\{\lambda_i \mu_j\}$ with corresponding eigenvectors $\{\mathbf{x}_i \otimes \mathbf{y}_j\}$.
9. If matrix A has size $n \times n$ and B has size $m \times m$, then $\det(A \otimes B) = (\det A)^m (\det B)^n$.
10. If f is an analytic matrix function and A has size $n \times n$, then
    (a) $f(I_n \otimes A) = I_n \otimes f(A)$, and
    (b) $f(A \otimes I_n) = f(A) \otimes I_n$.
11. $\operatorname{tr}(A \otimes B) = (\operatorname{tr} A)(\operatorname{tr} B)$.
12. If A is similar to C and B is similar to D, then $A \otimes B$ is similar to $C \otimes D$.
13. If $C(t) = A(t) \otimes B(t)$, then $\frac{dC}{dt} = \frac{dA}{dt} \otimes B + A \otimes \frac{dB}{dt}$.

2.4.21 MATRIX PSEUDOSPECTRA

Let A be a complex $n \times n$ matrix. The spectra (or eigenvalues) of A, denoted $\Lambda(A)$, is the collection of values such that $zI - A$ is singular; that is $\det(zI - A) = 0$.

$$\Lambda(A) = \{z \in C \mid \det(zI - A) = 0\}$$

By convention the norm of $(zI - A)^{-1}$ is infinite at the eigenvalues.

The pseudospectra of A, denoted $\Lambda_\epsilon(A)$, is those values for which the norm of $zI - A$ is very large; the following definitions are equivalent:

$$\begin{aligned} \Lambda_\epsilon(A) &= \left\{ z \in C \mid \left|\left|(zI - A)^{-1}\right|\right| \geq \epsilon^{-1} \right\} \\ &= \{ z \in C \mid z \in \Lambda(A + E) \qquad \text{for some } E \text{ with } ||E|| \leq \epsilon \} \\ &= \{ z \in C \mid ||(A - zI)\mathbf{v}|| \leq \epsilon \quad \text{for some } \mathbf{v} \in C^n \text{ with } ||\mathbf{v}|| = 1 \} \end{aligned}$$

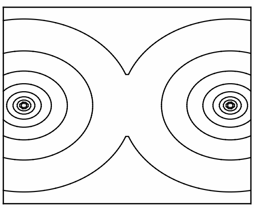
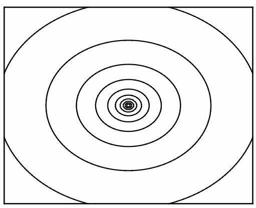
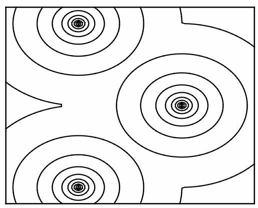
where $||\cdot||$ is a matrix norm induced by a vector norm.

A pseudospectral image shows outlines of the $\Lambda_\epsilon(A)$ regions in the complex plane. Normal matrices (e.g., those for which $A^*A = AA^*$; such as unitary, Hermitian, orthogonal, and symmetric matrices) have pseudospectral images that are qualitatively different than those of non-normal matrices. The images can be used, for example, to estimate how accurately the eigenvalues have been computed.

In the figures below $\Lambda_\epsilon(A)$ is shown for $\epsilon = 10^{-1}, 10^{-2}, \ldots, 10^{-10}$. The axes are different in each image.

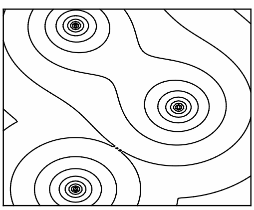
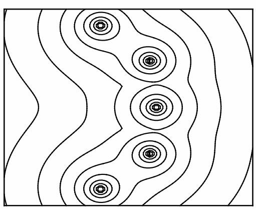
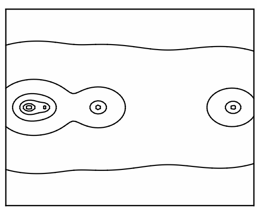
- Some normal matrices

$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ $\begin{bmatrix} 1+i & 0 & 0 \\ 0 & 1-i & 0 \\ 0 & 0 & 2 \end{bmatrix}$

- Some non-normal matrices

$\begin{bmatrix} 1+i & 0 & 1 \\ 0 & 1-i & 0 \\ 0 & 0 & 2 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ -1 & 1 & 1 & 1 & 1 \\ 0 & -1 & 1 & 1 & 1 \\ 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}$ $\begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 0 & 2 & 3 & 3 & 3 \\ 0 & 0 & 3 & 4 & 4 \\ 0 & 0 & 0 & 4 & 5 \end{bmatrix}$

2.5 ABSTRACT ALGEBRA

2.5.1 DEFINITIONS

1. A *binary operation on a set* S is a function $\star : S \times S \to S$.
2. An *algebraic structure* $(S, \star_1, \ldots, \star_n)$ consists of a non-empty set S with one or more binary operations $\star_i$ defined on S. If the operations are understood, then the binary operations need not be mentioned explicitly.
3. The *order* $|S|$ of an algebraic structure S is the number of elements in S.
4. A binary operation $\star$ on an algebraic structure $(S, \star)$ may have the following properties:
   (a) *Associative*: $a \star (b \star c) = (a \star b) \star c$ for all $a, b, c \in S$.
   (b) *Identity*: there exists an element $e \in S$ (*identity element* of S) such that $e \star a = a \star e = a$ for all $a \in S$.
   (c) *Inverse*: $a^{-1} \in S$ is an *inverse* of a if $a \star a^{-1} = a^{-1} \star a = e$.
   (d) *Commutative* (or *Abelian*): if $a \star b = b \star a$ for all $a, b \in S$.
5. A *semigroup* $(S, \star)$ consists of a non-empty set S and an associative binary operation $\star$ on S.
6. A *monoid* $(S, \star)$ consists of a non-empty set S with an identity element and an associative binary operation $\star$.

2.5.1.1 Examples of semigroups and monoids

1. The sets $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ (natural numbers), $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$ (integers), $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers), and $\mathbb{C}$ (complex numbers) where $\star$ is either addition or multiplication are semigroups and monoids.
2. The set of positive integers under addition is a semigroup but not a monoid.
3. If A is any non-empty set, then the set of all functions $f : A \to A$ where $\star$ is the composition of functions is a semigroup and a monoid.
4. Given a set S, the set of all strings of elements of S, where $\star$ is the concatenation of strings, is a monoid (the identity is λ, the empty string).

2.5.2 GROUPS

1. A *group* $(G, \star)$ consists of a set G with a binary operation $\star$ defined on G such that $\star$ satisfies the associative, identity, and inverse laws. *Note:* The operation $\star$ is often written as $+$ (an *additive group*) or as $\cdot$ or $\times$ (a *multiplicative group*).
   (a) If $+$ is used, the identity is written 0 and the inverse of a is written $-a$. Usually, in this case, the group is commutative. The following notation $na = \underbrace{a + \ldots + a}_{n \text{ times}}$ is then used.
   (b) If multiplicative notation is used, $a \star b$ is often written ab, the identity is often written 1, and the inverse of a is written a^{-1}.

2. The *order* of $a \in G$ is the smallest positive integer n such that $a^n = 1$ where $a^n = a \cdot a \cdots a$ (n times) (or $a + a + \cdots + a = 0$ if G is written additively). If there is no such integer, the element has *infinite order*. In a finite group of order n, each element has some order k (depending on the particular element) and it must be that k divides n.
3. $(H, \star)$ is a subgroup of $(G, \star)$ if $H \subseteq G$ and $(H, \star)$ is a group (using the same binary operation as in $(G, \star)$).
4. The *cyclic subgroup* $\langle a \rangle$ generated by $a \in G$ is the subgroup $\{a^n \mid n \in \mathbb{Z}\} = \{\ldots, a^{-2} = (a^{-1})^2, a^{-1}, a^0 = e, a, a^2, \ldots\}$. The element a is a *generator* of $\langle a \rangle$. A group G is *cyclic* if there is $a \in G$ such that $G = \langle a \rangle$.
5. If H is a subgroup of a group G, then a *left* [*right*] *coset* of H in G is the set $aH = \{ah \mid h \in H\}$ $[Ha = \{ha \mid h \in H\}]$.
6. A *normal subgroup* of a group G is a subgroup H such that $aH = Ha$ for all $a \in G$.
7. A *simple group* is a group $G \neq \{e\}$ with only G and $\{e\}$ as normal subgroups.
8. If H is a normal subgroup of G, then the *quotient group* (or *factor group*) *of G modulo H* is the group $G/H = \{aH \mid a \in G\}$, with binary operation $aH \cdot bH = (ab)H$.
9. A finite group G is *solvable* if there is a sequence of subgroups $G_1 = G, G_2, \ldots, G_{k-1}$, with $G_k = \{e\}$, such that each G_{i+1} is a normal subgroup of G_i and G_i/G_{i+1} is Abelian.

2.5.2.1 Facts about groups

1. The identity element is unique.
2. Each element has exactly one inverse.
3. Each of the equations $a \star x = b$ and $x \star a = b$ has exactly one solution, $x = a^{-1} \star b$ and $x = b \star a^{-1}$.
4. $(a^{-1})^{-1} = a$.
5. $(a \star b)^{-1} = b^{-1} \star a^{-1}$.
6. The *left* (respectively *right*) *cancellation law* holds in all groups: If $a \star b = a \star c$ then $b = c$ (respectively, if $b \star a = c \star a$ then $b = c$).
7. *Lagrange's theorem*: If G is a finite group and H is a subgroup of G, then the order of H divides the order of G.
8. Every group of prime order is Abelian and hence simple.
9. Every cyclic group is Abelian.
10. Every Abelian group is solvable.
11. *Feit–Thompson theorem*: All groups of odd order are solvable. Hence, all finite non-Abelian simple groups have even order.
12. Finite simple groups are of the following types:
    (a) $\mathbb{Z}_p$ (p prime)
    (b) A group of Lie type
    (c) A_n ($n \geq 5$)
    (d) Sporadic groups (see table on page 120)

2.5.2.2 Examples of groups

1. $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$, with $\star$ the addition of numbers, are additive groups.
2. For n a positive integer, $n\mathbb{Z} = \{nz \mid z \in \mathbb{Z}\}$ is an additive group.
3. $\mathbb{Q} - \{0\} = \mathbb{Q}^*$, $\mathbb{R} - \{0\} = \mathbb{R}^*$, $\mathbb{C} - \{0\} = \mathbb{C}^*$, with $\star$ the multiplication of numbers, are multiplicative groups.
4. $\mathbb{Z}_n = \mathbb{Z}/n\mathbb{Z} = \{0, 1, 2, \dots, n-1\}$ is a group where $\star$ is addition modulo n.
5. $\mathbb{Z}_n^* = \{k \mid k \in \mathbb{Z}_n, k$ has a multiplicative inverse (under multiplication modulo n) in $\mathbb{Z}_n\}$ is a group under multiplication modulo n. If p is prime, $\mathbb{Z}_p^*$ is cyclic. If p is prime and $a \in \mathbb{Z}_p^*$ has order (index) $p-1$, then a is a *primitive root modulo* p.
6. If $(G_1, \star_1), (G_2, \star_2), \dots, (G_n, \star_n)$ are groups, the (*direct*) *product group* is $(G_1 \times G_2 \times \cdots \times G_n, \star) = \{(a_1, a_2, \dots, a_n) \mid a_i \in G_i, i = 1, 2, \dots, n\}$ where $\star$ is defined by
$$(a_1, a_2, \dots, a_n) \star (b_1, b_2, \dots, b_n) = (a_1 \star_1 b_1, a_2 \star_2 b_2, \dots, a_n \star_n b_n).$$
7. All $m \times n$ matrices with real entries form a group under addition of matrices.
8. All $n \times n$ matrices with real entries and non-zero determinants form a group under matrix multiplication.
9. All 1–1, onto functions $f : S \to S$ (*permutations* of S), where S is any non-empty set, form a group under composition of functions. See Section 2.5.3.
   In particular, if $S = \{1, 2, 3, \dots, n\}$, the group of permutations of S is called the *symmetric group*, S_n. In S_n, each permutation can be written as a product of cycles. A *cycle* is a permutation $\sigma = (i_1\ i_2\ \cdots\ i_k)$, where $\sigma(i_1) = i_2, \sigma(i_2) = i_3, \dots, \sigma(i_k) = i_1$. Each cycle of length greater than 1 can be written as a product of transpositions (cycles of length 2). A permutation is *even* (*odd*) if it can be written as the product of an even (odd) number of transpositions. (Every permutation is either even or odd.) The set of all even permutations in S_n is a normal subgroup, A_n, of S_n. The group A_n is called the *alternating group* on n elements.
10. Given a regular n-gon, the *dihedral group* D_n is the group of all symmetries of the n-gon, that is, the group generated by the set of all rotations around the center of the n-gon through angles of $360k/n$ degrees (where $k = 0, 1, 2, \dots, n-1$), together with all reflections in lines passing through a vertex and the center of the n-gon, using composition of functions. Alternately, $D_n = \{a^i b^j \mid i = 0, 1;\ j = 0, 1, \dots, n-1;\ aba^{-1} = b^{-1}\}$.

2.5.2.3 Matrix classes that are groups

In the following, the group operation is ordinary matrix multiplication:

1.	$GL(n, \mathbb{C})$	all complex non-singular $n \times n$ matrices
2.	$GL(n, \mathbb{R})$	all real non-singular $n \times n$ matrices
3.	$O(n)$	*orthogonal group*, all $n \times n$ matrices A with $AA^{\mathrm{T}} = I$
4.	$SL(n, \mathbb{C})$	all complex $n \times n$ matrices of determinant 1, also called the *unimodular group* or the *special linear group*
5.	$SL(n, \mathbb{R})$	all real $n \times n$ matrices of determinant 1
6.	$SO(2)$	planar rotations: matrices of the form $\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$
7.	non-Abelian	matrices of the form $\begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$ with $a, b, c \in R$

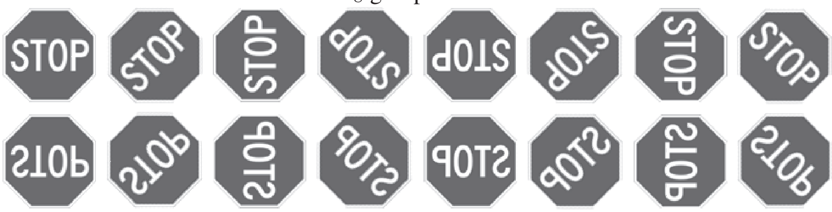
2.5.3 PERMUTATION GROUPS

Name	Symbol	Order	Definition
Symmetric group	S_p	$p!$	All permutations on $\{1, 2, \ldots, p\}$
Alternating group	A_p	$p!/2$	All even permutations on $\{1, 2, \ldots, p\}$
Cyclic group	C_p	p	Generated by $(12 \cdots p)$
Dihedral group	D_p	$2p$	Generated by $(12 \cdots p)$ and $(1p)(2\,p-1)(3\,p-2)\cdots$
Identity group	E_p	1	$(1)(2)\cdots(p)$ is the only permutation

EXAMPLE For $p = 3$ elements, the identity permutation is (123), and:

$$\begin{aligned}
A_3 &= \{(123), (231), (312)\},\\
C_3 &= \{(123), (231), (312)\},\\
D_3 &= \{(231), (213), (132), (321), (312), (123)\},\\
E_3 &= \{(123)\} \text{ and}\\
S_3 &= \{(231), (213), (132), (321), (312), (123)\}.
\end{aligned}$$

EXAMPLE The 16 elements in the D_8 group can be illustrated as



2.5.3.1 Creating new permutation groups

Assume A has permutations $\{X_i\}$, order n, degree d, B has permutations $\{Y_j\}$, order m, degree e, and $C = C(A, B)$ have permutations $\{W_k\}$, order p, degree f.

Name	Definition	Permutation	Order	Degree
Sum	$C = A + B$	$W = X \cup Y$	$p = mn$	$f = d + e$
Product	$C = A \times B$	$W = X \times Y$	$p = mn$	$f = de$
Composition	$C = A[B]$	$W = X \times Y$	$p = mn^d$	$f = de$
Power	$C = B^A$	$W = Y^X$	$p = mn$	$f = e^d$

2.5.3.2 Polya theory

Let π be a permutation. Define Inv (π) to be the number of invariant elements (i.e., mapped to themselves) in π. Define cyc (π) as the number of cycles in π. Suppose π has b_1 cycles of length 1, b_2 cycles of length 2, ..., b_k cycles of length k in its unique cycle decomposition. Then π can be encoded as the expression $x_1^{b_1} x_2^{b_2} \cdots x_k^{b_k}$. Summing these expressions for all permutations in the group G, and normalizing by the number of elements in G results in the *cycle index* of the group G:

$$P_G(x_1, x_2, \ldots, x_l) = \frac{1}{|G|} \sum_{\pi \in G} \left(x_1^{b_1} x_2^{b_2} \cdots x_k^{b_k} \right). \tag{2.5.1}$$

1. *Burnside's lemma*: Let G be a group of permutations of a set A, and let S be the equivalence relation on A induced by G. Then the number of equivalence classes in A is given by $\dfrac{1}{|G|} \displaystyle\sum_{\pi \in G} \operatorname{Inv}(\pi)$.
2. *Special case of Polya's theorem*: Let R be an m element set of colors. Let G be a group of permutations $\{\pi_1, \pi_2, \ldots\}$ of the set A. Let $C(A, R)$ be the set of colorings of the elements of A using colors in R. Then the number of distinct colorings in $C(A, R)$ is given by

$$\frac{1}{|G|} \left[m^{\operatorname{cyc}(\pi_1)} + m^{\operatorname{cyc}(\pi_2)} + \ldots \right].$$

3. *Polya's theorem*: Let G be a group of permutations on a set A with cycle index $P_G(x_1, x_2, \ldots, x_k)$. Let $C(A, R)$ be the collection of all colorings of A using colors in R. If w is a weight assignment on R, then the pattern inventory of colorings in $C(A, R)$ is given by

$$P_G \left(\sum_{r \in R} w(r), \sum_{r \in R} w^2(r), \cdots \sum_{r \in R} w^k(r) \right).$$

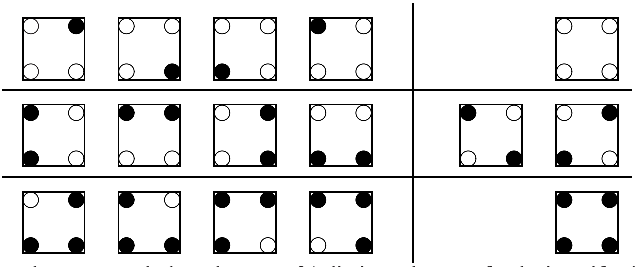
EXAMPLES

1. Consider necklaces constructed of $2k$ beads. Since a necklace can be flipped over, the appropriate permutation group is $G = \{\pi_1, \pi_2\}$ with $\pi_1 = (1)(2)\ldots(2k)$ and $\pi_2 = \begin{pmatrix}1 & 2k\end{pmatrix} \begin{pmatrix}2 & 2k-1\end{pmatrix} \begin{pmatrix}3 & 2k-2\end{pmatrix} \ldots \begin{pmatrix}k & k+1\end{pmatrix}$. Hence, $\operatorname{cyc}(\pi_1) = 2k$, $\operatorname{cyc}(\pi_2) = k$, and the cycle index is $P_G(x_1, x_2) = \left(x_1^{2k} + x_2^k\right)/2$. Using r colors, the number of distinct necklaces is $(r^{2k} + r^k)/2$.

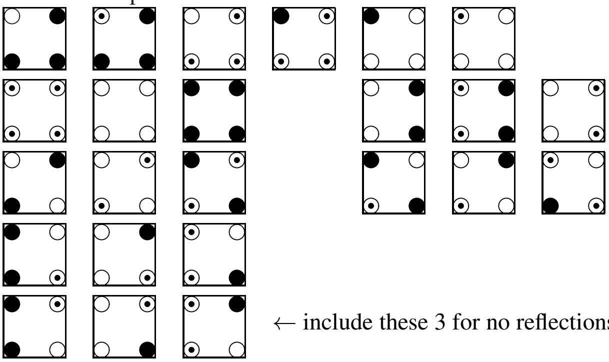
   For a 4-bead necklace ($k = 2$) using $r = 2$ colors (say $w_1 = b$ for "black" and $w_2 = w$ for "white"), the $(2^4 + 2^2)/2 = 10$ different necklaces are $\{bbbb\}$, $\{bbbw\}$, $\{bbwb\}$, $\{bbww\}$, $\{bwbw\}$, $\{bwwb\}$, $\{wbbw\}$, $\{bwww\}$, $\{wbww\}$, and $\{wwww\}$. The pattern inventory of colorings, $P_G(\sum_i w_i, \sum_i w_i^2) = \left((b+w)^4 + (b^2+w^2)^2\right)/2 = b^4 + 2b^3w + 4b^2w^2 + 2bw^3 + w^4$, tells how many colorings of each type there are.

2. Consider coloring the corners of a square. If the squares can be rotated, but not reflected, then the number of distinct colorings, using k colors, is $\frac{1}{4}\left(k^4 + k^2 + 2k\right)$. If the squares can be rotated and reflected, then the number of distinct colorings, using k colors, is $\frac{1}{8}\left(k^4 + 2k^3 + 3k^2 + 2k\right)$.

   (a) If $k = 2$ colors are used, then there are 6 distinct classes of colorings whether reflections are allowed, or not. These classes are the same in both cases. The 16 colorings of a square with 2 colors form 6 distinct classes as shown:



(b) If $k = 3$ colors are used, then there are 21 distinct classes of colorings if reflections are allowed, and 24 distinct classes of colorings if reflections are not allowed. Shown below are representative elements of each of these classes:



3. Number of distinct corner colorings of regular polygons using rotations and reflections, or rotations only, with no more than k colors:

object	rotations & reflections $k=2$	$k=3$	$k=4$	rotations only $k=2$	$k=3$	$k=4$
triangle	4	10	20	4	11	24
square	6	21	55	6	24	70
pentagon	8	39	136	8	51	208
hexagon	13	92	430	14	130	700

4. Coloring regular 2- and 3-dimensional objects with no more than k colors:

(a) tetrahedron

i. corners of a tetrahedron $\frac{1}{12}\left(k^6 + 3k^4 + 8k^2\right)$
ii. edges of a tetrahedron $\frac{1}{12}\left(k^6 + 3k^4 + 8k^2\right)$
iii. faces of a tetrahedron $\frac{1}{12}\left(k^4 + 11k^2\right)$

(b) cube

i. corners of a cube $\frac{1}{24}\left(k^8 + 17k^4 + 6k^2\right)$
ii. edges of a cube $\frac{1}{24}\left(k^{12} + 6k^7 + 3k^6 + 8k^4 + 6k^3\right)$
iii. faces of a cube $\frac{1}{24}\left(k^6 + 3k^4 + 12k^3 + 8k^2\right)$

(c) corners of a triangle

i. with rotations $\frac{1}{3}\left(k^3 + 2k\right)$
ii. with rotations and reflections $\frac{1}{6}\left(k^3 + 3k^2 + 2k\right)$

(d) corners of a square

i. with rotations $\frac{1}{4}\left(k^4+k^2+2k\right)$

ii. with rotations and reflections $\frac{1}{8}\left(k^4+2k^3+3k^2+2k\right)$

(e) corners of an n-gon

i. with rotations $\frac{1}{n}\sum_{d|n}\phi(d)k^{\frac{n}{d}}$

ii. with rotations and reflections (n odd) $\frac{1}{2n}\sum_{d|n}\phi(d)k^{\frac{n}{d}}+\frac{1}{2}k^{\frac{n+1}{2}}$

iii. with rotations and reflections (n even)

$$\frac{1}{2n}\sum_{d|n}\phi(d)k^{\frac{n}{d}}a+\frac{1}{4}\left(k^{\frac{n}{2}}+k^{\frac{n+2}{2}}\right)$$

5. The cycle index $P(x_1, x_2, \dots)$ and number of black-white colorings of regular objects under all permutations:

(a) corners of a triangle

i. cycle index $\frac{1}{6}\left(x_1^3+3x_1x_2+2x_3\right)$

ii. pattern inventory $1b^3+1b^2w+1bw^2+1w^3$

(b) corners of a square

i. cycle index $\frac{1}{8}\left(x_1^4+2x_2^2x_2+3x_2^2+2x_4\right)$

ii. pattern inventory $1b^4+1b^3w+2b^2w^2+1bw^3+1w^4$

(c) corners of a pentagon

i. cycle index $\frac{1}{10}\left(x_1^5+4x_5+5x_1x_2^2\right)$

ii. pattern inventory $1b^5+1b^4w+2b^3w^2+2b^2w^3+1bw^4+1w^5$

(d) corners of a cube

i. cycle index $\frac{1}{24}\left(x_1^8+6x_4^2+9x_2^4+8x_1^2x_3^2\right)$

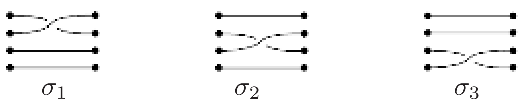
ii. pattern inventory

$$b^8+b^7w+3b^6w^2+3b^5w^3+7b^4w^4+3b^3w^5+3b^2w^6+bw^7+w^8$$

Note that the pattern inventory for the black-white colorings is given by $P((b+w),(b^2+w^2),(b^3+w^3),\dots)$.

2.5.4 BRAID GROUPS

A braid group represents how threads can be braided together. Consider the case $n=4$; given 4 threads we can switch the order of two adjacent threads in 3 ways:



These operations can be combined via composition. We note the properties:

- $\sigma_1\sigma_3=\sigma_3\sigma_1$
- $\sigma_1\sigma_2\sigma_1=\sigma_2\sigma_1\sigma_2$
- $\sigma_2\sigma_3\sigma_2=\sigma_3\sigma_2\sigma_3$

Abstracting these relations, the braid group on n strands (B_n) consists of the elements (or generators) $\{\sigma_1,\dots,\sigma_{n-1}\}$ with the following relations:

$$\sigma_i\sigma_j=\sigma_j\sigma_i \quad \text{for} \quad |i-j|\geq 2$$

$$\sigma_i\sigma_{i+1}\sigma_i=\sigma_{i+1}\sigma_i\sigma_{i+1} \quad \text{for} \quad 1\leq i\leq n-2$$

2.5.5 RINGS

2.5.5.1 Definitions

1. A *ring* $(R, +, \cdot)$ consists of a non-empty set R and two binary operations, $+$ and $\cdot$, such that $(R, +)$ is an Abelian group, the operation $\cdot$ is associative, and the *left distributive law* $a(b + c) = (ab) + (ac)$ and the *right distributive law* $(a + b)c = (ac) + (bc)$ hold for all a, b, $c \in R$.
2. A subset S of a ring R is a *subring* of R if S is a ring using the same operations used in R with the same unit.
3. A ring R is a *commutative* ring if the multiplication operation is commutative: $ab = ba$ for all $a, b \in R$.
4. A ring R is a *ring with unity* if there is an element 1 (called *unity*) such that $a1 = 1a = a$ for all $a \in R$.
5. A *unit* in a ring with unity is an element a with a multiplicative inverse a^{-1} (that is, $aa^{-1} = a^{-1}a = 1$).
6. If $a \neq 0$, $b \neq 0$, and $ab = 0$, then a is a *left divisor of zero* and b is a *right divisor of zero*.
7. A subset I of a ring $(R, +, \cdot)$ is a (two-sided) *ideal* of R if $(I, +)$ is a subgroup of $(R, +)$ and I is closed under left and right multiplication by elements of R (if $x \in I$ and $r \in R$, then $rx \in I$ and $xr \in I$).
8. An ideal $I \subseteq R$ is
   - (a) *Proper*: if $I \neq \{0\}$ and $I \neq R$
   - (b) *Maximal*: if I is proper and if there is no proper ideal properly containing I
   - (c) *Prime*: if $ab \in I$ implies that a or $b \in I$
   - (d) *Principal*: if I is the intersection of all ideals containing an element $a \in R$
9. If I is an ideal in a ring R, then a *coset* is a set $r + I = \{r + a \mid a \in I\}$.
10. If I is an ideal in a ring R, then the *quotient ring* is the ring $R/I = \{r + I \mid r \in R\}$, where $(r+I)+(s+I) = (r+s)+I$ and $(r+I)(s+I) = (rs)+I$.
11. An *integral domain* $(R, +, \cdot)$ is a commutative ring with unity such that cancellations hold: if $ab = ac$ then $b = c$ (respectively, if $ba = ca$ then $b = c$) for all a, b, $c \in R$, where $a \neq 0$. (Equivalently, an integral domain is a commutative ring with unity that has no divisors of zero.)
12. If R is an integral domain, then a non-zero element $r \in R$ that is not a unit is *irreducible* if $r = ab$ implies that either a or b is a unit.
13. If R is an integral domain, a non-zero element $r \in R$ that is not a unit is a *prime* if, whenever $r|ab$, then either $r|a$ or $r|b$ ($x|y$ means that there is an element $z \in R$ such that $y = zx$.).
14. A *unique factorization domain* (UFD) is an integral domain such that every non-zero element that is not a unit can be written uniquely as the product of irreducible elements (except for factors that are units and except for the order in which the factor appears).
15. A *principal ideal domain* (PID) is an integral domain in which every ideal is a principal ideal.

16. A *division ring* is a ring in which every non-zero element has a multiplicative inverse (that is, every non-zero element is a unit). (Equivalently, a division ring is a ring in which the non-zero elements form a multiplicative group.) A non-commutative division ring is called a *skew field*.

2.5.5.2 Facts about rings

1. The set of all units of a ring is a group under the ring multiplication operation.
2. Every principal ideal domain is a unique factorization domain.
3. If R is a commutative ring with unity, then
   (a) every maximal ideal is a prime ideal.
   (b) R is a field if and only if the only ideals of R are R and $\{0\}$.
   (c) I is a maximal ideal if and only if R/I is a field.
4. If R is a commutative ring with unity and $I \neq R$ is an ideal, then R/I is an integral domain if and only if I is a prime ideal.
5. If $f(x) \in F[x]$ (where F is a field) and the ideal generated by $f(x)$ is not $\{0\}$, then the ideal is maximal if and only if $f(x)$ is irreducible over F.
6. There are exactly four normed division rings, they have dimensions 1, 2, 4, and 8. They are the real numbers, the complex numbers, the quaternions, and the octonions. The quaternions are non-commutative and the octonions are non-associative.

2.5.5.3 Examples of rings

1. $\mathbb{Z}$ (integers), $\mathbb{Q}$ (rational numbers), $\mathbb{R}$ (real numbers), and $\mathbb{C}$ (complex numbers) are rings, with ordinary addition and multiplication of numbers.
2. $\mathbb{Z}_n$ is a ring, with addition and multiplication modulo n.
3. If $\sqrt{n}$ is not an integer, then $\mathbb{Z}[\sqrt{n}] = \{a + b\sqrt{n} \mid a, b \in \mathbb{Z}\}$, where $(a + b\sqrt{n}) + (c + d\sqrt{n}) = (a + c) + (b + d)\sqrt{n}$ and $(a + b\sqrt{n})(c + d\sqrt{n}) = (ac + nbd) + (ad + bc)\sqrt{n}$ is a ring.
4. The set of *Gaussian integers* $\mathbb{Z}[i] = \{a + bi \mid a, b \in \mathbb{Z}\}$ is a ring, with the usual definitions of addition and multiplication of complex numbers.
5. The *polynomial ring* in one variable over a ring R is the ring $R[x] = \{a_n x^n + \cdots + a_1 x + a_0 \mid a_i \in R;\ i = 0, 1, \ldots, n;\ n \in \mathcal{N}\}$. (Elements of $R[x]$ are added and multiplied using the usual rules for addition and multiplication of polynomials.) The *degree* of a polynomial $a_n x^n + \cdots + a_1 x + a_0$ with $a_n \neq 0$ is n. A polynomial is *monic* if $a_n = 1$. A polynomial $f(x)$ is *irreducible over* R if $f(x)$ cannot be factored as a product of polynomials in $R[x]$ of degree less than the degree of $f(x)$. A monic irreducible polynomial $f(x)$ of degree k in $\mathbb{Z}_p[x]$ (p prime) is *primitive* if the order of x in $\mathbb{Z}_p[x]/(f(x))$ is $p^k - 1$, where $(f(x)) = \{f(x)g(x) \mid g(x) \in \mathbb{Z}_p[x]\}$ (the ideal generated by $f(x)$). For example, the polynomial $x^2 + 1$ is
   (a) Irreducible in $\mathbb{R}[x]$ because $x^2 + 1$ has no real root
   (b) Reducible in $\mathbb{C}[x]$ because $x^2 + 1 = (x - i)(x + i)$
   (c) Reducible in $\mathbb{Z}_2[x]$ because $x^2 + 1 = (x + 1)^2$
   (d) Reducible in $\mathbb{Z}_5[x]$ because $x^2 + 1 = (x + 2)(x + 3)$

6. The *division ring of quaternions* is the ring $(\{a + bi + cj + dk \mid a, b, c, d \in \mathbb{R}\}, +, \cdot)$, where operations are carried out using the rules for polynomial addition and multiplication and the defining relations for the quaternion group Q.
7. Every octonion is a real linear combination of the unit octonions $\{1, e_1, e_2, e_3, e_4, e_5, e_6, e_7\}$. Their properties include: (a) $e_i^2 = -1$; (b) $e_i e_j = -e_j e_i$ when $i \neq j$; (c) the index doubling identity: $e_i e_j = e_k \implies e_{2i} e_{2j} = e_{2k}$; and (d) the index cycling identity: $e_i e_j = e_k \implies e_{i+1} e_{j+1} = e_{k+1}$ where the indices are computed modulo 7. The full multiplication table is as follows:

	1	e_1	e_2	e_3	e_4	e_5	e_6	e_7
1	1	e_1	e_2	e_3	e_4	e_5	e_6	e_7
e_1	e_1	-1	e_4	e_7	$-e_2$	e_6	$-e_5$	$-e_3$
e_2	e_2	$-e_4$	-1	e_5	e_1	$-e_3$	e_7	$-e_6$
e_3	e_3	$-e_7$	$-e_5$	-1	e_6	e_2	$-e_4$	e_1
e_4	e_4	e_2	$-e_1$	$-e_6$	-1	e_7	e_3	$-e_5$
e_5	e_5	$-e_6$	e_3	$-e_2$	$-e_7$	-1	e_1	e_4
e_6	e_6	e_5	$-e_7$	e_4	$-e_3$	$-e_1$	-1	e_2
e_7	e_7	e_3	e_6	$-e_1$	e_5	$-e_4$	$-e_2$	-1

8. The following table gives examples of rings with additional properties:

Ring	Commutative ring with unity	Integral domain	Principal ideal domain	Euclidean domain	Division ring	Field
$\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$	yes	yes	yes	yes	yes	yes
$\mathbb{Z}$	yes	yes	yes	yes	no	no
$\mathbb{Z}_p$ (p prime)	yes	yes	yes	yes	yes	yes
$\mathbb{Z}_n$ (n composite)	yes	no	no	no	no	no
$\mathbb{Z}[x]$	yes	yes	no	no	no	no
$\mathcal{M}_{n\times n}$	no	no	no	no	no	no

2.5.6 FIELDS

2.5.6.1 Definitions

1. A *field* $(F, +, \cdot)$ is a commutative ring with unity such that each non-zero element of F has a multiplicative inverse (equivalently, a field is a commutative division ring).
2. A *finite field* is a field that contains a finite number of elements.
3. The *characteristic* of a field (or a ring) is the smallest positive integer n such that $1 + 1 + \cdots + 1 = 0$ (n summands). If no such n exists, the field has characteristic 0 (or characteristic ∞).
4. Field K is an *extension field* of the field F if F is a subfield of K (i.e., $F \subseteq K$, and F is a field using the same operations used in K).

2.5.6.2 Examples of fields

1. $\mathbb{Q}$, $\mathbb{R}$, and $\mathbb{C}$ with ordinary addition and multiplication are fields.
2. $\mathbb{Z}_p$ (p a prime) is a field under addition and multiplication modulo p.
3. $F[x]/(f(x))$ is a field, provided that F is a field and $f(x)$ is a non-constant polynomial irreducible in $F[x]$.

2.5.7 QUADRATIC FIELDS

2.5.7.1 Definitions

1. A complex number is an *algebraic integer* if it is a root of a polynomial with integer coefficients that has a leading coefficient of 1.
2. If d is a square-free integer, then $\mathbb{Q}(\sqrt{d}) = \{a + b\sqrt{d}\}$, where a and b are rational numbers, is called a *quadratic field*. If $d > 0$ then $\mathbb{Q}(\sqrt{d})$ is a *real quadratic field*; if $d < 0$ then $\mathbb{Q}(\sqrt{d})$ is an *imaginary quadratic field*.
3. The *integers* of an algebraic number field are the algebraic integers that belong to this number field.
4. If $\{\alpha, \beta, \gamma\}$ are integers in $\mathbb{Q}(\sqrt{d})$ such that $\alpha\gamma = \beta$, then we say that α *divides* β; written $\alpha|\beta$.
5. An integer ϵ in $\mathbb{Q}(\sqrt{d})$ is a *unit* if it divides 1.
6. If $\alpha = a + b\sqrt{d}$ then
   (a) the *conjugate* of α is $\overline{\alpha} = a - b\sqrt{d}$.
   (b) the *norm* of α is $N(\alpha) = \alpha\overline{\alpha} = a^2 - db^2$.
7. If α is an integer of $\mathbb{Q}(\sqrt{d})$ and if ϵ is a unit of $\mathbb{Q}(\sqrt{d})$, then the number $\epsilon\alpha$ is an *associate* of α. A *prime* in $\mathbb{Q}(\sqrt{d})$ is an integer of $\mathbb{Q}(\sqrt{d})$ that is only divisible by the units and its associates.
8. A quadratic field $\mathbb{Q}(\sqrt{d})$ is a *Euclidean field* if, given integers α and β in $\mathbb{Q}(\sqrt{d})$ with $\beta \neq 0$, there are integers γ and δ in $\mathbb{Q}(\sqrt{d})$ such that $\alpha = \gamma\beta + \delta$ and $|N(\delta)| < |N(\beta)|$.
9. A quadratic field $\mathbb{Q}(\sqrt{d})$ has the *unique factorization property* if, whenever α is a non-zero, non-unit, integer in $\mathbb{Q}(\sqrt{d})$ with $\alpha = \epsilon\pi_1\pi_2\cdots\pi_r = \epsilon'\pi_1'\pi_2'\cdots\pi_s'$ where ϵ and ϵ' are units, then $r = s$ and the primes π_i and π_j' can be paired off into pairs of associates.

2.5.7.2 Facts about quadratic fields

1. The integers of $\mathbb{Q}(\sqrt{d})$ are of the form
   (a) $a + b\sqrt{d}$, with a and b integers, if $d \equiv 2 \mod 4$ or $d \equiv 3 \mod 4$.
   (b) $a + b\left(\frac{\sqrt{d}-1}{2}\right)$, with a and b integers, if $d \equiv 1 \mod 4$.
2. Norms are positive in imaginary quadratic fields, but not necessarily positive in real quadratic fields. It is always true that $N(\alpha\beta) = N(\alpha)N(\beta)$.
3. If α is an integer in $\mathbb{Q}(\sqrt{d})$ and $N(\alpha)$ is an integer that is prime, then α is prime.
4. The number of units in $\mathbb{Q}(\sqrt{d})$ is as follows:
   (a) If $d = -3$, there are 6 units: ± 1, $\pm\frac{-1+\sqrt{-3}}{2}$, and $\pm\frac{-1-\sqrt{-3}}{2}$.
   (b) If $d = -1$, there are 4 units: ± 1 and $\pm i$.
   (c) If $d < 0$ and $d \neq -1$ and $d \neq -3$ there are 2 units: ± 1.
   (d) If $d > 0$ there are infinitely many units. There is a fundamental unit, ϵ_0, such that all other units have the form $\pm\epsilon_0^n$ where n is an integer.
5. The quadratic field $\mathbb{Q}(\sqrt{d})$ is Euclidean if and only if d is one of the following: $-11, -7, -3, -2, -1, 2, 3, 5, 6, 7, 11, 13, 17, 19, 21, 29, 33, 37, 41, 57, 73$.

6. If $d < 0$ then the imaginary quadratic field $\mathbb{Q}\left(\sqrt{d}\right)$ has the unique factorization property if and only if d is one of the following: -1, -2, -3, -7, -11, -19, -43, -163.
7. Of the 60 real quadratic fields $\mathbb{Q}(\sqrt{d})$ with $2 \leq d \leq 100$, exactly 38 of them have the unique factorization property: $d = 2$, 3, 5, 6, 7, 11, 13, 14, 17, 19, 21, 22, 23, 29, 31, 33, 37, 38, 41, 43, 46, 47, 53, 57, 59, 61, 62, 67, 69, 71, 73, 77, 83, 86, 89, 93, 94, and 97.

2.5.7.3 Examples of quadratic fields

1. The algebraic integers of $\mathbb{Q}(\sqrt{-1})$ are of the form $a + bi$ where a and b are integers; they are called the *Gaussian integers*.
2. The number $1+\sqrt{2}$ is a fundamental unit of $\mathbb{Q}(\sqrt{2})$. Hence, all units in $\mathbb{Q}(\sqrt{2})$ have the form $\pm(1+\sqrt{2})^n$ for $n = 0, \pm1, \pm2, \ldots$.
3. The field $\mathbb{Q}\left(\sqrt{-5}\right)$ is not a unique factorization domain. This is illustrated by $6 = 2 \cdot 3 = (1+\sqrt{-5}) \cdot (1-\sqrt{-5})$, yet each of $\{2, 3, 1+\sqrt{-5}, 1-\sqrt{-5}\}$ is prime in this field.
4. The field $\mathbb{Q}\left(\sqrt{10}\right)$ is not a unique factorization domain. This is illustrated by $6 = 2 \cdot 3 = (4+\sqrt{10}) \cdot (4-\sqrt{10})$, yet each of $\{2, 3, 4+\sqrt{10}, 4-\sqrt{10}\}$ is prime in this field.

2.5.8 FINITE FIELDS

2.5.8.1 Facts about finite fields

1. If p is prime, then the ring $\mathbb{Z}_p$ is a finite field.
2. If p is prime and n is a positive integer, then there is exactly one field (up to isomorphism) with p^n elements. This field is denoted $GF(p^n)$ or F_{p^n} and is called a *Galois field*. (See the table on page 127.)
3. For F a finite field, there is a prime p and a positive integer n such that F has p^n elements. The prime number p is the characteristic of F. The field F is a *finite extension* of $\mathbb{Z}_p$, that is, F is a finite dimensional vector space over $\mathbb{Z}_p$.
4. If F is a finite field, then the set of non-zero elements of F under multiplication is a cyclic group. A generator of this group is a *primitive element*.
5. There are $\phi(p^n-1)/n$ primitive polynomials of degree n ($n > 1$) over $GF(p)$, where ϕ is the Euler ϕ-function. (See table on page 46.)
6. There are $(\sum_{j|k} \mu(k/j)p^{nj})/k$ irreducible polynomials of degree k over $GF(p^n)$, where μ is the Möbius function.
7. If F is a finite field where $|F| = k$ and $p(x)$ is a polynomial of degree n irreducible over F, then the field $F[x]/(p(x))$ has order k^n. If α is a root of $p(x) \in F[x]$ of degree $n \geq 1$, then $F[x]/(p(x)) = \{c_{n-1}\alpha^{n-1} + \cdots + c_1\alpha + c_0 \mid c_i \in F \text{ for all } i\}$.
8. When p is a prime, $GF(p^n)$ can be viewed as a vector space of dimension n over F_p. A basis of F_{p^n} of the form $\{\alpha, \alpha^p, \alpha^{p^2}, \ldots, \alpha^{p^{n-1}}\}$ is called a *normal basis*. If α is a primitive element of F_{p^n}, then the basis is said to be a *primitive normal basis*. Such an α satisfies a primitive normal polynomial of degree n over F_p.

2.5.9 HOMOMORPHISMS AND ISOMORPHISMS

2.5.9.1 Definitions

1. A *group homomorphism* from group G_1 to group G_2 is a function $\varphi : G_1 \to G_2$ such that $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in G_1$. *Note*: $a\varphi$ is often written instead of $\varphi(a)$.
2. A *character* of a group G is a group homomorphism $\chi : G \to \mathbb{C}^*$ (non-zero complex numbers under multiplication). (See table on page 128.)
3. A *ring homomorphism* from ring R_1 to ring R_2 is a function $\varphi : R_1 \to R_2$ such that $\varphi(a + b) = \varphi(a) + \varphi(b)$ and $\varphi(ab) = \varphi(a)\varphi(b)$ for all $a, b \in R_1$.
4. An *isomorphism* from group (ring) S_1 to group (ring) S_2 is a group (ring) homomorphism $\varphi : S_1 \to S_2$ that is 1-1 and onto S_2. If an isomorphism exists, then S_1 is said to be *isomorphic* to S_2. Write $S_1 \cong S_2$. (See the table on page 119 for numbers of non-isomorphic groups and the table on page 121 for examples of groups of orders less than 16.)
5. An *automorphism* of S is an isomorphism $\varphi : S \to S$.
6. The *kernel* of a group homomorphism $\varphi : G_1 \to G_2$ is $\varphi^{-1}(e) = \{g \in G_1 \mid \varphi(g) = e\}$. The *kernel* of a ring homomorphism $\varphi : R_1 \to R_2$ is $\varphi^{-1}(0) = \{r \in R_1 \mid \varphi(r) = 0\}$.

2.5.9.2 Facts about homomorphisms and isomorphisms

1. If $\varphi : G_1 \to G_2$ is a group homomorphism, then $\varphi(G_1)$ is a subgroup of G_2.
2. *Fundamental homomorphism theorem for groups*: If $\varphi : G_1 \to G_2$ is a group homomorphism with kernel K, then K is a normal subgroup of G_1 and $G_1/K \cong \varphi(G_1)$.
3. If G is a cyclic group of infinite order, then $G \cong (\mathbb{Z}, +)$.
4. If G is a cyclic group of order n, then $G \cong (\mathbb{Z}_n, +)$.
5. If p is prime, then there is only one group of order p, the group $(\mathbb{Z}_p,+)$.
6. *Cayley's theorem*: If G is a finite group of order n, then G is isomorphic to some subgroup of the group of permutations on n objects.
7. $\mathbb{Z}_m \times \mathbb{Z}_n \cong \mathbb{Z}_{mn}$ if and only if m and n are relatively prime.
8. If $n = n_1 \cdot n_2 \cdot \ldots \cdot n_k$ where each n_i is a power of a different prime, then $\mathbb{Z}_n \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$.
9. *Fundamental theorem of finite Abelian groups*: Every finite Abelian group G (order ≥ 2) is isomorphic to a product of cyclic groups where each cyclic group has order a power of a prime. That is, there is a unique set $\{n_1, \ldots, n_k\}$ where each n_i is a power of some prime such that $G \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k}$.
10. *Fundamental theorem of finitely generated Abelian groups*: If G is a finitely generated Abelian group, then there is a unique integer $n \geq 0$ and a unique set $\{n_1, \ldots, n_k\}$ where each n_i is a power of some prime such that $G \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_k} \times \mathbb{Z}^n$ (G is *finitely generated* if there are $a_1, a_2, \ldots, a_n \in G$ such that every element of G can be written as $a_{k_1}^{\epsilon_1} a_{k_2}^{\epsilon_2} \cdots a_{k_j}^{\epsilon_j}$ where $k_i \in \{1, \ldots, n\}$ (the k_i are not necessarily distinct) and $\epsilon_i \in \{1, -1\}$).
11. *Fundamental homomorphism theorem for rings*: If $\varphi : R_1 \to R_2$ is a ring homomorphism with kernel K, then K is an ideal in R_1 and $R_1/K \cong \varphi(R_1)$.

2.5.10 TABLES

2.5.10.1 Number of non-isomorphic groups of different orders

The $10n + k$ entry is is found by looking at row n_ and the column _k. For example, there are 267 non-isomorphic groups with 64 elements.

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
0_		1	1	1	2	1	2	1	5	2
1_	2	1	5	1	2	1	14	1	5	1
2_	5	2	2	1	15	2	2	5	4	1
3_	4	1	51	1	2	1	14	1	2	2
4_	14	1	6	1	4	2	2	1	52	2
5_	5	1	5	1	15	2	13	2	2	1
6_	13	1	2	4	267	1	4	1	5	1
7_	4	1	50	1	2	3	4	1	6	1
8_	52	15	2	1	15	1	2	1	12	1
9_	10	1	4	2	2	1	231	1	5	2
10_	16	1	4	1	14	2	2	1	45	1
11_	6	2	43	1	6	1	5	4	2	1
12_	47	2	2	1	4	5	16	1	2328	2
13_	4	1	10	1	2	5	15	1	4	1
14_	11	1	2	1	197	1	2	6	5	1
15_	13	1	12	2	4	2	18	1	2	1
16_	238	1	55	1	5	2	2	1	57	2
17_	4	5	4	1	4	2	42	1	2	1
18_	37	1	4	2	12	1	6	1	4	13
29_	4	1	1543	1	2	2	12	1	10	1
20_	52	2	2	2	12	2	2	2	51	1
21_	12	1	5	1	2	1	177	1	2	2
22_	15	1	6	1	197	6	2	1	15	1
23_	4	2	14	1	16	1	4	2	4	1
24_	208	1	5	67	5	2	4	1	12	1
25_	15	1	46	2	2	1	56092	1	6	1
26_	15	2	2	1	39	1	4	1	4	1
27_	30	1	54	5	2	4	10	1	2	4
28_	40	1	4	1	4	2	4	1	1045	2
29_	4	2	5	1	23	1	14	5	2	1
30_	49	2	2	1	42	2	10	1	9	2
31_	6	1	61	1	2	4	4	1	4	1
32_	1640	1	4	1	176	2	2	2	15	1
33_	12	1	4	5	2	1	228	1	5	1
34_	15	1	18	5	12	1	2	1	12	1
35_	10	14	195	1	4	2	5	2	2	1

2.5.10.2 Number of non-isomorphic Abelian groups of different orders

The $10n + k$ entry is is found by looking at row n_ and the column _k.

	_0	_1	_2	_3	_4	_5	_6	_7	_8	_9
0_		1	1	1	2	1	1	1	3	2
1_	1	1	2	1	1	1	5	1	2	1
2_	2	1	1	1	3	2	1	3	2	1
3_	1	1	7	1	1	1	4	1	1	1
4_	3	1	1	1	2	2	1	1	5	2
5_	2	1	2	1	3	1	3	1	1	1

2.5.10.3 List of all sporadic finite simple groups

These are the sporadic finite simple groups that are not in any of the standard classes (see page 107).

M_{11}	$2^4 \cdot 3^2 \cdot 5 \cdot 11$
M_{12}	$2^6 \cdot 3^3 \cdot 5 \cdot 11$
M_{22}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11$
M_{23}	$2^7 \cdot 3^2 \cdot 5 \cdot 7 \cdot 11 \cdot 23$
M_{24}	$2^{10} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11 \cdot 23$
J_1	$2^3 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 19$
J_2	$2^7 \cdot 3^3 \cdot 5^2 \cdot 7$
J_3	$2^7 \cdot 3^5 \cdot 5 \cdot 17 \cdot 19$
J_4	$2^{21} \cdot 3^3 \cdot 5 \cdot 7 \cdot 11^3 \cdot 23 \cdot 29 \cdot 31 \cdot 37 \cdot 43$
HS	$2^9 \cdot 3^2 \cdot 5^3 \cdot 7 \cdot 11$
Mc	$2^7 \cdot 3^6 \cdot 5^3 \cdot 11$
Suz	$2^{13} \cdot 3^7 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$
Ru	$2^{14} \cdot 3^3 \cdot 5^3 \cdot 7 \cdot 13 \cdot 29$
He	$2^{10} \cdot 3^3 \cdot 5^2 \cdot 7^3 \cdot 17$
Ly	$2^8 \cdot 3^7 \cdot 5^6 \cdot 7 \cdot 11 \cdot 31 \cdot 37 \cdot 67$
ON	$2^9 \cdot 3^4 \cdot 5 \cdot 7^3 \cdot 11 \cdot 19 \cdot 31$
.1	$2^{21} \cdot 3^9 \cdot 5^4 \cdot 7^2 \cdot 11 \cdot 13 \cdot 23$
.2	$2^{18} \cdot 3^6 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$
.3	$2^{10} \cdot 3^7 \cdot 5^3 \cdot 7 \cdot 11 \cdot 23$
$M(22)$	$2^{17} \cdot 3^9 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13$
$M(23)$	$2^{18} \cdot 3^{13} \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 23$
$M(24)'$	$2^{21} \cdot 3^{16} \cdot 5^2 \cdot 7^3 \cdot 11 \cdot 13 \cdot 23 \cdot 29$
F_5	$2^{15} \cdot 3^{10} \cdot 5^3 \cdot 7^2 \cdot 13 \cdot 19 \cdot 31$
F_3	$2^{14} \cdot 3^6 \cdot 5^6 \cdot 7 \cdot 11 \cdot 19$
F_2	$2^{41} \cdot 3^{13} \cdot 5^6 \cdot 7^2 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 23 \cdot 31 \cdot 47$
F_1	$2^{46} \cdot 3^{20} \cdot 5^9 \cdot 7^6 \cdot 11^2 \cdot 13^3 \cdot 17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$

2.5.10.4 Names of groups of small order

Order n	Distinct groups of order n
1	$\{e\}$
2	C_2
3	C_3
4	$C_2 \times C_2$, C_4
5	C_5
6	C_6, D_3
7	C_7
8	$C_2 \times C_2 \times C_2$, $C_2 \times C_4$, C_8, Q, D_4
9	$C_3 \times C_3$, C_9
10	C_{10}, D_5
11	C_{11}
12	$C_2 \times C_6$, C_{12}, A_4, D_6, $T = C_3 \times C_4$
13	C_{13}
14	C_{14}, D_7
15	C_{15}

2.5.10.5 Representations of groups of small order

In all cases the identity element, $\{1\}$, forms a subgroup of order 1.

1. **C_2, the cyclic group of order 2**
   Generator: a with relation $a^2 = 1$

	1	a
1	1	a
a	a	1

Elements

(a) order 2: a

Subgroups

(a) order 2: $\{1, a\}$

2. **C_3, the cyclic group of order 3**
   Generator: a with relation $a^3 = 1$

	1	a	a^2
1	1	a	a^2
a	a	a^2	1
a^2	a^2	1	a

Elements

(a) order 3: a, a^2

Subgroups

(a) order 3: $\{1, a, a^2\}$

3. **C_4, the cyclic group of order 4**
Generator: a with relation $a^4 = 1$

	1	a	a^2	a^3
1	1	a	a^2	a^3
a	a	a^2	a^3	1
a^2	a^2	a^3	1	a
a^3	a^3	1	a	a^2

Elements

(a) order 4: a, a^3
(b) order 2: a^2

Subgroups

(a) order 4: $\{1, a, a^2, a^3\}$
(b) order 2: $\{1, a^2\}$

4. **V, the Klein four group**
Generators: a, b with relations $a^2 = 1$, $b^2 = 1$, $ba = ab$:

	1	a	b	ab
1	1	a	b	ab
a	a	1	ab	b
b	b	ab	1	a
ab	ab	b	a	1

Elements

(a) order 2: a, b, ab

Subgroups

(a) order 4: $\{1, a, b, ab\}$
(b) order 2: $\{1, a\}$, $\{1, b\}$, $\{1, ab\}$

5. **C_5, the cyclic group of order 5**
Generator: a with relation $a^5 = 1$

	1	a	a^2	a^3	a^4
1	1	a	a^2	a^3	a^4
a	a	a^2	a^3	a^4	1
a^2	a^2	a^3	a^4	1	a
a^3	a^3	a^4	1	a	a^2
a^4	a^4	1	a	a^2	a^3

Elements

(a) order 5: a, a^2, a^3, a^4

Subgroups

(a) order 5: $\{1, a, a^2, a^3, a^4\}$

6. **C_6, the cyclic group of order 6**
Generator: a with relation $a^6 = 1$

	1	a	a^2	a^3	a^4	a^5
1	1	a	a^2	a^3	a^4	a^5
a	a	a^2	a^3	a^4	a^5	1
a^2	a^2	a^3	a^4	a^5	1	a
a^3	a^3	a^4	a^5	1	a	a^2
a^4	a^4	a^5	1	a	a^2	a^3
a^5	a^5	1	a	a^2	a^3	a^4

Elements

(a) order 6: a, a^5
(b) order 3: a^2, a^4
(c) order 2: a^3

Subgroups

(a) order 6: $\{1, a, a^2, a^3, a^4, a^5\}$
(b) order 3: $\{1, a^2, a^4\}$
(c) order 2: $\{1, a^3\}$

7. **S_3, the symmetric group with three elements**
Generators: a, b with relations $a^3 = 1$, $b^2 = 1$, $ba = a^{-1}b$

	1	a	a^2	b	ab	a^2b
1	1	a	a^2	b	ab	a^2b
a	a	a^2	1	ab	a^2b	b
a^2	a^2	1	a	a^2b	b	ab
b	b	a^2b	ab	1	a^2	a
ab	ab	b	a^2b	a	1	a^2
a^2b	a^2b	ab	b	a^2	a	1

Elements

(a) order 3: a, a^2
(b) order 2: b, ab, a^2b

Subgroups

(a) order 6: $\{1, a, a^2, b, ab, a^2b\}$ (is a normal subgroup)
(b) order 3: $\{1, a, a^2\}$ (is a normal subgroup)
(c) order 2: $\{1, b\}$, $\{1, ab\}$, $\{1, a^2b\}$

8. **C_7, the cyclic group of order 7**
Generator: a with relation $a^7 = 1$

	1	a	a^2	a^3	a^4	a^5	a^6
1	1	a	a^2	a^3	a^4	a^5	a^6
a	a	a^2	a^3	a^4	a^5	a^6	1
a^2	a^2	a^3	a^4	a^5	a^6	1	a
a^3	a^3	a^4	a^5	a^6	1	a	a^2
a^4	a^4	a^5	a^6	1	a	a^2	a^3
a^5	a^5	a^6	1	a	a^2	a^3	a^4
a^6	a^6	1	a	a^2	a^3	a^4	a^5

Elements

(a) order 7: a, a^2, a^3, a^4, a^5, a^6

Subgroups

(a) order 7: $\{1, a, a^2, a^3, a^4, a^5, a^6\}$

9. **C_8, the cyclic group of order 8**
Generator: a with relation $a^8 = 1$

	1	a	a^2	a^3	a^4	a^5	a^6	a^7
1	1	a	a^2	a^3	a^4	a^5	a^6	a^7
a	a	a^2	a^3	a^4	a^5	a^6	a^7	1
a^2	a^2	a^3	a^4	a^5	a^6	a^7	1	a
a^3	a^3	a^4	a^5	a^6	a^7	1	a	a^2
a^4	a^4	a^5	a^6	a^7	1	a	a^2	a^3
a^5	a^5	a^6	a^7	1	a	a^2	a^3	a^4
a^6	a^6	a^7	1	a	a^2	a^3	a^4	a^5
a^7	a^7	1	a	a^2	a^3	a^4	a^5	a^6

Elements

(a) order 8: a, a^3, a^5, a^7
(b) order 4: a^2, a^6
(c) order 2: a^4

Subgroups

(a) order 8: $\{1, a, a^2, a^3, a^4, a^5, a^6, a^7\}$
(b) order 4: $\{1, a^2, a^4, a^6\}$
(c) order 2: $\{1, a^4\}$

10. **$C_4 \times C_2$, the direct product of a cyclic group of order 4 and a cyclic group of order 2**

Generators: a, b with relations $a^4 = 1$, $b^2 = 1$, and $ba = ab$

	1	a	a^2	a^3	b	ab	a^2b	a^3b
1	1	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	1	ab	a^2b	a^3b	b
a^2	a^2	a^3	1	a	a^2b	a^3b	b	ab
a^3	a^3	1	a	a^2	a^3b	b	ab	a^2b
b	b	ab	a^2b	a^3b	1	a	a^2	a^3
ab	ab	a^2b	a^3b	b	a	a^2	a^3	1
a^2b	a^2b	a^3b	b	ab	a^2	a^3	1	a
a^3b	a^3b	b	ab	a^2b	a^3	1	a	a^2

Elements

(a) order 4: a, a^3, ab, a^3b
(b) order 2: a^2, b, a^2b

Subgroups

(a) order 8: $\{1, a, a^2, a^3, b, ab, a^2b, a^3b\}$
(b) order 4: $\{1, a, a^2, a^3\}$ $\{1, ab, a^2, a^3b\}$ $\{1, a^2, b, a^2b\}$
(c) order 2: $\{1, a^2\}$, $\{1, b\}$, $\{1, a^2b\}$

11. **$C_2 \times C_2 \times C_2$, the direct product of 3 cyclic groups of order 2**

Generators: a, b, c with relations $a^2 = 1$, $b^2 = 1$, $c^2 = 1$, $ba = ab$, $ca = ac$, $cb = bc$

	1	a	b	ab	c	ac	bc	abc
1	1	a	b	ab	c	ac	bc	abc
a	a	1	ab	b	ac	c	abc	bc
b	b	ab	1	a	bc	abc	c	ac
ab	ab	b	a	1	abc	bc	ac	c
c	c	ac	bc	abc	1	a	b	ab
ac	ac	c	abc	bc	a	1	ab	b
bc	bc	abc	c	ac	b	ab	1	a
abc	abc	bc	ac	c	ab	b	a	1

Elements

(a) order 2: a, b, ab, c, ac, bc, abc

Subgroups

(a) order 8: $\{1, a, b, ab, c, ac, bc, abc\}$
(b) order 4: $\{1, a, b, ab\}$, $\{1, a, c, ac\}$, $\{1, a, bc, abc\}$, $\{1, b, c, bc\}$, $\{1, b, ac, abc\}$, $\{1, ab, c, abc\}$, $\{1, ab, ac, bc\}$
(c) order 2: $\{1, a\}$, $\{1, b\}$, $\{1, ab\}$, $\{1, c\}$, $\{1, ac\}$, $\{1, bc\}$, $\{1, abc\}$

12. **D_4, the dihedral group of order 8**
Generators: a, b with relations $a^4 = 1$, $b^2 = 1$, $ba = a^{-1}b$

	1	a	a^2	a^3	b	ab	a^2b	a^3b
1	1	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	1	ab	a^2b	a^3b	b
a^2	a^2	a^3	1	a	a^2b	a^3b	b	ab
a^3	a^3	1	a	a^2	a^3b	b	ab	a^2b
b	b	a^3b	a^2b	ab	1	a^3	a^2	a
ab	ab	b	a^3b	a^2b	a	1	a^3	a^2
a^2b	a^2b	ab	b	a^3b	a^2	a	1	a^3
a^3b	a^3b	a^2b	ab	b	a^3	a^2	a	1

Elements

(a) order 4: a, a^3
(b) order 2: a^2, b, ab, a^2b, a^3b

Normal subgroups

(a) order 8: $\{1, a, a^2, a^3, b, ab, a^2b, a^3b\}$
(b) order 4: $\{1, a^2, b, a^2b\}$, $\{1, a, a^2, a^3\}$, $\{1, a^2, ab, a^3b\}$
(c) order 2: $\{1, a^2\}$

Additional subgroups

(a) order 2: $\{1, b\}$, $\{1, a^2b\}$, $\{1, ab\}$, $\{1, a^3b\}$

13. **Q, the quaternion group (of order 8)**
Generators: a, b with relations $a^4 = 1$, $b^2 = a^2$, $ba = a^{-1}b$

	1	a	a^2	a^3	b	ab	a^2b	a^3b
1	1	a	a^2	a^3	b	ab	a^2b	a^3b
a	a	a^2	a^3	1	ab	a^2b	a^3b	b
a^2	a^2	a^3	1	a	a^2b	a^3b	b	ab
a^3	a^3	1	a	a^2	a^3b	b	ab	a^2b
b	b	a^3b	a^2b	ab	a^2	a	1	a^3
ab	ab	b	a^3b	a^2b	a^3	a^2	a	1
a^2b	a^2b	ab	b	a^3b	1	a^3	a^2	a
a^3b	a^3b	a^2b	ab	b	a	1	a^3	a^2

Elements

(a) order 4: a, a^3, b, ab, a^2b, a^3b
(b) order 2: a^2

Subgroups (all of them are normal subgroups)

(a) order 8: $\{1, a, a^2, a^3, b, ab, a^2b, a^3\}$
(b) order 4: $\{1, a, a^2, a^3\}$, $\{1, b, a^2, a^2b\}$, $\{1, ab, a^2, a^3b\}$
(c) order 2: $\{1, a^2\}$

Notes

- Q can be defined as the set $\{1, -1, i, -i, j, -j, k, -k\}$ where multiplication is defined by: $i^2 = j^2 = k^2 = -1$, $ij = -ji = k$, $jk = -kj = i$, and $ki = -ik = j$. This is an alternate group representation with:
  - Elements
    - order 4: i, $-i$, j, $-j$, k, $-k$
    - order 2: -1
  - Subgroups (all of them are normal subgroups)
    - order 8: $\{1, -1, i, -i, j, -j, k, -k\}$
    - order 4: $\{1, i, -1, -i\}$, $\{1, j, -1, -j\}$, $\{1, k, -1, -k\}$
    - order 2: $\{1, -1\}$
- Q can be defined as the group composed of the 8 matrices:

$$+1_q = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, -1_q = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, +i_q = \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix}, -i_q = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix},$$
$$+j_q = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}, -j_q = \begin{bmatrix} 0 & -i \\ -i & 0 \end{bmatrix}, +k_q = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, -k_q = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$$

  where a subscript of q indicates a quaternion element, $i^2 = -1$, and matrix multiplication is the group operation.

2.5.10.6 Small finite fields

In the following, the entries under α^i denote the coefficient of powers of α. For example, the last entry of the $p(x) = x^3 + x^2 + 1$ table is 1 1 0. That is: $\alpha^6 \equiv 1\alpha^2 + 1\alpha^1 + 0\alpha^0$ modulo $p(\alpha)$, where the coefficients are taken modulo 2.

$q = 4$	$x^2 + x + 1$
i	α^i
0	0 1
1	1 0
2	1 1

$q = 8$	$x^3 + x + 1$
i	α^i
0	0 0 1
1	0 1 0
2	1 0 0
3	0 1 1
4	1 1 0
5	1 1 1
6	1 0 1

$q = 8$	$x^3 + x^2 + 1$
i	α^i
0	0 0 1
1	0 1 0
2	1 0 0
3	1 0 1
4	1 1 1
5	0 1 1
6	1 1 0

$q = 16$		$x^4 + x + 1$	
i	α^i	7	1 0 1 1
0	0 0 0 1	8	0 1 0 1
1	0 0 1 0	9	1 0 1 0
2	0 1 0 0	10	0 1 1 1
3	1 0 0 0	11	1 1 1 0
4	0 0 1 1	12	1 1 1 1
5	0 1 1 0	13	1 1 0 1
6	1 1 0 0	14	1 0 0 1

$q = 16$		$x^4 + x^3 + 1$	
i	α^i	7	0 1 1 1
0	0 0 0 1	8	1 1 1 0
1	0 0 1 0	9	0 1 0 1
2	0 1 0 0	10	1 0 1 0
3	1 0 0 0	11	1 1 0 1
4	1 0 0 1	12	0 0 1 1
5	1 0 1 1	13	0 1 1 0
6	1 1 1 1	14	1 1 0 0

2.5.10.7 Addition and multiplication tables for F_2, F_3, F_4, and F_8

(a) F_2 addition and multiplication:

+	0	1
0	0	1
1	1	0

·	0	1
0	0	0
1	0	1

(b) F_3 addition and multiplication:

+	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

·	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

(c) F_4 addition and multiplication (using $\beta = \alpha + 1$):

+	0	1	α	β
0	0	1	α	β
1	1	0	β	α
α	α	β	0	1
β	β	α	1	0

·	0	1	α	β
0	0	0	0	0
1	0	1	α	β
α	0	α	β	1
β	0	β	1	α

(d) F_8 addition and multiplication (using strings of 0s and 1s to represent the polynomials: $0 = 000$, $1 = 001$, $\alpha = 010$, $\alpha + 1 = 011$, $\alpha^2 = 100$, $\alpha^2 + \alpha = 110$, $\alpha^2 + 1 = 101$, $\alpha^2 + \alpha + 1 = 111$):

+	000	001	010	011	100	101	110	111
000	000	001	010	011	100	101	110	111
001	001	000	011	010	101	100	111	110
010	010	011	000	001	110	111	100	101
011	011	010	001	000	111	110	101	100
100	100	101	110	111	000	001	010	011
101	101	100	111	110	001	000	011	010
110	110	111	100	101	010	011	000	001
111	111	110	101	100	011	010	001	000

·	000	001	010	011	100	101	110	111
000	000	000	000	000	000	000	000	000
001	000	001	010	011	100	101	110	111
010	000	010	100	110	011	001	111	101
011	000	011	110	101	111	100	001	010
100	000	100	011	111	110	010	101	001
101	000	101	001	100	010	111	011	110
110	000	110	111	001	101	011	010	100
111	000	111	101	010	001	110	100	011

2.5.10.8 Linear characters

A *linear character* of a finite group G is a homomorphism from G to the multiplicative group of the non-zero complex numbers. Let χ be a linear character of G, let ι be the identity of G and assume $g, h \in G$. Then

$$\chi(\iota) = 1 \qquad \chi(g \star h) = \chi(g)\chi(h) \qquad \chi(g^{-1}) = \overline{\chi(g)}$$
$$\chi(g) = e^{2\pi ik/n} \quad \text{for some integer } k\text{, where } n \text{ is the order of } g$$

The trivial character of G maps every element to 1. If χ_1 and χ_2 are two linear characters, then so is $\chi = \chi_1\chi_2$ defined by $\chi(g) = \chi_1(g)\chi_2(g)$. The linear characters form a group.

Group	Characters
C_n	For $m = 0, 1, \ldots, n-1$, $\quad \chi_m : k \mapsto e^{2\pi ikm/n}$
G Abelian	$G \cong \mathbb{Z}_{n_1} \times \mathbb{Z}_{n_2} \times \cdots \times \mathbb{Z}_{n_j}$ with each n_i a power of a prime. For $m_j = 0, 1, \ldots, n_j - 1$ and $g_j = (0, \ldots, 0, 1, 0, \ldots, 0)$ $\chi_{m_1,m_2,\ldots,m_n} : g_j \mapsto e^{2\pi im_j/n_j}$
D_n dihedral	For $x = \pm 1$, $y = \begin{cases} \pm 1 & \text{if } n \text{ even,} \\ 1 & \text{if } n \text{ odd,} \end{cases}$ $\chi_{x,y} : a \mapsto x, b \mapsto y$. (See definition of D_n.)
Quaternions	For $x \pm 1$ and $y = \pm 1$ or ∓ 1, $\chi_{x,y} : \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mapsto x, \qquad \chi_{x,y} : \begin{bmatrix} -i & 0 \\ 0 & i \end{bmatrix} \mapsto y$
S_n symmetric	The trivial character and sgn, where sgn is the signum function on permutations. (See Section 5.12.2.)

2.5.10.9 Indices and power residues

For $\mathbb{Z}_n^*$ the following table lists the index (order) of a and the *power residues* a, a^2, …, $a^{\text{index}(a)} = 1$ for each element a, where $(a, n) = 1$.

Group	Element	Index	Power residues
$\mathbb{Z}_2^*$	1	1	1
$\mathbb{Z}_3^*$	1	1	1
	2	2	2,1
$\mathbb{Z}_4^*$	1	1	1
	3	2	3,1
$\mathbb{Z}_5^*$	1	1	1
	2	4	2,4,3,1
	3	4	3,4,2,1
	4	2	4,1
$\mathbb{Z}_6^*$	1	1	1
	5	2	5,1

Group	Element	Index	Power residues
$\mathbb{Z}_7^*$	1	1	1
	2	3	2,4,1
	3	6	3,2,6,4,5,1
	4	3	4,2,1
	5	6	5,4,6,2,3,1
	6	2	6,1
$\mathbb{Z}_8^*$	1	1	1
	3	2	3,1
	5	2	5,1
	7	2	7,1

2.5.10.10 Power residues in $\mathbb{Z}_p$

For prime $p < 40$, the following table lists the minimal primitive root a and the power residues of a. These can be used to find $a^m \pmod{p}$ for any $(a,p) = 1$. For example, to find $3^7 \pmod{11}$ ($a = 3, m = 7$), look in row $p = 11$ until the power of a that is equal to 3 is found. In this case $2^8 \equiv 3 \pmod{11}$. This means that $3^7 \equiv (2^8)^7 \equiv 2^{56} \equiv (2^{10})^5 \cdot 2^6 \equiv 2^6 \equiv 9 \pmod{11}$.

p	a	Power residues	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9
3	2	0.	1	2	1							
5	2	0.	1	2	4	3	1					
7	3	0.	1	3	2	6	4	5	1			
11	2	0.	1	2	4	8	5	10	9	7	3	6
		1.	1									
13	2	0.	1	2	4	8	3	6	12	11	9	5
		1.	10	7	1							
17	3	0.	1	3	9	10	13	5	15	11	16	14
		1.	8	7	4	12	2	6	1			
19	2	0.	1	2	4	8	16	13	7	14	9	18
		1.	17	15	11	3	6	12	5	10	1	
23	5	0.	1	5	2	10	4	20	8	17	16	11
		1.	9	22	18	21	13	19	3	15	6	7
		2.	12	14	1							

2.5.10.11 Primitive normal polynomials

Degree	Primitive normal polynomials		
n	$p = 2$	$p = 3$	$p = 5$
2	$x^2 + x + 1$	$x^2 + x + 2$	$x^2 + x + 2$
3	$x^3 + x^2 + 1$	$x^3 + 2x^2 + 1$	$x^3 + x^2 + 2$
4	$x^4 + x^3 + 1$	$x^4 + x^3 + 2$	$x^4 + x^3 + 4x + 2$
5	$x^5 + x^4 + x^2 + x + 1$	$x^5 + 2x^4 + 1$	$x^5 + 2x^4 + 3$
6	$x^6 + x^5 + 1$	$x^6 + x^5 + x^3 + 2$	$x^6 + x^5 + 2$
7	$x^7 + x^6 + 1$	$x^7 + x^6 + x^2 + 1$	$x^7 + x^6 + 2$

2.5.10.12 Table of primitive monic polynomials

In the table below, the elements in each string are the coefficients of the polynomial after the highest power of x. (For example, 564 represents $x^3 + 5x^2 + 6x + 4$.)

Field	Degree	Primitive polynomials					
F_2	1	0	1				
	2	11					
	3	011	101				
	4	0011	1001				
	5	00101	01001	01111	10111	11011	11101
	6	000101	011011	100001	100111	101101	110011
F_3	1	0	1				
	2	12	22				
	3	021	121	201	211		
	4	0012	0022	1002	1122	1222	2002
		2112	2212				
F_5	1	0	2	3			
	2	12	23	33	42		
	3	032	033	042	043	102	113
		143	203	213	222	223	242
		302	312	322	323	343	403
		412	442				
F_7	1	0	2	4			
	2	13	23	25	35	45	53
		55	63				
	3	032	052	062	112	124	152
		154	214	242	262	264	304
		314	322	334	352	354	362
		422	432	434	444	504	524
		532	534	542	552	564	604
		612	632	644	654	662	664

2.5.10.13 Table of irreducible polynomials in $\mathbb{Z}_2[x]$

Each polynomial is represented by its coefficients (which are either 0 or 1), beginning with the highest power. For example, $x^4 + x + 1$ is represented as 10011.

degree 1:	10	11				
degree 2:	111					
degree 3:	1011	1101				
degree 4:	10011	11001	11111			
degree 5:	100101	101001	101111	110111	111011	111101
degree 6:	1000011	1001001	1010111	1011011	1100001	1100111
	1101101	1110011	1110101			
degree 8:	100011011	100011101	100101011	100101101	100111001	100111111
	101001101	101011111	101100011	101100101	101101001	101110001
	101110111	101111011	110000111	110001011	110001101	110011111
	110100011	110101001	110110001	110111101	111000011	111001111
	111010111	111011101	111100111	111110011	111110101	111111001

2.5.10.14 Table of primitive roots

The number of integers not exceeding and relatively prime to the integer n is $\phi(n)$ (see page 46). These integers form a group under multiplication module n; the group is cyclic if and only if $n = 1, 2, 4$ or n is of the form p^k or $2p^k$, where p is an odd prime. The number g is a *primitive root* of n if it generates that group, i.e., if $\{g, g^2, \ldots, g^{\phi(n)}\}$ are distinct modulo n. There are $\phi(\phi(n))$ primitive roots of n.

1. If g is a primitive root of p and $g^{p-1} \not\equiv 1 \pmod{p^2}$, then g is a primitive root of p^k for all k.
2. If $g^{p-1} \equiv 1 \pmod{p^2}$ then $g + p$ is a primitive root of p^k for all k.
3. If g is a primitive root of p^k, then either g or $g + p^k$, whichever is odd, is a primitive root of $2p^k$.
4. If g is a primitive root of n, then g^k is a primitive root of n if and only if k and $\phi(n)$ are relatively prime, i.e., $(\phi(n), k) = 1$.

In the following table,

- g denotes the least primitive root of p
- G denotes the least negative primitive root of p
- ϵ denotes whether 10, -10, or both, are primitive roots of p

p	$p-1$	g	G	ϵ	p	$p-1$	g	G	ϵ
3	2	2	-1	—	5	2^2	2	-2	—
7	$2 \cdot 3$	3	-2	10	11	$2 \cdot 5$	2	-3	—
13	$2^2 \cdot 3$	2	-2	—	17	2^4	3	-3	± 10
19	$2 \cdot 3^2$	2	-4	10	23	$2 \cdot 11$	5	-2	10
29	$2^2 \cdot 7$	2	-2	± 10	31	$2 \cdot 3 \cdot 5$	3	-7	-10
37	$2^2 \cdot 3^2$	2	-2	—	41	$2^3 \cdot 5$	6	-6	—
43	$2 \cdot 3 \cdot 7$	3	-9	-10	47	$2 \cdot 23$	5	-2	10
53	$2^2 \cdot 13$	2	-2	—	59	$2 \cdot 29$	2	-3	10
61	$2^2 \cdot 3 \cdot 5$	2	-2	± 10	67	$2 \cdot 3 \cdot 11$	2	-4	-10
71	$2 \cdot 5 \cdot 7$	7	-2	-10	73	$2^3 \cdot 3^2$	5	-5	—
79	$2 \cdot 3 \cdot 13$	3	-2	—	83	$2 \cdot 41$	2	-3	-10
89	$2^3 \cdot 11$	3	-3	—	97	$2^5 \cdot 3$	5	-5	± 10
101	$2^2 \cdot 5^2$	2	-2	—	103	$2 \cdot 3 \cdot 17$	5	-2	—
107	$2 \cdot 53$	2	-3	-10	109	$2^2 \cdot 3^3$	6	-6	± 10
113	$2^4 \cdot 7$	3	-3	± 10	127	$2 \cdot 3^2 \cdot 7$	3	-9	—
131	$2 \cdot 5 \cdot 13$	2	-3	10	137	$2^3 \cdot 17$	3	-3	—
139	$2 \cdot 3 \cdot 23$	2	-4	—	149	$2^2 \cdot 37$	2	-2	± 10
151	$2 \cdot 3 \cdot 5^2$	6	-5	-10	157	$2^2 \cdot 3 \cdot 13$	5	-5	—
163	$2 \cdot 3^4$	2	-4	-10	167	$2 \cdot 83$	5	-2	10
173	$2^2 \cdot 43$	2	-2	—	179	$2 \cdot 89$	2	-3	10
181	$2^2 \cdot 3^2 \cdot 5$	2	-2	± 10	191	$2 \cdot 5 \cdot 19$	19	-2	-10
193	$2^6 \cdot 3$	5	-5	± 10	197	$2^2 \cdot 7^2$	2	-2	—
199	$2 \cdot 3^2 \cdot 11$	3	-2	-10	211	$2 \cdot 3 \cdot 5 \cdot 7$	2	-4	—
223	$2 \cdot 3 \cdot 37$	3	-9	10	227	$2 \cdot 113$	2	-3	-10

2.5.10.15 Table of factorizations of $x^n - 1$

n	Factorization of $x^n - 1$ mod 2
1	$-1 + x$
2	$(-1 + x)\,(1 + x)$
3	$(-1 + x)\,(1 + x + x^2)$
4	$(-1 + x)\,(1 + x)\,(1 + x^2)$
5	$(-1 + x)\,(1 + x + x^2 + x^3 + x^4)$
6	$(-1 + x)\,(1 + x)\,(1 - x + x^2)\,(1 + x + x^2)$
7	$(-1 + x)\,(1 + x + x^2 + x^3 + x^4 + x^5 + x^6)$
8	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 + x^4)$
9	$(-1 + x)\,(1 + x + x^2)\,(1 + x^3 + x^6)$
10	$(-1 + x)\,(1 + x)\,(1 - x + x^2 - x^3 + x^4)\,(1 + x + x^2 + x^3 + x^4)$
11	$(-1 + x)\,(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10})$
12	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 - x + x^2)\,(1 + x + x^2)\,(1 - x^2 + x^4)$
13	$(-1 + x)\left(\sum_{k=0}^{12} x^k\right)$
14	$(-1 + x)\,(1 + x)\,(1 - x + x^2 - x^3 + x^4 - x^5 + x^6)$ $(1 + x + x^2 + x^3 + x^4 + x^5 + x^6)$
15	$(-1 + x)\,(1 + x + x^2)\,(1 + x + x^2 + x^3 + x^4)$ $(1 - x + x^3 - x^4 + x^5 - x^7 + x^8)$
16	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 + x^4)\,(1 + x^8)$
17	$(-1 + x)\left(\sum_{k=0}^{16} x^k\right)$
18	$(-1 + x)\,(1 + x)\,(1 - x + x^2)\,(1 + x + x^2)\,(1 - x^3 + x^6)$ $(1 + x^3 + x^6)$
19	$(-1 + x)\left(\sum_{k=0}^{18} x^k\right)$
20	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 - x + x^2 - x^3 + x^4)$ $(1 + x + x^2 + x^3 + x^4)\,(1 - x^2 + x^4 - x^6 + x^8)$
21	$(-1 + x)\,(1 + x + x^2)\,(1 + x + x^2 + x^3 + x^4 + x^5 + x^6)$ $(1 - x + x^3 - x^4 + x^6 - x^8 + x^9 - x^{11} + x^{12})$
22	$(-1 + x)\,(1 + x)$ $(1 - x + x^2 - x^3 + x^4 - x^5 + x^6 - x^7 + x^8 - x^9 + x^{10})$ $(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10})$
23	$(-1 + x)\left(\sum_{k=0}^{22} x^k\right)$
24	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 - x + x^2)\,(1 + x + x^2)\,(1 + x^4)$ $(1 - x^2 + x^4)\,(1 - x^4 + x^8)$
25	$(-1 + x)\,(1 + x + x^2 + x^3 + x^4)\,(1 + x^5 + x^{10} + x^{15} + x^{20})$
26	$(-1 + x)\,(1 + x)$ $(1 - x + x^2 - x^3 + x^4 - x^5 + x^6 - x^7 + x^8 - x^9 + x^{10} - x^{11} + x^{12})$ $(1 + x + x^2 + x^3 + x^4 + x^5 + x^6 + x^7 + x^8 + x^9 + x^{10} + x^{11} + x^{12})$
27	$(-1 + x)\,(1 + x + x^2)\,(1 + x^3 + x^6)\,(1 + x^9 + x^{18})$
28	$(-1 + x)\,(1 + x)\,(1 + x^2)\,(1 - x + x^2 - x^3 + x^4 - x^5 + x^6)$ $(1 + x + x^2 + x^3 + x^4 + x^5 + x^6)\,(1 - x^2 + x^4 - x^6 + x^8 - x^{10} + x^{12})$
29	$(-1 + x)\left(\sum_{k=0}^{28} x^k\right)$

Chapter 3

Discrete Mathematics

3.1 SETS

A *set* can be thought of as a collection of objects. Examples of sets are:

1. The people in Boston on January 1, 2018
2. The real numbers between 0 and 1 inclusive
3. The numbers 1, 2, 3, and 4
4. All of the formulas in this book

Commonly used sets of numbers include the following:

$\mathbb{N}$	positive integers (also known as natural numbers)
$\mathbb{Z}$	integers
$\mathbb{Q}$	rational numbers (quotients of integers)
$\mathbb{R}$	real numbers
$[-\infty, +\infty]$	extended reals
$\mathbb{C}$	complex numbers

Note that: $\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}$.

Formal set theory can be defined by the (Zermelo–Fraenkell) axioms:

1. Axiom of Extensionality
2. Axiom of the Unordered Pair
3. Axiom of Subsets
4. Axiom of the Sum Set
5. Axiom of the Power Set
6. Axiom of Infinity
7. Axiom of Replacement
8. Axiom of Foundation
9. Axiom of Choice

3.1.1 SET OPERATIONS AND RELATIONS

If x is an element of a set A, then we write $x \in A$ (read "x is in A.") If x is not in A, we write $x \notin A$. When considering sets, a set U called the *universe*, is chosen, from which all elements are taken. The *null set* or *empty set* $\emptyset$ is the set containing no elements. Thus, $x \notin \emptyset$ for all $x \in U$.

Relation	**Read as**	**Definition**
$A \subseteq B$	A is contained in B	All elements of A are also elements of B
$A = B$	A equals B	$(A \subseteq B)$ and $(B \subseteq A)$

Basic set operations include:

1. The **union** of the sets A and B, denoted $A \cup B$, are all the elements that are in A or are in B. The **intersection** of the sets A and B, denoted $A \cap B$, are all the elements in both A and B.
   Distributive laws: Intersection and union distribute over each other. That is:
   $$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C) \end{aligned} \tag{3.1.1}$$
   for all sets A, B, C.

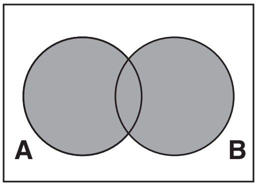
2. Given the set X, the **power set** of X is $\{S \mid S \subseteq X\}$, the set of all subsets of X. The power set of X is written as as $\mathcal{P}(X)$ or 2^X.
   For example, if $X = \{0, 1\}$ then $\mathcal{P}(\{0, 1\}) = \{\emptyset, \{0\}, \{1\}, \{0, 1\}\}$. Note that $\mathcal{P}(\emptyset) = \{\emptyset\}$; the power set of the empty set has one element.
3. If S and X are sets, then the **complement** of S in X, is represented in many different ways: $X \backslash S = X - S = S' = \{x \in X \text{ and } x \notin S\}$
   For example, $\{a, b, c\} \backslash \{c, d\} = \{a, b, c\} - \{c, d\} = \{c, d\}' = \{a, b\}$
   When S is a subset of X, then the complement of S in X can be written $\complement S$.
   (a) $S \setminus T = S \cap \complement T$.
   (b) $\complement\complement S = S$
   (c) *De Morgan's Laws for sets*: The complement of a union is the intersection of complements, and vice versa
   $$\complement\left(\bigcup_{\lambda\in\Lambda} S_\lambda\right) = \bigcap_{\lambda\in\Lambda} (\complement S_\lambda), \qquad \complement\left(\bigcap_{\lambda\in\Lambda} S_\lambda\right) = \bigcup_{\lambda\in\Lambda} (\complement S_\lambda).$$
   (d) There is an order-reversing *duality* between statements about collections of sets and statements about the complements of those sets:
   $$A \subseteq B \quad \Longleftrightarrow \quad \complement B \supseteq \complement A.$$
4. The **symmetric difference** of two sets A and B (represented as $A \oplus B$ or $A \Delta B$) is the set of elements that are in A or B and are not in both A and B (i.e., the union minus the intersection).
   $$\begin{aligned} A \triangle B &= (A \backslash B) \cup (B \backslash A) \\ &= (A \cap \complement B) \cup (\complement A \cap B) \\ &= (A \cup B) \setminus (A \cap B) \end{aligned}$$
   For example, $\{a, b, c\} \triangle \{b, c, d\} = \{a, d\}$. Note that:
   (a) $A \triangle B = B \triangle A$
   (b) $A \triangle A = \emptyset$
   (c) $A \triangle \emptyset = A$
   (d) If $\{A, B\}$ are subsets of X, then $\complement(A \triangle B) = A \triangle (\complement B) = (\complement A) \triangle B$.
   (e) $(C \triangle A) \triangle B = C \triangle (A \triangle B) = \{x \mid x \text{ is in one or three of } \{A, B, C\}\}$.

3.1.2 CONNECTION BETWEEN SETS AND PROBABILITY

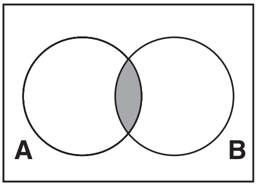
Set concept	Probability concept
Set	Event
Set containing a single element	Indecomposable, elementary, or atomic event
Set with more than one element	Compound event
Universal set or space	Sample space
Complement of a set	Non-occurrence of an event
Function on the universal set	Random variable
Measure of a set	Probability of an event
Integral with respect to the measure	Expectation or expected value

3.1.3 VENN DIAGRAMS

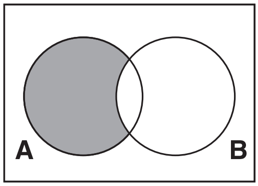
The operations and relations on sets can be illustrated by *Venn diagrams*. The diagrams below show a few possibilities.



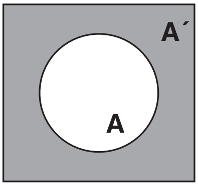
$A \cup B$



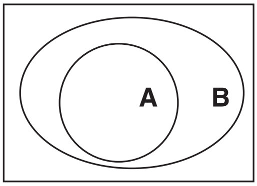
$A \cap B$



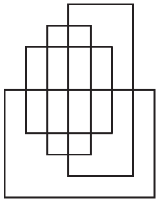
$A - B$



A'



$A \subseteq B$



Venn diagrams can be constructed to show combinations of many events. Each of the 2^4 regions created by the rectangles in the diagram to the right represents a different combination.

3.1.4 PARADOXES AND THEOREMS OF SET THEORY

3.1.4.1 Russell's paradox

Around 1900, Bertrand Russell presented a paradox, paraphrased as follows: since the elements of sets can be arbitrary, sets can contain sets as elements. Therefore, a set can possibly be a member of itself. (For example, the set of all sets would be a member of itself. Another example is the collection of all sets that can be described in fewer than 50 words.) Now let A be the set of all sets which are *not* members of themselves. Then if A is a member of itself, it is not a member of itself. And if A is not a member of itself, then by definition, A is a member of itself. This paradox led to a more careful consideration of how sets should be defined.

3.1.4.2 Infinite sets and the continuum hypothesis

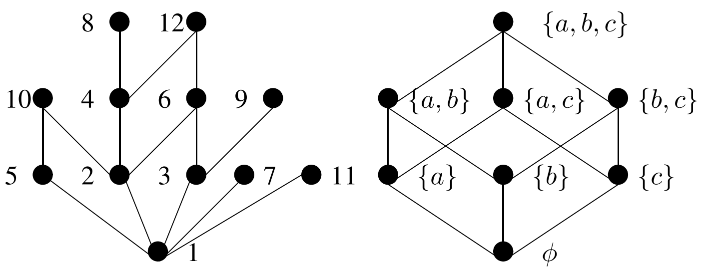
Georg Cantor showed how the number of elements of infinite sets can be counted, much as finite sets. He used the symbol $\aleph_0$ (read "aleph null") for the number of integers and introduced larger infinite numbers such as $\aleph_1$, $\aleph_2$, and so on. A few of his results were as follows (note: the "+" sign below is a set operation)

$$\aleph_0 + \aleph_0 = \aleph_0, \qquad (\aleph_0)^2 = \aleph_0, \qquad \mathbf{c} = 2^{\aleph_0} = \aleph_0^{\aleph_0} > \aleph_0.$$

where $\mathbf{c}$ is the cardinality of real numbers. The *continuum hypothesis* asks whether or not $\mathbf{c} = \aleph_1$, the first infinite cardinal greater than $\aleph_0$. In 1963, Cohen showed this result is independent of the other axioms of set theory: "... the truth or falsity of the continuum hypothesis ... cannot be determined by set theory as we know it today."

FIGURE 3.1

Left: Hasse diagram for integers up to 12 with $x \preceq y$ meaning "the number x divides the number y." Right: Hasse diagram for the power set of $\{a, b, c\}$ with $x \preceq y$ meaning "the set x is a subset of the set y."



3.1.5 PARTIALLY ORDERED SETS

Consider a set S and a relation on it. Given any two elements x and y in S, we can determine whether or not x is "related" to y; if it is, "$x \preceq y$." The relation "$\preceq$" will be a *partial order* on S if it satisfies the following three conditions:

reflexive	$s \preceq s$ for every $s \in S$,
antisymmetric	$s \preceq t$ and $t \preceq s$ imply $s = t$, and
transitive	$s \preceq t$ and $t \preceq u$ imply $s \preceq u$.

If $\preceq$ is a partial order on S, then the pair $(S, \preceq)$ is called a *partially ordered set* or a *poset*. Given the partial order $\preceq$ on the set S, define the relation $\prec$ by

$$x \prec y \qquad \text{if and only if} \qquad x \preceq y \text{ and } x \neq y.$$

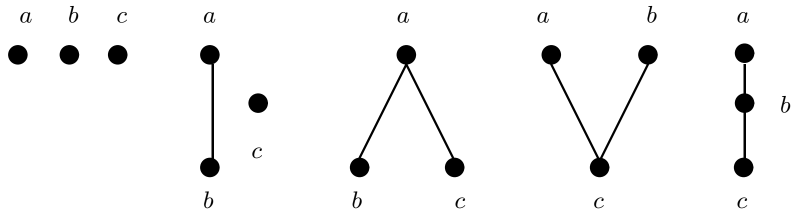
We say that the element t *covers* the element s if $s \prec t$ and there is no element u with $s \prec u \prec t$. A *Hasse diagram* of the poset $(S, \preceq)$ is a figure consisting of the elements of S with a line segment directed generally upward from s to t whenever t covers s. (See Figure 3.1.)

Two elements x and y in a poset $(S, \preceq)$ are said to be *comparable* if either $x \preceq y$ or $y \preceq x$. If every pair of elements in a poset is comparable, then $(S, \preceq)$ is a *chain*. An *antichain* is a poset in which no two elements are comparable (i.e., $x \preceq y$ if and only if $x = y$ for all x and y in the antichain). A *maximal chain* is a chain that is not properly contained in another chain (and similarly for a *maximal antichain*).

EXAMPLES

1. Let S be the set of natural numbers up to 12 and let "$x \preceq y$" mean "the number x divides the number y." Then $(S, \preceq)$ is a poset with the Hasse diagram shown in Figure 3.1 (left). Observe that the elements 2 and 4 are comparable, but elements 2 and 5 are not comparable.
2. Let S be the set of all subsets of the set $\{a, b, c\}$ and let "$x \preceq y$" mean "the set x is contained in the set y." Then $(S, \preceq)$ is a poset with the Hasse diagram shown in Figure 3.1 (right).

3. There are 16 different isomorphism types of posets of size 4. There are 5 different isomorphism types of posets of size 3, shown below:



3.1.6 INCLUSION/EXCLUSION

Let $\{a_1, a_2, \ldots, a_r\}$ be properties that the elements of a set may or may not have. If the set has N objects, then the number of objects having exactly m properties (with $m \leq r$), e_m, is given by

$$e_m = s_m - \binom{m+1}{1} s_{m+1} + \binom{m+2}{2} s_{m+2} - \binom{m+3}{3} s_{m+3} + \cdots \qquad (3.1.2)$$
$$\cdots + (-1)^p \binom{m+p}{p} s_{m+p} \cdots + (-1)^{r-m} \binom{m+(r-m)}{(r-m)} s_r.$$

Here $s_t = \sum N(a_{i_1} a_{i_2} \cdots a_{i_t})$ is the number of elements that have any t distinct properties. When $m = 0$, this is the usual inclusion/exclusion rule:

$$e_0 = s_0 - s_1 + s_2 - \cdots + (-1)^r s_r = N \qquad (3.1.3)$$
$$- \sum_i N(a_i) + \sum_{i,\, j \text{ distinct}} N(a_i a_j) - \sum_{i,\, j,\, k \text{ distinct}} N(a_i a_j a_k) + \cdots + (-1)^r N(a_1 a_2 \ldots a_r)$$

EXAMPLE Hatcheck problem: distribute n hats to their n owners. Let a_i be the property that the i^{th} person receives their own hat. Then $N(a_1 a_2 \ldots a_r) = (n-r)!$ and so $s_t = \sum N(a_{i_1} a_{i_2} \cdots a_{i_t}) = \binom{n}{t}(n-t)!$. Therefore the number of ways in which exactly one person gets their own hat back is $e_1 = s_1 - \binom{2}{1} s_2 + \binom{3}{2} s_3 - \ldots = n! \left[1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^{n-1}}{(n-1)!}\right] = n\, D_{n-1}$. (See page 144 for D_n.)

3.1.7 PIGEONHOLE PRINCIPLE

Pigeonhole Principle

If n items are put into m pigeonholes with $n > m$, then at least one pigeonhole must contain more than one item.

Generalized Pigeonhole Principle

If n items are put into m containers, then

- at least one container must hold no fewer than $\lceil \frac{n}{m} \rceil$ objects;
- at least one container must hold no more than $\lfloor \frac{n}{m} \rfloor$ items.

where $\lceil \cdot \rceil$ and $\lfloor \cdot \rfloor$ are the ceiling and floor functions.

EXAMPLES

- If you select five numbers from the integers 1 to 8, then two of them must add up to nine.
- In any group of two or more people, there must be at least two people who have the same number of friends in the group.

3.2 COMBINATORICS

3.2.1 SAMPLE SELECTION

There are four different ways in which a sample of r elements can be obtained from a set of n distinguishable objects:

$$\begin{aligned} C(n,r) &= \binom{n}{r} = \frac{n!}{r!\,(n-r)!}, \\ P(n,r) &= (n)_r = n^{\underline{r}} = \frac{n!}{(n-r)!}, \\ C^R(n,r) &= C(n+r-1,r) = \frac{(n+r-1)!}{r!(n-1)!}, \text{ and} \\ P^R(n,r) &= n^r. \end{aligned} \tag{3.2.1}$$

EXAMPLE There are four ways to choose a 2-element sample from the set $\{a, b\}$:

r-combination	$C(2,2) = 1$	$\Rightarrow$	ab
r-permutation	$P(2,2) = 2$	$\Rightarrow$	ab and ba
r-combination with replacement	$C^R(2,2) = 3$	$\Rightarrow$	aa, ab, and bb
r-permutation with replacement	$P^R(2,2) = 4$	$\Rightarrow$	aa, ab, ba, and bb

3.2.2 BALLS INTO CELLS

There are eight different ways in which n balls can be placed into k cells:

Distinguish the balls?	Distinguish the cells?	Can cells be empty?	Number of ways to place n balls into k cells
Yes	Yes	Yes	k^n
Yes	Yes	No	$k!\left\{ {n \atop k} \right\}$
Yes	No	Yes	$\left\{ {n \atop 1} \right\} + \left\{ {n \atop 2} \right\} + \ldots + \left\{ {n \atop k} \right\}$
Yes	No	No	$\left\{ {n \atop k} \right\}$
No	Yes	Yes	$C(k+n-1,n) = \binom{k+n-1}{n}$
No	Yes	No	$C(n-1,k-1) = \binom{n-1}{k-1}$
No	No	Yes	$p_1(n) + p_2(n) + \ldots + p_k(n)$
No	No	No	$p_k(n)$

where $\left\{ {n \atop k} \right\}$ is the Stirling subset number (see page 147) and $p_k(n)$ is the number of partitions of the number n into exactly k integer pieces (see page 145).

Given n distinguishable balls and k distinguishable cells, the number of ways in which we can place n_1 balls into cell 1, n_2 balls into cell 2, ..., n_k balls into cell k, is given by the multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_k}$ (see page 143).

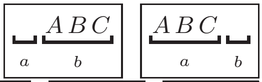
EXAMPLE

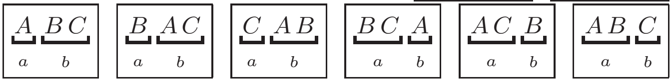
Consider placing $n = 3$ balls into $k = 2$ cells. Let $\{A, B, C\}$ denote the names of the balls (when needed) and $\{a, b\}$ denote the names of the cells (when needed). A cell will be denoted like this. Begin with: Are the balls distinguishable?

1. Yes, the balls are distinguishable. Are the cells distinguishable?

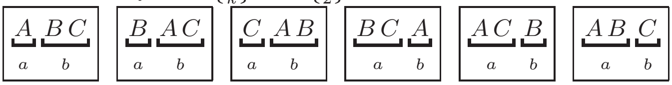
   (a) Yes, the cells are distinguishable. Can the cells be empty?

   i. Yes. Number of ways is $k^n = 2^3 = 8$:



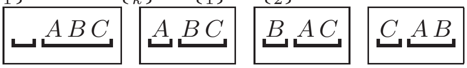


   ii. No. Number of ways is $k!\left\{{n \atop k}\right\} = 2!\left\{{3 \atop 2}\right\} = 6$:



   (b) No, the cells are not distinguishable. Can the cells be empty?

   i. Yes. Number of ways is $\left\{{n \atop 1}\right\} + \ldots + \left\{{n \atop k}\right\} = \left\{{3 \atop 1}\right\} + \left\{{3 \atop 2}\right\} = 1 + 3 = 4$:

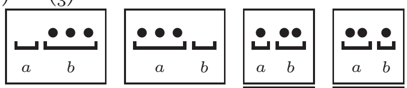


   ii. No. Number of ways is $\left\{{n \atop k}\right\} = \left\{{3 \atop 2}\right\} = 3$:

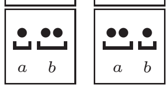
2. No, the balls are not distinguishable. Are the cells distinguishable?

   (a) Yes, the cells are distinguishable. Can the cells be empty?

   i. Yes. Number of ways is $\binom{k+n-1}{n} = \binom{4}{3} = 4$:

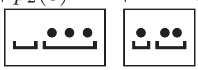


   ii. No. Number of ways is $\binom{n-1}{k-1} = \binom{2}{1} = 2$:

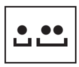


   (b) No, the cells are not distinguishable. Can the cells be empty?

   i. Yes. Number of ways is $p_1(n) + \ldots + p_k(n) = p_1(3) + p_2(3) = 1 + 1 = 2$:



   ii. No. Number of ways is $p_k(n) = p_2(3) = 1$:



3.2.3 BINOMIAL COEFFICIENTS

The binomial coefficient $C(n,m) = \binom{n}{m}$ is the number of ways of choosing m objects from a collection of n distinct objects without regard to order:

EXAMPLE For the 5-element set $\{a,b,c,d,e\}$ there are $\binom{5}{3} = \frac{5!}{3!2!} = 10$ subsets containing exactly three elements. They are:

$$\begin{array}{lllll} \{a,b,c\}, & \{a,b,d\}, & \{a,b,e\}, & \{a,c,d\}, & \{a,c,e\}, \\ \{a,d,e\}, & \{b,c,d\}, & \{b,c,e\}, & \{b,d,e\}, & \{c,d,e\}. \end{array}$$

Properties of binomial coefficients include:

1. $\displaystyle\binom{n}{m} = \frac{n!}{m!(n-m)!} = \frac{n(n-1)\cdots(n-m+1)}{m!} = \binom{n}{n-m}$.
2. $\displaystyle\binom{n}{0} = \binom{n}{n} = 1$ and $\displaystyle\binom{n}{1} = n$.
3. $\displaystyle\binom{2n}{n} = \frac{2^n(2n-1)!!}{n!} = \frac{2^n(2n-1)(2n-3)\cdots 3\cdot 1}{n!}$.
4. If n and m are integers, and $m > n$, then $\binom{n}{m} = 0$.
5. The recurrence relation: $\binom{n+1}{m} = \binom{n}{m} + \binom{n}{m-1}$.
6. Two generating functions for binomial coefficients are $\sum_{m=0}^{n} \binom{n}{m} x^m = (1+x)^n$ for $n = 1, 2, \ldots$, and $\sum_{n=m}^{\infty} \binom{n}{m} x^{n-m} = (1-x)^{-m-1}$.
7. The *Vandermonde convolution* is $\displaystyle\binom{x+y}{n} = \sum_{k=0}^{n} \binom{x}{k}\binom{y}{n-k}$.

3.2.3.1 Pascal's triangle

The binomial coefficients $\binom{n}{k}$ can be arranged in a triangle in which each number is the sum of the two numbers above it. For example $\binom{3}{2} = \binom{2}{1} + \binom{2}{2}$.

$$\begin{array}{c} 1 \\ 1\quad 1 \\ 1\quad 2\quad 1 \\ 1\quad 3\quad 3\quad 1 \\ 1\quad 4\quad 6\quad 4\quad 1 \\ 1\quad 5\quad 10\quad 10\quad 5\quad 1 \\ 1\quad 6\quad 15\quad 20\quad 15\quad 6\quad 1 \\ 1\quad 7\quad 21\quad 35\quad 35\quad 21\quad 7\quad 1 \\ 1\quad 8\quad 28\quad 56\quad 70\quad 56\quad 28\quad 8\quad 1 \end{array} \qquad \begin{array}{c} \binom{0}{0} \\ \binom{1}{0}\quad \binom{1}{1} \\ \binom{2}{0}\quad \binom{2}{1}\quad \binom{2}{2} \\ \binom{3}{0}\quad \binom{3}{1}\quad \binom{3}{2}\quad \binom{3}{3} \\ \binom{4}{0}\quad \binom{4}{1}\quad \binom{4}{2}\quad \binom{4}{3}\quad \binom{4}{4} \end{array}$$

3.2.3.2 Binomial coefficient relationships

The binomial coefficients satisfy

$$
\begin{array}{lll}
[a] & \binom{n+1}{m+1} = \binom{n}{m} + \binom{n}{m+1}, & \\
[b] & \binom{n}{m} = \binom{n}{n-m}, & \\
[c] & \binom{n+m+1}{n+1} = \binom{n}{n} + \binom{n+1}{n} + \binom{n+2}{n} + \ldots + \binom{n+m}{n}, & \\
[d] & \binom{2n}{n} = \binom{n}{0}^2 + \binom{n}{1}^2 + \ldots + \binom{n}{n}^2, & \\
[e] & \binom{m+n}{p} = \binom{m}{0}\binom{n}{p} + \binom{m}{1}\binom{n}{p-1} + \ldots + \binom{m}{p}\binom{n}{0}, & \\
[f] & 2^n = \binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{n}, & \\
[g] & 0 = \binom{n}{0} - \binom{n}{1} + \ldots + (-1)^n\binom{n}{n} & \text{for } n \geq 1, \\
[h] & 2^{n-1} = \binom{n}{0} + \binom{n}{2} + \binom{n}{4} + \ldots & \text{for } n \geq 1, \\
[i] & 2^{n-1} = \binom{n}{1} + \binom{n}{3} + \binom{n}{5} + \ldots & \text{for } n \geq 1, \\
[j] & 0 = 1\binom{n}{1} - 2\binom{n}{2} + \ldots + (-1)^{n+1}n\binom{n}{n} & \text{for } n \geq 1.
\end{array}
$$

3.2.4 MULTINOMIAL COEFFICIENTS

The multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_k}$ (also written $C(n; n_1, n_2, \ldots, n_k)$) is the number of ways of choosing n_1 objects, then n_2 objects, ..., then n_k objects from a collection of $n = \sum_{j=1}^{k} n_j$ distinct objects without regard to order. The multinomial coefficient is numerically evaluated as

$$\binom{n}{n_1, n_2, \ldots, n_k} = \frac{n!}{n_1!\, n_2! \cdots n_k!}. \tag{3.2.2}$$

EXAMPLE The number of ways to choose 2 objects, then 1 object, then 1 object from the set $\{a, b, c, d\}$ is $\binom{4}{2,1,1} = \frac{4!}{2!\,1!\,1!} = \frac{24}{2} = 12$; they are as follows (vertical bars show the ordered selections):

$$
\begin{array}{llll}
|\, ab \,|\, c \,|\, d \,|, & |\, ab \,|\, d \,|\, c \,|, & |\, ac \,|\, b \,|\, d \,|, & |\, ac \,|\, d \,|\, b \,|, \\
|\, ad \,|\, b \,|\, c \,|, & |\, ad \,|\, c \,|\, b \,|, & |\, bc \,|\, a \,|\, d \,|, & |\, bc \,|\, d \,|\, a \,|, \\
|\, bd \,|\, a \,|\, c \,|, & |\, bd \,|\, c \,|\, a \,|, & |\, cd \,|\, a \,|\, b \,|, & |\, cd \,|\, b \,|\, a \,|.
\end{array}
$$

To count the number of ordered selections assume that there are p types of objects with n_i indistinguishable objects of type i (for $i = 1, 2, \ldots, p$). Then the number of distinguishable permutations of length k, with up to n_i objects of type i, is the coefficient of $x^k/k!$ in the exponential generating function $\left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{n_1}}{n_1!}\right)\left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{n_2}}{n_2!}\right)\ldots\left(1 + x + \frac{x^2}{2!} + \cdots + \frac{x^{n_p}}{n_p!}\right)$.

EXAMPLE The number of 2 letter words formed from the set $\{A, A, B, B, C\}$ is the coefficient of $\frac{x^2}{2!}$ in $\left(1 + x + \frac{x^2}{2!}\right)\left(1 + x + \frac{x^2}{2!}\right)(1 + x) = 1 + 3x + 4x^2 + \ldots$, or 8. They are: $\{AA, AB, AC, BA, BB, BC, CA, CB\}$.

3.2.5 ARRANGEMENTS AND DERANGEMENTS

The number of ways to arrange n distinct objects in a row is $n!$; this is the number of permutations of n objects. For example, for the three objects $\{a, b, c\}$, the number of arrangements is $3! = 6$. These permutations are: abc, bac, cab, acb, bca, and cba.

The number of ways to arrange n objects (assuming that there are k types of objects and n_i copies of each object of type i) is the multinomial coefficient $\binom{n}{n_1, n_2, \ldots, n_k}$. For example, for the set $\{a, a, b, c\}$ the parameters are $n = 4$, $k = 3$, $n_1 = 2$, $n_2 = 1$, and $n_3 = 1$. Hence, there are $\binom{4}{2,1,1} = \frac{4!}{2!\,1!\,1!} = 12$ arrangements; they are

$$\begin{array}{cccccc} aabc, & aacb, & abac, & abca, & acab, & acba, \\ baac, & baca, & bcaa, & caab, & caba, & cbaa. \end{array}$$

A *derangement* is a permutation of objects, in which object i is not in the i^{th} location. For example, all of the derangements of $\{1, 2, 3, 4\}$ are

$$\begin{array}{ccc} 2143, & 2341, & 2413, \\ 3142, & 3412, & 3421, \\ 4123, & 4312, & 4321. \end{array}$$

The number of derangements of n elements, D_n, satisfies the recursion relation $D_n = (n-1)(D_{n-1} + D_{n-2})$ with the initial values $D_1 = 0$ and $D_2 = 1$. Hence,

$$D_n = n! \left(1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!} \right).$$

The numbers D_n are also called *sub-factorials* or *rencontres numbers*. For large values of n, $D_n/n! \sim e^{-1} \approx 0.37$. Hence more than one of every three permutations is a derangement.

n	1	2	3	4	5	6	7	8	9	10
D_n	0	1	2	9	44	265	1854	14833	133496	1334961

3.2.6 CATALAN NUMBERS

The Catalan numbers are $C_n = \frac{1}{n+1}\binom{2n}{n} \approx \frac{4^n}{n^{3/2}}$ for large n. One recurrence relation is:

$$C_n = C_0 C_{n-1} + C_1 C_{n-2} + \ldots + C_{n-1} C_0. \tag{3.2.3}$$

n	0	1	2	3	4	5	6	7	8	9	10
C_n	1	1	2	5	14	42	132	429	1,430	4,862	16,796

The Catalan numbers count many things. For example, the number of permutations of $\{1, \ldots, n\}$ that avoid the pattern "123."

EXAMPLE Given a product of n terms, the number of ways to pair terms keeping the original order is C_{n-1}. For example, with $n = 4$, there are $C_3 = 5$ ways to group 4 terms, they are: $(AB)(CD)$, $((AB)C)D$, $(A(BC))D$, $A((BC)D)$, and $A(B(CD))$.

3.2.7 PARTITIONS

A partition of the number n is a representation of n as the sum of positive integral parts. The number of partitions of n is denoted $p(n)$. For example: $5 = 4+1 = 3+2 = 3+1+1 = 2+2+1 = 2+1+1+1 = 1+1+1+1+1$ so that $p(5) = 7$.

1. The number of partitions of n into exactly m parts is equal to the number of partitions of n into parts the largest of which is exactly m; this is denoted $p_m(n)$. For example, $p_2(5) = 2$ and $p_3(5) = 2$. Note that $\sum_m p_m(n) = p(n)$.
2. The number of partitions of n into at most m parts is equal to the number of partitions of n into parts which do not exceed m.
3. The partition generating function is $$1 + \sum_{n=1}^{\infty} p(n)x^n = \prod_{k=1}^{\infty} \left(\frac{1}{1-x^k} \right)$$
4. $p(n) \sim \dfrac{1}{4n\sqrt{3}} \exp\left(\pi \sqrt{\dfrac{2n}{3}} \right)$ as $n \to \infty$

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$p(n)$	1	2	3	5	7	11	15	22	30	42	56	77	101	135	176	231

A table of $p_m(n)$ values. The columns sum to $p(n)$.

m	$n=1$	2	3	4	5	6	7	8	9	10	11
1	1	1	1	1	1	1	1	1	1	1	1
2		1	1	2	2	3	3	4	4	5	5
3			1	1	2	3	4	5	7	8	10
4				1	1	2	3	5	6	9	11
5					1	1	2	3	5	7	10
6						1	1	2	3	5	7
7							1	1	2	3	5
8								1	1	2	3
9									1	1	2

3.2.8 BELL NUMBERS

The Bell number B_n denotes the number of partitions of a set with n elements. For example: the 5 ways to partition the 3-element set $\{a,b,c\}$ are: $\{(a),(b),(c)\}$, $\{(a),(b,c)\}$, $\{(b),(a,c)\}$, $\{(c),(a,b)\}$, and $\{(a,b,c)\}$. The Bell numbers may be written in terms of the Stirling subset numbers: $B_n = \sum_{m=1}^{n} \left\{ {n \atop m} \right\}$.

n	1	2	3	4	5	6	7	8	9	10
B_n	1	2	5	15	52	203	877	4140	21147	115975

1. A generating function for Bell numbers is $\sum_{n=0}^{\infty} B_n x^n = \exp\left(e^x - 1\right) - 1$. This gives *Dobinski's formula*: $B_n = e^{-1} \sum_{m=0}^{\infty} m^n/m!$.
2. For large values of n, $B_n \sim n^{-1/2}[\lambda(n)]^{n+1/2} e^{\lambda(n)-n-1}$ where $\lambda(n)$ is defined by the relation: $\lambda(n) \log \lambda(n) = n$; see the Lambert function.

3.2.9 STIRLING CYCLE NUMBERS

The number $\left[{n \atop k}\right]$, called a *Stirling cycle number*, is the number of permutations of n symbols which have exactly k non-empty cycles. These are also called the signed Stirling numbers of the first kind.

EXAMPLE For the 4-element set $\{a, b, c, d\}$, there are $\left[{4 \atop 2}\right] = 11$ permutations containing exactly 2 cycles. They are:

$$\begin{pmatrix}1\,2\,3\,4\\2\,3\,1\,4\end{pmatrix} = (123)(4), \quad \begin{pmatrix}1\,2\,3\,4\\3\,1\,2\,4\end{pmatrix} = (132)(4), \quad \begin{pmatrix}1\,2\,3\,4\\3\,2\,4\,1\end{pmatrix} = (134)(2),$$

$$\begin{pmatrix}1\,2\,3\,4\\4\,2\,1\,3\end{pmatrix} = (143)(2), \quad \begin{pmatrix}1\,2\,3\,4\\2\,4\,3\,1\end{pmatrix} = (124)(3), \quad \begin{pmatrix}1\,2\,3\,4\\4\,1\,3\,2\end{pmatrix} = (142)(3),$$

$$\begin{pmatrix}1\,2\,3\,4\\1\,3\,4\,2\end{pmatrix} = (234)(1), \quad \begin{pmatrix}1\,2\,3\,4\\1\,4\,2\,3\end{pmatrix} = (243)(1), \quad \begin{pmatrix}1\,2\,3\,4\\2\,1\,4\,3\end{pmatrix} = (12)(34),$$

$$\begin{pmatrix}1\,2\,3\,4\\3\,4\,1\,2\end{pmatrix} = (13)(24), \quad \begin{pmatrix}1\,2\,3\,4\\4\,3\,2\,1\end{pmatrix} = (14)(23).$$

3.2.9.1 Properties of Stirling cycle numbers $\left[{n \atop k}\right]$

1. $\left[{n \atop k}\right] = (n-1)\left[{n-1 \atop k}\right] + n\left[{n-1 \atop k-1}\right]$ for $k > 0$.
2. $\left[{n \atop 0}\right] = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0 \end{cases}$
3. $\left[{n \atop 1}\right] = (n-1)!$
4. $\left[{n \atop 2}\right] = (n-1)!\, H_{n-1}$
5. $\left[{n \atop n-1}\right] = \binom{n}{2}$
6. $\sum_{k=0}^{n} \left[{n \atop k}\right] = n!$
7. $\left[{n \atop k}\right] = \sum_{m=0}^{n-k} (-1)^m \binom{n-1+m}{n-k+m}\binom{2n-k}{n-m-k}\left\{{n-m-k \atop m}\right\}$

   where $\left\{{n-m-k \atop k}\right\}$ is a Stirling subset number.
8. $\sum_{n=0}^{\infty} s(n,k)\frac{x^n}{n!} = \frac{(\log(1+x))^k}{k!}$ for $|x| < 1$. Here $s(n,k)$ is a *Stirling number of the first kind* and can be written as $s(n,k) = (-1)^{n-k}\left[{n \atop k}\right]$.
9. The factorial polynomial, defined as $x^{(n)} = x(x-1)\cdots(x-n+1)$ with $x^{(0)} = 1$, can be written as

   $$x^{(n)} = \sum_{k=0}^{n} s(n,k)x^k = s(n,1)x + s(n,2)x^2 + \ldots + s(n,n)x^n \qquad (3.2.4)$$

   For example: $x^{(3)} = x(x-1)(x-2) = 2x - 3x^2 + x^3 = \left[{3 \atop 1}\right]x - \left[{3 \atop 2}\right]x^2 + \left[{3 \atop 3}\right]x^3$.

3.2.10 STIRLING SUBSET NUMBERS

The *Stirling subset number*, $\left\{ n \atop k \right\}$, is the number of ways to partition n into k blocks. Equivalently, it is the number of ways that n distinguishable balls can be placed into k indistinguishable cells, with no cell empty. These are also called the Stirling numbers of the second kind.

EXAMPLE Placing the 4 distinguishable balls $\{a, b, c, d\}$ into 2 indistinguishable cells, so that no cell is empty can be done in $\left\{ 4 \atop 2 \right\} = 7$ ways. These are (vertical bars delineate the cells)

$$\begin{array}{llll} \mid ab \mid cd \mid, & \mid ad \mid bc \mid, & \mid ac \mid bd \mid, & \mid a \mid bcd \mid, \\ \mid b \mid acd \mid, & \mid c \mid abd \mid, & \mid d \mid abc \mid . & \end{array}$$

3.2.10.1 Properties of Stirling subset numbers $\left\{ n \atop k \right\}$

1. Stirling subset numbers are also called *Stirling numbers of the second kind*, and are denoted by $S(n, k)$.

2. $\left\{ n \atop 0 \right\} = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n \neq 0. \end{cases}$

3. $\left\{ n \atop 1 \right\} = \left\{ n \atop n \right\} = 1.$

4. $\left\{ n \atop 2 \right\} = 2^{n-1} - 1.$

5. $\displaystyle\sum_{n=0}^{\infty} \left\{ n \atop k \right\} \frac{x^n}{n!} = \frac{(e^x - 1)^k}{k!}.$

6. $\left\{ n \atop k \right\} = k \left\{ n-1 \atop k \right\} + \left\{ n-1 \atop k-1 \right\}$ for $k > 0$.

7. $\left\{ n \atop k \right\} = \displaystyle\frac{1}{k!} \sum_{m=0}^{k} (-1)^{k-m} \binom{k}{m} m^n.$

8. Ordinary powers can be expanded in terms of factorial polynomials. If $n > 0$, then

$$x^n = \sum_{k=0}^{n} \left\{ n \atop k \right\} x^{(k)} = \left\{ n \atop 0 \right\} x^{(0)} + \left\{ n \atop 1 \right\} x^{(1)} + \ldots + \left\{ n \atop n \right\} x^{(n)}. \tag{3.2.5}$$

For example,

$$\begin{aligned} x^3 &= \left\{ 3 \atop 0 \right\} x^{(0)} + \left\{ 3 \atop 1 \right\} x^{(1)} + \left\{ 3 \atop 2 \right\} x^{(2)} + \left\{ 3 \atop 3 \right\} x^{(3)} \\ &= \left\{ 3 \atop 0 \right\} + \left\{ 3 \atop 1 \right\} x + \left\{ 3 \atop 2 \right\} x(x-1) + \left\{ 3 \atop 3 \right\} x(x-1)(x-2) \\ &= 0 + x + 3(x^2 - x) + (x^3 - 3x^2 + 2x). \end{aligned} \tag{3.2.6}$$

3.2.11 GENERATING FUNCTIONS

Generating functions can solve counting problems by clever construction of the generating function and appropriate interpretation of its coefficients and the exponents.

Suppose you have 1 dime and 2 nickels. For each example below we create an appropriate generating function, then interpret the terms.

1. How many different coin combinations are there? Define
$$G_1(x) = \underbrace{\left(x^{0\cdot d} + x^{1\cdot d}\right)}_{\text{either 0 or 1 dime}} \underbrace{\left(x^{0\cdot n} + x^{1\cdot n} + x^{2\cdot n}\right)}_{\text{either 0, 1, or 2 nickels}} = \left(x^0 + x^d\right)\left(x^0 + x^n + x^{2n}\right)$$
$$= x^0 + x^n + x^{2n} + x^d + x^{d+n} + x^{d+2n}$$
We interpret the exponents, sequentially, as: "no coins", "one nickel", "two nickels", "a dime", "a dime and a nickel", "a dime and two nickels".
2. How many different combinations of coins can be obtained? Define
$$G_2(x) = \underbrace{\left(x^0 + x^1\right)}_{\text{either 0 or 1 dime}} \underbrace{\left(x^0 + x^1 + x^2\right)}_{\text{either 0, 1, or 2 nickels}} = (1+x)\left(1 + x + x^2\right)$$
$$= 1 + 2x + 2x^2 + x^3 = 1x^0 + 2x^1 + 2x^2 + x^3$$
We interpret the coefficients and exponents, sequentially, as: "1 way to have 0 coins", "2 ways to have 1 coin" (a nickel or a dime), "2 ways to have 2 coins" (a nickel and a dime or two nickels), "1 way to have 3 coins" (one dime and two nickels), "0 ways to have 4 coins", etc. Note: G_2 is G_1 with $\{n = d = 1\}$.
3. How many different amounts of money can be obtained using these coins? Define
$$G_3(x) = \underbrace{\left(x^0 + x^{10}\right)}_{\text{dimes: 0 or 10 cents}} \underbrace{\left(x^0 + x^5 + x^{10}\right)}_{\text{nickels: 0, 5, or 10 cents}} = \left(1 + x^{10}\right)\left(1 + x^5 + x^{10}\right)$$
$$= 1 + x^5 + 2x^{10} + x^{15} + x^{20} = 1x^0 + x^5 + 2x^{10} + x^{15} + x^{20}$$
We interpret the coefficients and exponents, sequentially, as: "1 way to have 0 cents" (no coins), "1 ways to have 5 cents" (a nickel), "2 ways to have 10 cents" (a dime or 2 nickels), "1 way to have 15 cents" (a dime and a nickel), "1 way to have 20 cents" (a dime and 2 nickels). G_2 is G_1 with $\{n = 5, d = 10\}$.

How many ways are there to convert a dollar into nickels and dimes? Define

$$G_4(x) = \underbrace{\left(x^0 + x^{10} + x^{20} + x^{30} + \ldots\right)}_{\text{using any number of dimes}} \underbrace{\left(x^0 + x^5 + x^{10} + x^{15} + \ldots\right)}_{\text{using any number of nickels}}$$
$$= \left(\sum_{k=0}^{\infty} x^{10k}\right)\left(\sum_{k=0}^{\infty} x^{5k}\right) = \left(\frac{1}{1-x^{10}}\right)\left(\frac{1}{1-x^5}\right)$$
$$= 1 + x^5 + 2x^{10} + 2x^{15} + 3x^{20} + 3x^{25} + \ldots + 11x^{100} + \ldots$$

So there are 11 different ways to convert a dollar into nickels and dimes. Note: There are 121 ways if quarters are also used, 242 ways if pennies and quarters are also used.

3.2.12 TABLES

3.2.12.1 Combinations $C(n,m) = \binom{n}{m}$

These tables contain the number of combinations of n distinct things taken m at a time, given by $C(n,m) = \binom{n}{m} = \frac{n!}{m!(n-m)!} = \binom{n}{n-m} = C(n, n-m)$.

	m										
n	0	1	2	3	4	5	6	7	8	9	10
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1
11	1	11	55	165	330	462	462	330	165	55	11
12	1	12	66	220	495	792	924	792	495	220	66
13	1	13	78	286	715	1287	1716	1716	1287	715	286
14	1	14	91	364	1001	2002	3003	3432	3003	2002	1001
15	1	15	105	455	1365	3003	5005	6435	6435	5005	3003
16	1	16	120	560	1820	4368	8008	11440	12870	11440	8008
17	1	17	136	680	2380	6188	12376	19448	24310	24310	19448
18	1	18	153	816	3060	8568	18564	31824	43758	48620	43758
19	1	19	171	969	3876	11628	27132	50388	75582	92378	92378
20	1	20	190	1140	4845	15504	38760	77520	125970	167960	184756

3.2.12.2 Stirling subset numbers $\left\{ {n \atop k} \right\}$

	k							
n	0	1	2	3	4	5	6	7
0	1							
1	0	1						
2	0	1	1					
3	0	1	3	1				
4	0	1	7	6	1			
5	0	1	15	25	10	1		
6	0	1	31	90	65	15	1	
7	0	1	63	301	350	140	21	1
8	0	1	127	966	1701	1050	266	28
9	0	1	255	3025	7770	6951	2646	462
10	0	1	511	9330	34105	42525	22827	5880

3.2.12.3 Permutations $P(n,m)$

These tables contain the number of permutations of n distinct things taken m at a time, given by $P(n,m) = n!/(n-m)! = n(n-1)\cdots(n-m+1)$.

	m								
n	0	1	2	3	4	5	6	7	8
0	1								
1	1	1							
2	1	2	2						
3	1	3	6	6					
4	1	4	12	24	24				
5	1	5	20	60	120	120			
6	1	6	30	120	360	720	720		
7	1	7	42	210	840	2520	5040	5040	
8	1	8	56	336	1680	6720	20160	40320	40320
9	1	9	72	504	3024	15120	60480	181440	362880
10	1	10	90	720	5040	30240	151200	604800	1814400
11	1	11	110	990	7920	55440	332640	1663200	6652800
12	1	12	132	1320	11880	95040	665280	3991680	19958400
13	1	13	156	1716	17160	154440	1235520	8648640	51891840
14	1	14	182	2184	24024	240240	2162160	17297280	121080960
15	1	15	210	2730	32760	360360	3603600	32432400	259459200

	m				
n	9	10	11	12	13
9	362880				
10	3628800	3628800			
11	19958400	39916800	39916800		
12	79833600	239500800	479001600	479001600	
13	259459200	1037836800	3113510400	6227020800	6227020800
14	726485760	3632428800	14529715200	43589145600	87178291200
15	1816214400	10897286400	54486432000	217945728000	653837184000

3.2.12.4 Stirling cycle numbers $\left[{n \atop k}\right]$

	k							
n	0	1	2	3	4	5	6	7
0	1							
1	0	1						
2	0	1	1					
3	0	2	3	1				
4	0	6	11	6	1			
5	0	24	50	35	10	1		
6	0	120	274	225	85	15	1	
7	0	720	1764	1624	735	175	21	1
8	0	5040	13068	13132	6769	1960	322	28
9	0	40320	109584	118124	67284	22449	4536	546
10	0	362880	1026576	1172700	723680	269325	63273	9450

3.3 GRAPHS

3.3.1 NOTATION

3.3.1.1 Notation for graphs

E	edge set	V	vertex set
G	graph	ϕ	incidence mapping

3.3.1.2 Graph invariants

$\mathrm{Aut}(G)$	automorphism group	$c(G)$	circumference
$d(u,v)$	distance between two vertices	$\mathrm{diam}(G)$	diameter
$\deg x$	degree of a vertex	$e(G)$	size
$\mathrm{ecc}(x)$	eccentricity	$\mathrm{gir}(G)$	girth
$\mathrm{rad}(G)$	radius	$P_G(x)$	chromatic polynomial
$Z(G)$	center	$\alpha(G)$	independence number
$\chi(G)$	chromatic number	$\chi'(G)$	chromatic index
$\delta(G)$	minimum degree	$\Delta(G)$	maximum degree
$\gamma(G)$	genus	$\tilde{\gamma}(G)$	crosscap number
$\kappa(G)$	vertex connectivity	$\lambda(G)$	edge connectivity
$\nu(G)$	crossing number	$\overline{\nu}(G)$	rectilinear crossing number
$\theta(G)$	thickness	$\Upsilon(G)$	arboricity
$\omega(G)$	clique number	$\lvert G\rvert$	order

3.3.1.3 Examples of graphs

C_n	cycle	O_n	odd graph
$\overline{K}_n$	empty graph	P_n	path
K_n	complete graph	Q_n	cube
$K_{m,n}$	complete bipartite graph	S_n	star
$K_n^{(m)}$	Kneser graphs	$T_{n,k}$	Turán graph
M_n	Möbius ladder	W_n	wheel

3.3.2 BASIC DEFINITIONS

There are two standard definitions of graphs, a general definition and a more common simplification. Except where otherwise indicated, this book uses the simplified definition, according to which a *graph* is an ordered pair (V, E) consisting of an arbitrary set V and a set E of 2-element subsets of V. Each element of V is called a *vertex* (plural *vertices*). Each element of E is called an *edge*.

According to the general definition, a *graph* is an ordered triple $G = (V, E, \phi)$ consisting of arbitrary sets V and E and an *incidence mapping* ϕ that assigns to each element $e \in E$ a non-empty set $\phi(e) \subseteq V$ of cardinality at most two. Again, the elements of V are called *vertices* and the elements of E are called *edges*. A *loop* is an edge e for which $|\phi(e)| = 1$. A graph has *multiple edges* if edges $e \neq e'$ exist for which $\phi(e) = \phi(e')$.

A (general) graph is *simple* if it has neither loops nor multiple edges. Because each edge in a simple graph can be identified with the two-element set $\phi(e) \subseteq V$, the simplified definition of graph given above is an alternative definition of a simple graph.

The word *multigraph* is used to discuss general graphs with multiple edges but no loops. Occasionally the word *pseudograph* is used to emphasize that the graphs under discussion may have both loops and multiple edges. Every graph $G = (V, E)$ considered here is *finite*, i.e., both V and E are finite sets.

Specialized graph terms include the following:

acyclic: A graph is *acyclic* if it has no cycles.

adjacency: Two distinct vertices v and w in a graph are *adjacent* if the pair $\{v, w\}$ is an edge. Two distinct edges are *adjacent* if their intersection is non-empty, i.e., if there is a vertex incident with both of them.

adjacency matrix: For an ordering $v_1, v_2, \ldots, v_n$ of the vertices of a graph $G = (V, E)$ of order $|G| = n$, there is a corresponding $n \times n$ *adjacency matrix* $A = (a_{ij})$ defined as follows:

$$a_{ij} = \begin{cases} 1 & \text{if } \{v_i, v_j\} \in E; \\ 0 & \text{otherwise.} \end{cases} \tag{3.3.1}$$

arboricity: The *arboricity* $\Upsilon(G)$ of a graph G is the minimum number of edge-disjoint spanning forests into which G can be partitioned.

automorphism: An *automorphism* of a graph is a permutation of its vertices that is an isomorphism.

automorphism group: The composition of two automorphisms is again an automorphism; with this binary operation, the automorphisms of a graph G form a group $\mathrm{Aut}(G)$ called the *automorphism group* of G.

ball: The *ball* of radius k about a vertex u in a graph is the set

$$B(u, k) = \{v \in V \mid d(u, v) \leq k\}. \tag{3.3.2}$$

See also *sphere* and *neighborhood*.

block: A *block* is a graph with no cut vertex. A *block* of a graph is a maximal subgraph that is a block.

bridge: A *bridge* is an edge in a connected graph whose removal would disconnect the graph.

cactus: A *cactus* is a connected graph, each of whose blocks is a cycle.

cage: An (r, n)-*cage* is a graph of minimal order among r-regular graphs with girth n. A $(3, n)$-cage is also called an n-cage.

center: The *center* $Z(G)$ of a graph $G = (V, E)$ consists of all vertices whose eccentricity equals the radius of G: $Z(G) = \{v \in V(G) \mid \text{ecc}(v) = \text{rad}(G)\}$. Each vertex in the center of G is called a *central vertex*.

characteristic polynomial: All adjacency matrices of a graph G have the same characteristic polynomial, which is called the *characteristic polynomial* of G.

chromatic index: The *chromatic index* $\chi'(G)$ is the least k for which there exists a proper k-coloring of the edges of G; in other words, it is the least number of matchings into which the edge set can be decomposed.

chromatic number: The *chromatic number* $\chi(G)$ of a graph G is the least k for which there exists a proper k-coloring of the vertices of G; in other words, it is the least k for which G is k-partite. See also *multipartite*.

chromatic polynomial: For a graph G of order $|G| = n$ with exactly k connected components, the *chromatic polynomial* of G is the unique polynomial $P_G(x)$ for which $P_G(m)$ is the number of proper colorings of G with m colors for each positive integer m.

circuit: A *circuit* in a graph is a trail whose first and last vertices are identical.

circulant graph: A graph G is a *circulant graph* if its adjacency matrix is a circulant matrix; that is, the rows are circular shifts of one another.

circumference: The circumference of a graph is the length of its longest cycle.

clique: A *clique* is a set S of vertices for which the induced subgraph $G[S]$ is complete.

clique number: The *clique number* $\omega(G)$ of a graph G is the largest cardinality of a clique in G.

coloring: A partition of the vertex set of a graph is called a *coloring*, and the blocks of the partition are called *color classes*. A coloring with k color classes is called a *k-coloring*. A coloring is *proper* if no two adjacent vertices belong to the same color class. See also *chromatic number* and *chromatic polynomial*.

complement: The *complement* $\overline{G}$ of a graph $G = (V, E)$ has vertex set V and edge set $\binom{V}{2} \setminus E$; that is, its edges are exactly the pairs of vertices that are not edges of G.

complete graph: A graph is *complete* if every pair of distinct vertices is an edge; K_n denotes a complete graph with n vertices.

component: A *component* of a graph is a maximal connected subgraph.

connectedness: A graph is said to be *connected* if each pair of vertices is joined by a walk; otherwise, the graph is *disconnected*. A graph is *k-connected* if it has order at least $k + 1$ and each pair of vertices is joined by k pairwise internally disjoint paths. If a graph is disconnected, then its complement is connected.

connectivity: The *connectivity* $\kappa(G)$ of G is the largest k for which G is k-connected.

contraction: To *contract* an edge $\{v, w\}$ of a graph G is to construct a new graph G' from G by removing the edge $\{v, w\}$ and identifying the vertices v and w. A graph G is *contractible* to a graph H if H can be obtained from G via the contraction of one or more edges of G.

cover: A set $S \subseteq V$ is a *vertex cover* if every edge of G is incident with some vertex in S. A set $T \subseteq E$ is an *edge cover* of a graph $G = (V, E)$ if each vertex of G is incident to at least one edge in T.

crosscap number: The *crosscap number* $\tilde{\gamma}(G)$ of a graph G is the least g for which G has an embedding in a non-orientable surface obtained from the sphere by adding g crosscaps. See also *genus*.

crossing: A *crossing* is a point lying in images of two edges of a drawing of a graph on a surface.

crossing number: The *crossing number* $\nu(G)$ of a graph G is the minimum number of crossings among all drawings of G in the plane. The *rectilinear crossing number* $\overline{\nu}(G)$ of a graph G is the minimum number of crossings among all drawings of G in the plane for which the image of each edge is a straight line segment.

cubic: A graph is a *cubic* graph if it is regular of degree 3.

cut: For each partition of the vertex V set of a graph $G = (V, E)$ into two disjoint blocks (V_1 and V_2), the set of all edges joining a vertex in V_1 to a vertex in V_2 is called a *cut*.

cut space: The *cut space* of a graph G is the subspace of the edge space of G spanned by the cut vectors.

cut vector: The *cut vector* corresponding to a cut C of a graph $G = (V, E)$ is the mapping $v\colon E \to GF(2)$ in the edge space of G

$$v(e) = \begin{cases} 1, & e \in C, \\ 0, & \text{otherwise}. \end{cases} \tag{3.3.3}$$

cut vertex: A *cut vertex* of a connected graph is a vertex whose removal, along with all edges incident with it, leaves a disconnected graph.

cycle: A *cycle* is a circuit, each pair of whose vertices other than the first and the last are distinct.

cycle space: The *cycle space* of a graph G is the subspace of the edge space of G consisting of all 1-chains with boundary 0. An indicator mapping of a set of edges with which each vertex is incident an even number of times is called a *cycle vector*. The cycle space is the span of the cycle vectors.

degree: The *degree* $\deg x$ of a vertex x in a graph is the number of vertices adjacent to it. The maximum and minimum degrees of vertices in a graph G are denoted $\Delta(G)$ and $\delta(G)$, respectively.

degree sequence: A sequence $(d_1, \ldots, d_n)$ is a *degree sequence* of a graph if there is some ordering $v_1, \ldots, v_n$ of the vertices for which d_i is the degree of v_i for each i.

diameter: The *diameter* of G is the maximum distance between two vertices of G; thus it is also the maximum eccentricity of a vertex in G.

digraph: A *digraph* is a directed graph in which each edge has a direction.

distance: The *distance* $d(u, v)$ between vertices u and v in a graph G is the minimum among the lengths of u, v-paths in G, or ∞ if there is no u, v-path.

drawing: A *drawing* of a graph G in a surface S consists of a one-to-one mapping from the vertices of G to points of S and a one-to-one mapping from the edges of G to open arcs in X so that (i) no image of an edge contains an image of some vertex, (ii) the image of each edge $\{v, w\}$ joins the images of v and w, (iii) the images of adjacent edges are disjoint, (iv) the images of two distinct edges never have more than one point in common, and (v) no point of the surface lies in the images of more than two edges.

eccentricity: The *eccentricity* $\mathrm{ecc}(x)$ of a vertex x in a graph G is the maximum distance from x to a vertex of G.

edge connectivity: The *edge connectivity* or *line connectivity* of G, denoted $\lambda(G)$, is the minimum number of edges whose removal results in a disconnected or trivial graph.

edge space: The *edge space* of a graph $G = (V, E)$ is the vector space of all mappings from E to the two-element field $GF(2)$. Elements of the edge space are called 1-*chains*.

embedding: An *embedding* of a graph G in a topological space X consists of an assignment of the vertices of G to distinct points of X and an assignment of the edges of G to disjoint open arcs in X so that no arc representing an edge contains some point representing a vertex and so that each arc representing an edge joins the points representing the vertices incident with the edge. See also *drawing*.

end vertex: A vertex of degree 1 in a graph is called an *end vertex*.

Eulerian circuits and trails: A trail or circuit that includes every edge of a graph is said to be *Eulerian*, and a graph is *Eulerian* if it has an Eulerian circuit.

even: A graph is *even* if the degree of every vertex is even.

factor of a graph: A *factor* of a graph G is a spanning subgraph of G. A factor in which every vertex has the same degree k is called a k-factor. If G_1, G_2, ..., G_k ($k \geq 2$) are edge-disjoint factors of the graph G, and if $\bigcup_{i=1}^{k} E(G_i) = E(G)$, then G is said to be *factored* into G_1, G_2, ..., G_k and we write $G = G_1 \oplus G_2 \oplus \cdots \oplus G_k$.

forest: A *forest* is an acyclic simple graph; see also *tree*.

genus: The *genus* $\gamma(G)$ (plural form *genera*) of a graph G is the least g for which G has an embedding in an orientable surface of genus g. See also *crosscap number*.

girth: The *girth* $\mathrm{gir}(G)$ of a graph G is the minimum length of a cycle in G, or ∞ if G is acyclic.

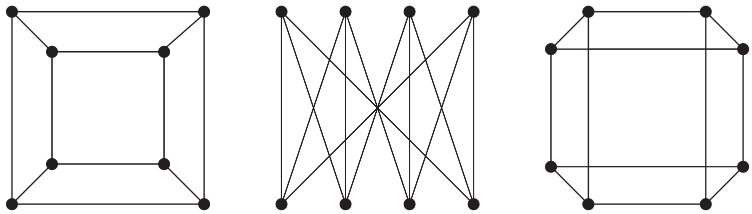
Hamiltonian cycles and paths: A path or cycle through all the vertices of a graph is said to be *Hamiltonian*. A graph is *Hamiltonian* if it has a Hamiltonian cycle.

homeomorphic graphs: Two graphs are *homeomorphic* to one another if there is a third graph of which each is a subdivision.

identification of vertices: To *identify* vertices v and w of a graph G is to construct a new graph G' from G by removing the vertices v and w and all the edges of G incident with them and introducing a new vertex u and new edges joining u to each vertex that was adjacent to v or to w in G. See also *contraction*.

incidence: A vertex v and an edge e are *incident* with one another if $v \in e$.

FIGURE 3.2
Three graphs that are isomorphic.



incidence matrix: For an ordering $v_1, v_2, \ldots, v_n$ of the vertices and an ordering $e_1, e_2, \ldots, e_m$ of the edges of a graph $G = (V, E)$ with order $|G| = n$ and size $e(G) = m$, there is a corresponding $n \times m$ *incidence matrix* $B = (b_{ij})$ defined as follows:

$$b_{ij} = \begin{cases} 1, & \text{if } v_i \text{ and } e_j \text{ are incident,} \\ 0, & \text{otherwise.} \end{cases} \tag{3.3.4}$$

independence number: The *independence number* $\alpha(G)$ of a graph $G = (V, E)$ is the largest cardinality of an independent subset of V.

independent set: A set $S \subseteq V$ is said to be *independent* if the induced subgraph $G[S]$ is empty. See also *matching*.

internally disjoint paths: Two paths in a graph with the same initial vertex v and terminal vertex w are *internally disjoint* if they have no internal vertex in common.

isolated vertex: A vertex is *isolated* if it is adjacent to no other vertex.

isomorphism: An *isomorphism* between graphs $G = (V_G, E_G)$ and $H = (V_H, E_H)$ is a bijective mapping $\psi\colon V_G \to V_H$ for which $\{x, y\} \in E_G$ if and only if $\{\psi(x), \psi(y)\} \in E_H$. If there is an isomorphism between G and H, then G and H are said to be *isomorphic* to one another; this is denoted as $G \cong H$. Figure 3.2 contains three graphs that are isomorphic.

labeled graph: Graph theorists sometimes speak loosely of *labeled graphs* of order n and *unlabeled graphs* of order n to distinguish between graphs with a fixed vertex set of cardinality n and the family of isomorphism classes of such graphs. Thus, one may refer to *labeled graphs* to indicate an intention to distinguish between any two graphs that are *distinct* (i.e., have different vertex sets and/or different edge sets). One may refer to *unlabeled graphs* to indicate the intention to view any two distinct but isomorphic graphs as the 'same' graph, and to distinguish only between non-isomorphic graphs.

matching: A *matching* in a graph is a set of edges, no two having a vertex in common. A *maximal matching* is a matching that is not a proper subset of any other matching. A *maximum matching* is a matching of greatest cardinality. For a matching M, an M*-alternating path* is a path whose every other edge belongs to M, and an M*-augmenting* path is an M-alternating path whose first and last edges do not belong to M. A matching *saturates* a vertex if the vertex belongs to some edge of the matching.

monotone graph property: A property $\mathcal{P}$ that a graph may or may not enjoy is said to be *monotone* if, whenever H is a graph enjoying $\mathcal{P}$, every supergraph G of H with $|G| = |H|$ also enjoys $\mathcal{P}$.

multipartite graph: A graph is k-*partite* if its vertex set can be partitioned into k disjoint sets called *color classes* in such a way that every edge joins vertices in two different color classes (see also *coloring*). A two-partite graph is called *bipartite*.

neighbor: Adjacent vertices v and w in a graph are said to be *neighbors* of one another.

neighborhood: The sphere $S(x, 1)$ is called the *neighborhood* of x, and the ball $B(x, 1)$ is called the *closed neighborhood* of x.

order: The *order* $|G|$ of a graph $G = (V, E)$ is the number of vertices in G; in other words, $|G| = |V|$.

path: A *path* is a walk whose vertices are distinct.

perfect graph: A graph is *perfect* if $\chi(H) = \omega(H)$ for all induced subgraphs H of G.

planarity: A graph is *planar* if it has an embedding in the plane.

radius: The *radius* $\mathrm{rad}(G)$ of a graph G is the minimum vertex eccentricity in G.

regularity: A graph is k-*regular* if each of its vertices has degree k. A graph is *strongly regular* with parameters (k, λ, μ) if (i) it is k-regular, (ii) every pair of adjacent vertices has exactly λ common neighbors, and (iii) every pair of non-adjacent vertices has exactly μ common neighbors. A graph $G = (V, E)$ of order $|G| \geq 3$ is called *highly regular* if there exists an $n \times n$ matrix $C = (c_{ij})$, where $2 \leq n < |G|$, called a *collapsed adjacency matrix*, so that, for each vertex v of G, there is a partition of V into n subsets $V_1 = \{v\}$, $V_2, \ldots, V_n$ so that every vertex $y \in V_j$ is adjacent to exactly c_{ij} vertices in V_i. Every highly regular graph is regular.

rooted graph: A *rooted graph* is an ordered pair (G, v) consisting of a graph G and a distinguished vertex v of G called the *root*.

self-complementary: A graph is *self-complementary* if it is isomorphic to its complement.

similarity: Two vertices u and v of a graph G are *similar* (in symbols $u \sim v$) if there is an automorphism α of G for which $\alpha(u) = v$. Similarly, two edges (u, v) and (a, b) in the graph G are *similar* if an automorphism α of G exists for which $\{\alpha(u), \alpha(v)\} = \{a, b\}$.

size: The *size* $e(G)$ of a graph $G = (V, E)$ is the number of edges of G, that is, $e(G) = |E|$.

spectrum: The *spectrum* of a graph G is the spectrum of its characteristic polynomial, i.e., the non-decreasing sequence of $|G|$ eigenvalues of the characteristic polynomial of G. Since adjacency matrices are real symmetric, their spectrum is real.

sphere: The *sphere* of radius k about a vertex u is the set

$$S(u,k) = \{v \in V \mid d(u,v) = k\}. \tag{3.3.5}$$

See also *ball* and *neighborhood*.

subdivision: To *subdivide* an edge $\{v,w\}$ of a graph G is to construct a new graph G' from G by removing the edge $\{v,w\}$ and introducing new vertices x_i and new edges $\{v,x_1\}$, $\{x_k,w\}$ and $\{x_i,x_{i+1}\}$ for $1 \le i < k$. A *subdivision* of a graph is a graph obtained by subdividing one or more edges of the graph.

subgraph: A graph $H = (V_H, E_H)$ is a *subgraph* of a graph $G = (V_G, E_G)$ (in symbols, $H \preceq G$), if $V_H \subseteq V_G$ and $E_H \subseteq E_G$. In that case, G is a *supergraph* of H (in symbols, $G \succeq H$). If $V_H = V_G$, then H is called a *spanning subgraph* of G. For each set $S \subseteq V_G$, the subgraph $G[S]$ of G *induced* by S is the unique subgraph of G with vertex set S for which every edge of G incident with two vertices in S is also an edge of $G[S]$.

symmetry: A graph is *vertex symmetric* if every pair of vertices is similar. A graph is *edge symmetric* if every pair of edges is similar. A graph is *symmetric* if it is both vertex and edge symmetric.

2-switch: For vertices v, w, x, y in a graph G for which $\{v,w\}$ and $\{x,y\}$ are edges, but $\{v,y\}$ and $\{x,w\}$ are not edges, the construction of a new graph G' from G via the removal of edges $\{v,w\}$ and $\{x,y\}$ together with the insertion of the edges $\{v,y\}$ and $\{x,w\}$ is called a 2-*switch*.

thickness: The *thickness* $\theta(G)$ of a graph G is the least k for which G is a union of k planar graphs.

trail: A *trail* in a graph is a walk whose edges are distinct.

tree: A *tree* is a connected forest, i.e., a connected acyclic graph. A spanning subgraph of a graph G that is a tree is called a *spanning tree* of G.

triangle: A 3-cycle is called a triangle.

trivial graph: A *trivial* graph is a graph with exactly one vertex and no edges.

vertex space: The *vertex space* of a graph G is the vector space of all mappings from V to the two-element field $GF(2)$. The elements of the vertex space are called 0-*chains*.

walk: A *walk* in a graph is an alternating sequence $v_0, e_1, v_1, \ldots, e_k, v_k$ of vertices v_i and edges e_i for which e_i is incident with v_{i-1} and with v_i for each i. Such a walk is said to have *length* k and to *join* v_0 and v_k. The vertices v_0 and v_k are called the *initial vertex* and *terminal vertex* of the walk; the remaining vertices are called *internal vertices* of the walk.

3.3.3 CONSTRUCTIONS

3.3.3.1 Operations on graphs

For graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, there are several binary operations that yield a new graph from G_1 and G_2. The table below gives the names of some of those operations and the orders and sizes of the resulting graphs.

Operation producing G		Order $\lvert G\rvert$	Size $e(G)$
Composition	$G_1[G_2]$	$= \lvert G_1\rvert \cdot \lvert G_2\rvert$	$= \lvert G_1\rvert\, e(G_2) + \lvert G_2\rvert\, e(G_1)$
Conjunction	$G_1 \wedge G_2$	$= \lvert G_1\rvert \cdot \lvert G_2\rvert$	
Edge sum[a]	$G_1 \oplus G_2$	$= \lvert G_1\rvert = \lvert G_2\rvert$	$\leq \ (e(G_1) + e(G_2))$
Join	$G_1 + G_2$	$= \lvert G_1\rvert + \lvert G_2\rvert$	$= e(G_1) + e(G_2) + \lvert G_1\rvert \cdot \lvert G_2\rvert$
Product	$G_1 \times G_2$	$= \lvert G_1\rvert \cdot \lvert G_2\rvert$	$= \lvert G_1\rvert e(G_2) + \lvert G_2\rvert e(G_1)$
Union	$G_1 \cup G_2$	$= \lvert G_1\rvert + \lvert G_2\rvert$	$= e(G_1) + e(G_2)$

[a] When applicable.

composition: For graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the *composition* $G = G_1[G_2]$ is the graph with vertex set $V_1 \times V_2$ whose edges are (1) the pairs $\{(u,v),(u,w)\}$ with $u \in V_1$ and $\{v,w\} \in E_2$ and (2) the pairs $\{(t,u),(v,w)\}$ for which $\{t,v\} \in E_1$.

conjunction: The conjunction $G_1 \wedge G_2$ of two graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ is the graph $G_3 = (V_3, E_3)$ for which $V_3 = V_1 \times V_2$ and for which vertices $\mathbf{e}_1 = (u_1, u_2)$ and $\mathbf{e}_2 = (v_1, v_2)$ in V_3 are adjacent in G_3 if and only if u_1 is adjacent to v_1 in G_1 and u_2 is adjacent to v_2 in G_2.

edge difference: For graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ with the same vertex set V, the *edge difference* $G_1 - G_2$ is the graph with vertex set V and edge set $E_1 \setminus E_2$.

edge sum: For graphs $G_1 = (V, E_1)$ and $G_2 = (V, E_2)$ with the same vertex set V, the *edge sum* of G_1 and G_2 is the graph $G_1 \oplus G_2$ with vertex set V and edge set $E_1 \cup E_2$. Sometimes the edge sum is denoted $G_1 \cup G_2$.

join: For graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 \cap V_2 = \emptyset$, the *join* $G_1 + G_2 = G_2 + G_1$ is the graph obtained from the union of G_1 and G_2 by adding edges joining each vertex in V_1 to each vertex in V_2.

power: For a graph $G = (V, E)$, the k^{th} *power* G^k is the graph with the same vertex set V whose edges are the pairs $\{u,v\}$ for which $d(u,v) \leq k$ in G. The *square* of G is G^2.

product: For graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$, the *product* $G_1 \times G_2$ has vertex set $V_1 \times V_2$; its edges are all of the pairs $\{(u,v),(u,w)\}$ for which $u \in V_1$ and $\{v,w\} \in E_2$ and all of the pairs $\{(t,v),(u,v)\}$ for which $\{t,u\} \in E_1$ and $v \in V_2$.

union: For graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ with $V_1 \cap V_2 = \emptyset$, the *union* of G_1 and G_2 is the graph $G_1 \cup G_2 = (V_1 \cup V_2, E_1 \cup E_2)$. The union is sometimes called the *disjoint union* to distinguish it from the *edge sum*.

3.3.3.2 Graphs described by one parameter

complete graph, K_n: A complete graph of order n is a graph isomorphic to the graph K_n with vertex set $\{1, 2, \ldots, n\}$ whose every pair of vertices is an edge. The graph K_n has size $\binom{n}{2}$ and is Hamiltonian. If G is a graph g of order n, then $K_n = G \oplus \overline{G}$.

cube, Q_n: An *n-cube* is a graph isomorphic to the graph Q_n whose vertices are the 2^n binary n-vectors and whose edges are the pairs of vectors that differ in exactly one place. It is an n-regular bipartite graph of order 2^n and size $n2^{n-1}$. An equivalent recursive definition is $Q_1 = K_2$ and $Q_n = Q_{n-1} \times K_2$.

cycle, C_n: A *cycle* of order n is a graph isomorphic to the graph C_n with vertex set $\{0, 1, \ldots, n-1\}$ whose edges are the pairs $\{v_i, v_{i+1}\}$ with $0 \leq i < n$ and arithmetic modulo n. The cycle C_n has size n and is Hamiltonian.
The graph C_n is a special case of a circulant graph. The graph C_3 is called a *triangle*, the graph C_4 is called a *square*.

empty graph: A graph is *empty* if it has no edges; $\overline{K}_n$ denotes an empty graph of order n.

Kneser graphs, $K_n^{(m)}$: For $n \geq 2m$, the *Kneser graph* $K_n^{(m)}$ is the complement of the intersection graph of the m-subsets of an n-set. The *odd graph* O_m is the Kneser graph $K_{2m+1}^{(m)}$. The *Petersen graph* is the odd graph $O_2 = K_5^{(2)}$.

ladder: A *ladder* is a graph of the form $P_n \times P_2$. The *Möbius ladder* M_n is the graph obtained from the ladder $P_n \times P_2$ by joining the opposite end vertices of the two copies of P_n.

path, P_n: A *path* of order n is a graph isomorphic to the graph P_n whose vertex set is $\{1, \ldots, n\}$ and whose edges are the pairs $\{v_i, v_{i+1}\}$ with $1 \leq i < n$. A path of order n has size $n-1$ and is a tree.

star, S_n: A *star* of order n is a graph isomorphic to the graph $S_n = K_{1,n}$. It has a vertex cover consisting of a single vertex, its size is n, and it is a complete bipartite graph and a tree.

wheel, W_n: The wheel W_n of order $n \geq 4$ consists of a cycle of order $n-1$ and an additional vertex adjacent to every vertex in the cycle. Equivalently, $W_n = C_{n-1} + K_1$. This graph has size $2(n-1)$.

3.3.3.3 Graphs described by two parameters

complete bipartite graph, $K_{n,m}$: The complete bipartite graph $K_{n,m}$ is the graph $\overline{K}_n + \overline{K}_m$. Its vertex set can be partitioned into two color classes of cardinalities n and m, respectively, so that each vertex in one color class is adjacent to every vertex in the other color class. The graph $K_{n,m}$ has order $n+m$ and size nm.

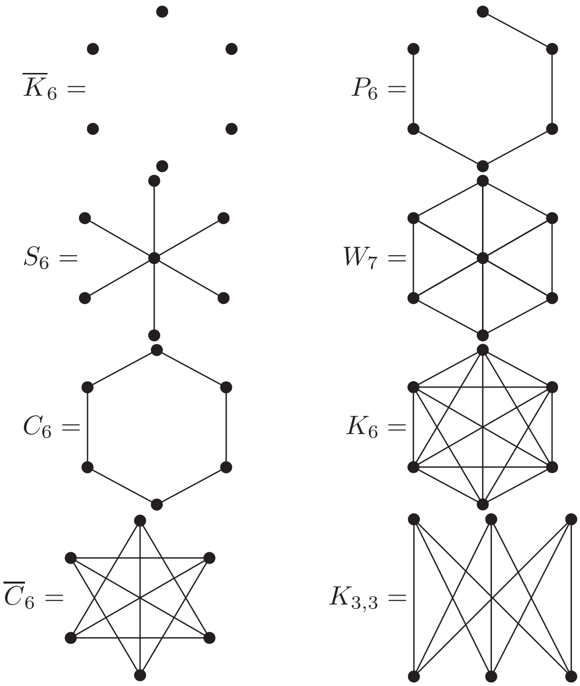
planar mesh: A graph of the form $P_n \times P_m$ is called a *planar mesh*.

prism: A graph of the form $C_m \times P_n$ is called a *prism*.

toroidal mesh: A graph of the form $C_m \times C_n$ with $m \geq 2$ and $n \geq 2$.

Turán graph, $T_{n,k}$: The *Turán graph* $T_{n,k}$ is the complete k-partite graph in which the cardinalities of any two color classes differ by, at most, one. It has $n - k\lfloor n/k \rfloor$ color classes of cardinality $\lfloor n/k \rfloor + 1$ and $k - n + k\lfloor n/k \rfloor$ color classes of cardinality $\lfloor n/k \rfloor$. Note that $\omega(T_{n,k}) = k$.

FIGURE 3.3
Examples of graphs with 6 or 7 vertices.



3.3.3.4 Graphs described by three or more parameters

Cayley graph: For a group Γ and a set X of generators of Γ, the *Cayley graph* of the pair (Γ, X) is the graph with vertex set Γ in which $\{\alpha, \beta\}$ is an edge if either $\alpha^{-1}\beta \in X$ or $\beta^{-1}\alpha \in X$.

complete multipartite graph, $K_{n_1,n_2,\ldots,n_k}$: The *complete k-partite graph* $K_{n_1,n_2,\ldots,n_k}$ is the graph $\overline{K}_{n_1} + \cdots + \overline{K}_{n_k}$. It is a k-partite graph with color classes V_i of cardinalities $|V_i| = n_i$ for which every pair of vertices in two distinct color classes is an edge. The graph $K_{n_1,n_2,\ldots,n_k}$ has order $\sum_{i=1}^{k} n_k$ and size $\sum_{1\le i<j\le k} n_i n_j$.

intersection graph: For a family $F = \{S_1, \ldots, S_n\}$ of subsets of a set S, the *intersection graph* of F is the graph with vertex set F in which $\{S_i, S_j\}$ is an edge if and only if $S_i \cap S_j \neq \emptyset$. Each graph G is an intersection graph of some family of subsets of a set of cardinality at most $\lfloor |G|^2/4 \rfloor$.

interval graph: An *interval graph* is an intersection graph of a family of intervals on the real line.

3.3.4 FUNDAMENTAL RESULTS

3.3.4.1 Walks and connectivity

1. Every x, y walk includes all the edges of some x, y path.
2. Some path in G has length $\delta(G)$.
3. Connectivity is a monotone graph property. If more edges are added to a connected graph, the new graph is itself connected.
4. A graph is disconnected if and only if it is the union of two graphs.
5. Every vertex of a graph lies in at least one block.
6. For every graph G, $0 \leq \kappa(G) \leq |G| - 1$.
7. For all integers a, b, c with $0 < a \leq b \leq c$, a graph G exists with $\kappa(G) = a$, $\lambda(G) = b$, and $\delta(G) = c$.
8. For any graph G, $\kappa(G) \leq \lambda(G) \leq \delta(G)$.
9. *Menger's theorem*: Suppose that G is a connected graph of order greater than k. Then G is k-connected if and only if it is impossible to disconnect G by removing fewer than k vertices, and G is k-edge connected if and only if it is impossible to disconnect G by removing fewer than k edges.
10. If G is a connected graph with a bridge, then $\lambda(G) = 1$. If G has order n and is r-regular with $r \geq n/2$, then $\lambda(G) = r$.

3.3.4.2 Circuits and cycles

1. *Euler's theorem*: A multigraph is Eulerian if and only if it is connected and even.

2. If G is Hamiltonian, and if G' is obtained from G by removing a non-empty set S of vertices, then the number of components of G' is at most $|S|$.

3. *Ore's theorem*: If G is a graph for which $\deg v + \deg w \geq |G|$ whenever v and w are non-adjacent vertices, then G is Hamiltonian.

4. *Dirac's theorem*: If G is a graph of order $|G| \geq 3$ and $\deg v \geq |G|/2$ for each vertex v, then G is Hamiltonian.

5. (Erdös–Chvátal) If $\alpha(G) \leq \kappa(G)$, then G is Hamiltonian.

6. Every 4-connected planar graph is Hamiltonian.

7. The following table indicates which graphs are Eulerian or Hamiltonian:

Graph	Is it Eulerian?	Is it Hamiltonian?
C_n	yes	yes, for $n \geq 1$
K_n	yes, for odd n	yes, for $n \geq 3$
$K_{m,n}$	yes, for m and n both even	yes, for $m = n$
Q_n	yes, for n even	yes, for $n \geq 2$
W_n	no	yes, for $n \geq 2$

3.3.4.3 Trees

1. A graph is a tree if and only if
   (a) it is acyclic and has size $|G| - 1$.
   (b) it is connected and has size $|G| - 1$.
   (c) each of its edges is a bridge.
   (d) each vertex of degree greater than 1 is a cut vertex.
   (e) each pair of its vertices is joined by exactly one path.
2. Every connected graph has a spanning tree.
3. Every tree of order greater than 1 has at least two end vertices.
4. The center of a tree consists of one vertex or two adjacent vertices.
5. For a graph G, every tree with at most $\delta(G)$ edges is a subgraph of G.
6. *Kirchhoff matrix-tree theorem*: Let G be a connected graph and let A be an adjacency matrix for G. Obtain a matrix M from $-A$ by replacing each term a_{ii} on the main diagonal with $\deg v_i$. Then all cofactors of M have the same value, which is the number of spanning trees of G.
7. *Nash–Williams arboricity theorem*: For a graph G and for each $n \leq |G|$, define $e_n(G) = \max\{e(H) : H \preceq G, \text{ and } |H| = n\}$. Then $\Upsilon(G) = \max_n \lceil e_n/(n-1) \rceil$.

3.3.4.4 Cliques and independent sets

1. A set $S \subseteq V$ is a vertex cover if and only if $V \setminus S$ is an independent set.
2. *Turán's theorem*: If $|G| = n$ and $\omega(G) \leq k$, then $e(G) \leq e(T_{n,k})$.
3. *Ramsey's theorem*: For all positive integers k and l, there is a least integer $R(k,l)$ for which every graph of order at least $R(k,l)$ has either a clique of cardinality k or an independent set of cardinality l. For $k \geq 2$ and $l \geq 2$, $R(k,l) \leq R(k,l-1) + R(k-1,l)$.

$R(k,l)$	$l=1$	2	3	4	5	6	7
$k=1$	1	1	1	1	1	1	1
2	1	2	3	4	5	6	7
3	1	3	6	9	14	18	23

3.3.4.5 Colorings and partitions

1. $P_G(x) = P_{G-e}(x) - P_{G\backslash e}(x)$.
2. (Appel–Haken) *Four-color theorem*: $\chi(G) \leq 4$ for every planar graph G.
3. (König) If G is bipartite, then $\chi'(G) = \Delta(G)$.
4. *Brooks' theorem*: If G is a connected graph that is neither a complete graph nor a cycle of odd length, then $\chi(G) \leq \Delta(G)$.
5. *Szekeres–Wilf theorem*: For every graph $G = (V, E)$,
$$\chi(G) \leq 1 + \max_{S \subseteq V} \delta(G[S]).$$
6. *Vizing's theorem*: For every graph G, $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$.
7. Every graph G is k-partite for some k; in particular, G is $|G|$-partite.

8. Every graph G has a bipartite subgraph H for which $e(H) \geq e(G)/2$.
9. For all positive integers g and c, a graph G exists such that $\chi(G) \geq c$ and $\text{gir}(G) \geq g$.
10. *Nordhaus–Gaddum bounds*: For every graph G,

$$2\sqrt{|G|} \leq \chi(G) + \chi(\bar{G}) \leq |G| + 1, \text{ and}$$
$$|G| \leq \chi(G)\chi(\bar{G}) \leq \left(\frac{|G|+1}{2}\right)^2.$$

11. The following table gives the chromatic numbers and chromatic polynomials of various graphs:

G	$\chi(G)$	$P_G(x)$
K_n	n	$x(x-1)\cdots(x-n+1)$
$\overline{K}_n$	1	x^n
T_n	2	$x(x-1)^{n-1}$
P_n	2	$x(x-1)^{n-1}$
C_4	2	$x(x-1)(x^2-3x+3)$

12. The following table gives the chromatic numbers and edge-chromatic numbers (i.e., edges are colored instead of vertices) of various graphs:

G	$\chi(G)$	$\chi_1(G)$
C_n with n even, $n \geq 2$	2	2
C_n with n odd, $n \geq 3$	3	3
K_n with n even, $n \geq 2$	n	$n-1$
K_n with n odd, $n \geq 3$	n	n
$K_{m,n}$ with $m, n \geq 1$	2	$\max(m,n)$
$K_{m_1,\ldots,m_k}$ with $m_i \geq 1$	k	$\max(m_1,\ldots,m_k)$
P_n	2	2
Petersen graph	3	4
W_n with n even, $n \geq 2$	3	n
W_n with n odd, $n \geq 3$	4	n

13. For each graph G of order $|G| = n$ and size $e(G) = m$ with exactly k components, the chromatic polynomial is of the form

$$P_G(x) = \sum_{i=0}^{n-k} (-1)^i a_i x^{n-i}, \tag{3.3.6}$$

with $a_0 = 1$, $a_1 = m$ and every a_i positive.
14. Not every polynomial is a chromatic polynomial. For example $P(x) = x^4 - 4x^3 + 3x^2$ is not a chromatic polynomial.
15. Sometimes a class of chromatic polynomials can only come from a specific class of graphs. For example:

    (a) If $P_G(x) = x^n$, then $G = \overline{K}_n$.
    (b) If $P_G(x) = (x)_n$, then $G = K_n$.

3.3.4.6 Distance

1. A metric space (X, d) is the metric space associated with a connected graph with vertex set X if and only if it satisfies two conditions: (i) $d(u, v)$ is a non-negative integer for all $u, v \in X$, and (ii) whenever $d(u, v) \geq 2$, some element of X lies between u and v. The edges of the graph are the pairs $\{u, v\} \subseteq X$ for which $d(u, v) = 1$. (In an arbitrary metric space (X, d), a point $v \in X$ is said to lie *between* distinct points $u \in X$ and $w \in X$ if it satisfies the *triangle equality* $d(u, w) = d(u, v) + d(v, w)$.)
2. If $G = (V, E)$ is connected, then distance is always finite, and d is a metric on V. Note that $\deg(x) = |S(x, 1)|$.
3. *Moore bound*: For every connected graph G,

$$|G| \leq 1 + \Delta(G) \sum_{i=1}^{\mathrm{diam}(G)} (\Delta(G) - 1)^i. \tag{3.3.7}$$

A graph for which the Moore bound holds exactly is called a *Moore graph* with parameters $(|G|, \Delta(G), \mathrm{diam}(G))$. Every Moore graph is regular. If G is a Moore graph with parameters (n, r, d), then $(n, r, d) = (n, n-1, 1)$ (in which case G is complete), $(n, r, d) = (2m+1, 2, m)$ (in which case G is a $(2m+1)$-cycle), or $(n, r, d) \in \{(10, 3, 2), (50, 7, 2), (3250, 57, 2)\}$.

3.3.4.7 Drawings, embeddings, planarity, and thickness

1. Every graph has an embedding in $\mathbb{R}^3$ for which the arcs representing edges are all straight line segments. Such an embedding can be constructed by using distinct points on the curve $\{(t, t^2, t^3) : 0 \leq t \leq 1\}$ as representatives for the vertices.
2. Every planar graph can be embedded in the plane so that every edge is a straight line segment; this is a *Fary embedding*.
3. For $n \geq 2$,

$$\gamma(Q_n) = (n-4)2^{n-3} + 1 \qquad \text{and} \qquad \tilde{\gamma}(Q_n) = (n-4)2^{n-2} + 2.$$

4. For $r, s \geq 2$,

$$\gamma(K_{r,s}) = \left\lceil \frac{(r-2)(s-2)}{4} \right\rceil \qquad \text{and} \qquad \tilde{\gamma}(K_{r,s}) = \left\lceil \frac{(r-2)(s-2)}{2} \right\rceil.$$

5. For $n \geq 3$,

$$\gamma(K_n) = \left\lceil \frac{(n-3)(n-4)}{12} \right\rceil \qquad \text{and} \qquad \tilde{\gamma}(K_n) = \left\lceil \frac{(n-3)(n-4)}{6} \right\rceil.$$

6. *Heawood map coloring theorem*: The greatest chromatic number among graphs of genus n is

$$\max\{\chi(G) \mid \gamma(G) = n\} = \left\lceil \frac{7 + \sqrt{1 + 48n}}{2} \right\rceil.$$

7. *Kuratowski's theorem*: A graph is planar if and only if it has no subgraph homeomorphic to K_5 or $K_{3,3}$.
8. A graph is planar if and only if it does not have a subgraph contractible to K_5 or $K_{3,3}$.
9. The graph K_n is non-planar if and only if $n \geq 5$.
10. The four-color theorem states that any planar graph is four colorable.
11. For every graph G of order $|G| \geq 3$, $\theta(G) \geq \left\lceil \dfrac{e(G)}{3|G|-6} \right\rceil$.
12. The complete graphs K_9 and K_{10} have thickness 3; for $n \notin \{9, 10\}$,

$$\theta(K_n) = \left\lfloor \frac{n+7}{6} \right\rfloor. \tag{3.3.8}$$

13. The n-cube has thickness $\theta(Q_n) = \lfloor n/4 \rfloor + 1$.
14. For every planar graph G, $\nu(G) = \overline{\nu}(G)$. That equality does not hold for all graphs: $\nu(K_8) = 18$, and $\overline{\nu}(K_8) = 19$.

3.3.4.8 Vertex degrees

1. *Handshaking lemma*: For every graph G, $\sum_{v \in V} \deg v = 2e(G)$.
2. Every 2-switch preserves the degree sequence.
3. If G and H have the same degree sequence, then H can be obtained from G via a sequence of 2-switches.
4. *Havel–Hakimi theorem*: The values $\{d_1, d_2, \dots, d_n\}$ with $d_1 \geq d_2 \geq \cdots \geq d_n > 0$ are a degree sequence if and only if the sequence obtained by deleting d_1 and subtracting 1 from each of the next d_1 largest values (i.e., $\{d_2 - 1, d_3 - 1, \dots, d_{d_1+1} - 1, d_{d_1+2}, \dots, d_n\}$) is a degree sequence.
5. *Erdos–Gallai theorem*: The values $\{d_1, d_2, \dots, d_n\}$ are a degree sequence if and only if the sum of vertex degrees is even and the sequence has the property, for each integer $r \leq n-1$: $\displaystyle\sum_{i=1}^{r} d_i \leq r(r-1) + \sum_{i=r+1}^{n} \min(r, d_i)$.

3.3.4.9 Algebraic methods

1. The bipartite graphs $K_{n,n}$ are circulant graphs.
2. For a graph G with exactly k connected components, the cycle space has dimension $e(G) - |G| + k$, and the cut space has dimension $|G| - k$.
3. In the k^{th} power $A^k = (a_{ij}^k)$ of the adjacency matrix, each entry a_{ij}^k is the number of v_i, v_j walks of length k.
4. The incidence matrix of a graph G is totally unimodular if and only if G is bipartite. (Unimodular matrices form a group under matrix multiplication.)
5. The smallest graph that is vertex symmetric, but is not edge symmetric, is the prism $K_3 \times K_2$. The smallest graph that is edge symmetric, but is not vertex symmetric, is $S_2 = P_3 = K_{1,2}$.
6. The spectrum of a disconnected graph is the union of the spectra of its components.
7. The sum of the eigenvalues in the spectrum of a graph is zero.

8. The number of distinct eigenvalues in the spectrum of a connected graph is greater than the diameter of the graph.
9. The largest eigenvalue in the spectrum of a graph G is, at most, $\Delta(G)$, with equality if and only if G is regular.
10. (Hoffman) If G is a connected graph of order n with spectrum $\lambda_1 \geq \cdots \geq \lambda_n$, then $\chi(G) \geq 1 - \lambda_1/\lambda_n$.
11. (Wilf) If G is a connected graph and its largest eigenvalue is λ, then $\chi(G) \leq 1 + \lambda$. Moreover, equality holds if and only if G is a complete graph or a cycle of odd length.
12. *Integrality condition*: If G is a strongly regular graph with parameters (k, λ, μ), then the quantities

$$\frac{1}{2}\left(|G| - 1 \pm \frac{(|G|-1)(\mu-\lambda) - 2k}{\sqrt{(\mu-\lambda)^2 + 4(k-\mu)}}\right) \tag{3.3.9}$$

are non-negative integers.
13. The following table gives the automorphism groups of various graphs:

G	$\mathrm{Aut}(G)$
C_n for $n \geq 3$	D_n
$\overline{K}_n$	S_n
K_n	S_n

14. A graph and its complement have the same group; $\mathrm{Aut}(G) = \mathrm{Aut}(\overline{G})$.
15. *Frucht's theorem*: Every finite group is the automorphism group of some graph.
16. If G and G' are edge isomorphic, then G and G' are not necessarily required to be isomorphic. For example, the graphs C_3 and S_3 are edge isomorphic, but not isomorphic. (Two graphs are edge isomorphic if edges are incident in one graph if and only if the corresponding edges are incident in the other graph.)
17. If the graph G has order n, then the order of its automorphism group $|\mathrm{Aut}(G)|$ is a divisor of $n!$. The order of the automorphism group equals $n!$ if and only if $G \simeq K_n$ or $G \simeq \overline{K}_n$.

3.3.4.10 Enumeration

1. The number of labeled graphs of order n is $2^{\binom{n}{2}}$.
2. The number of labeled graphs of order n and size m is $\binom{\binom{n}{2}}{m}$.
3. The number of different ways in which a graph G of order n can be labeled is $n!/|\mathrm{Aut}(G)|$.
4. *Cayley's formula*: The number of labeled trees of order n is n^{n-2}.
5. The number of labeled trees of order n with exactly t end vertices is $\frac{n!}{t!}S(n-2, n-t)$ for $2 \leq t \leq n-1$, where $S(,)$ is a Stirling number of the second kind.
6. Note the generating functions
   (a) $T(x) = \sum_{n=0}^{\infty} T_n x^n = x \exp\left(\sum_{r=1}^{\infty} \frac{1}{r} T(x^r)\right)$
   (b) $t(x) = \sum_{n=0}^{\infty} t_n x^n = T(x) - \frac{1}{2}\left[T^2(x) - T(x^2)\right]$.

7. The number of graphs, and other objects, of different orders.

Order	Graphs	Digraphs	Trees (t_n)	Rooted trees (T_n)
1	1	1	1	1
2	2	3	1	1
3	4	16	1	2
4	11	218	2	4
5	34	9608	3	9
6	156	1540944	6	20
7	1044	882033440	11	48
8	12346	1793359192848	23	115
9	274668		47	286
10	12005168		106	719

8. The number of isomorphism classes of digraphs with n vertices and m arcs.

m	$n = 1$	2	3	4	5
0	1	1	1	1	1
1		2	6	12	20
2		1	15	66	190
3			20	220	1,140
4			15	495	4,845
5			6	792	15,504
6			1	924	38,760
7				792	77,520

9. The number of isomorphism classes of graphs of order n having various properties:

n	1	2	3	4	5	6	7	8
All	1	2	4	11	34	156	1 044	12 346
Connected	1	1	2	6	21	112	853	11 117
Even	1	1	2	3	7	16	54	243
Eulerian	1	0	1	1	4	8	37	184
Blocks	0	1	1	3	10	56	468	7 123
Trees	1	1	1	2	3	6	11	23
Rooted trees	1	1	2	4	9	20	48	115

10. The number of labeled graphs of order n having various properties:

n	1	2	3	4	5	6	7	8
All	1	2	8	64	1 024	2^{15}	2^{21}	2^{28}
Connected	1	1	4	38	728	26 704	1 866 256	251 548 592
Even	1	1	2	8	64	1 024	2^{15}	2^{21}
Trees	1	1	3	16	125	1 296	16 807	262 144

11. The number of isomorphism classes of graphs of order n and size m.

m	$n=1$	2	3	4	5	6	7	8	9
0	1	1	1	1	1	1	1	1	1
1		1	1	1	1	1	1	1	1
2			1	2	2	2	2	2	1
3			1	3	4	5	5	5	5
4				2	6	9	10	11	11
5				1	6	15	21	24	25
6				1	6	21	41	56	63
7					4	24	65	115	148
8					2	24	97	221	345
9					1	21	131	402	771
10					1	15	148	663	1637

3.3.4.11 Descriptions of graphs with few vertices

The small graphs can be described in terms of the operations on page 159. Let $\mathcal{G}_{n,m}$ denote the family of isomorphism classes of graphs of order n and size m. Then

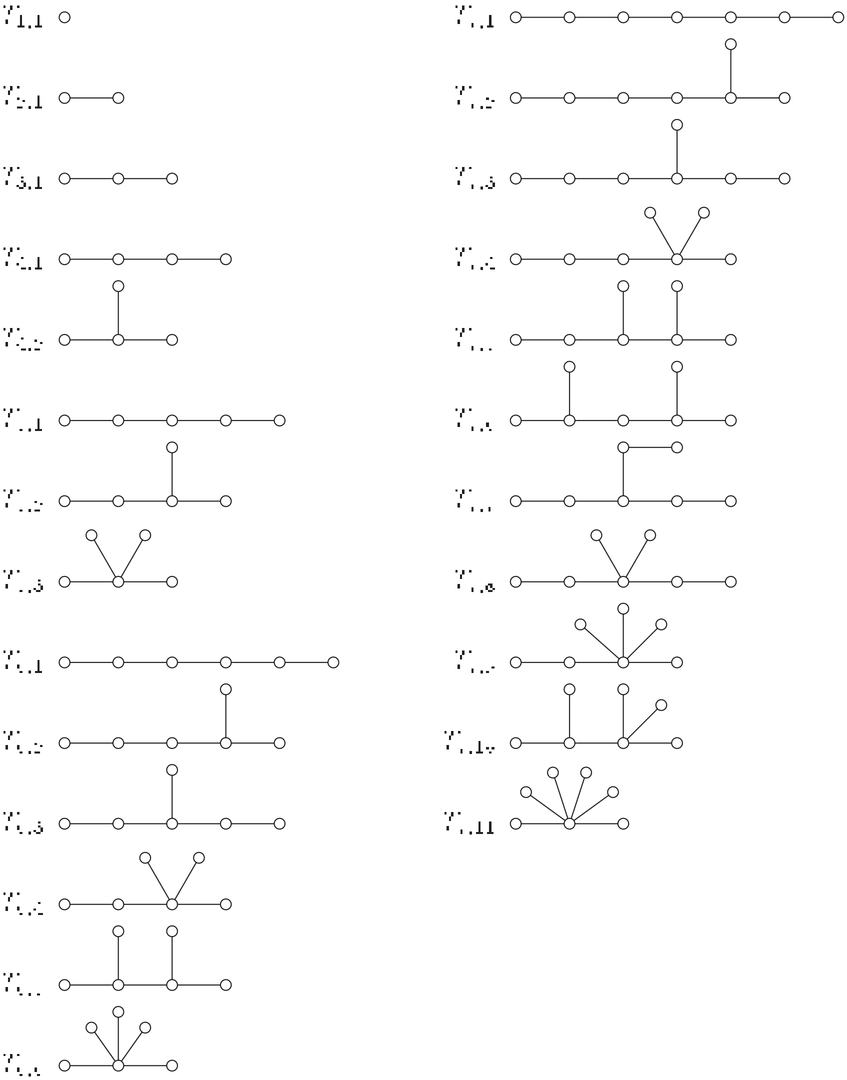
$$\mathcal{G}_{1,0} = \{K_1\} \qquad \mathcal{G}_{4,0} = \{\overline{K}_4\}$$
$$\mathcal{G}_{2,0} = \{\overline{K}_2\} \qquad \mathcal{G}_{4,1} = \{P_2 \cup \overline{K}_2\}$$
$$\mathcal{G}_{2,1} = \{K_2\} \qquad \mathcal{G}_{4,2} = \{P_3 \cup \overline{K}_1, P_2 \cup P_2\}$$
$$\mathcal{G}_{3,0} = \{\overline{K}_3\} \qquad \mathcal{G}_{4,3} = \{P_4, K_3 \cup \overline{K}_1, K_{1,3}\}$$
$$\mathcal{G}_{3,1} = \{K_2 \cup K_1\} \qquad \mathcal{G}_{4,4} = \{C_4, (K_2 \cup K_1) + K_1\}$$
$$\mathcal{G}_{3,2} = \{P_3\} \qquad \mathcal{G}_{4,5} = \{K_4 - e\}$$
$$\mathcal{G}_{3,3} = \{K_3\} \qquad \mathcal{G}_{4,6} = \{K_4\}$$

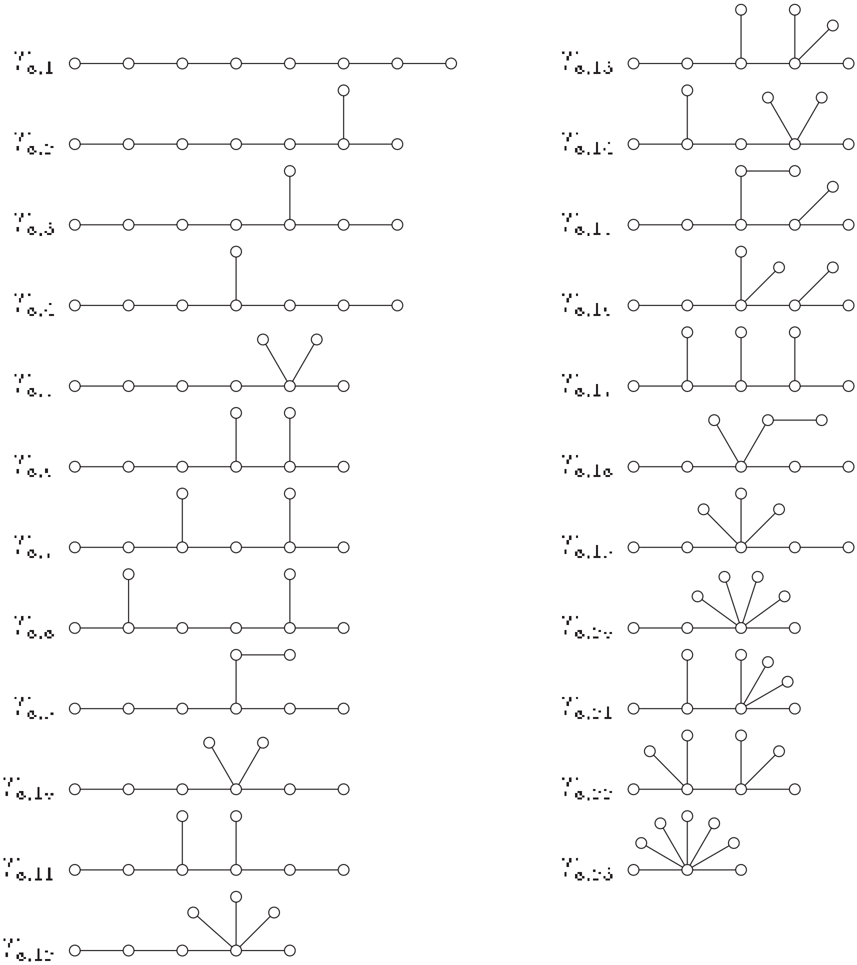
3.3.4.12 Matchings

1. A matching M is a maximum matching if and only if there is no M-augmenting path.
2. *Hall's theorem (Marriage theorem)*: For a set of vertices V in a graph G, let $N_G(V)$ be the set of all vertices adjacent to some element of X. A bipartite graph G with bipartition (S, T), with $|S| = |T|$, has a perfect matching if and only if $|X| \leq |N_G(X)|$ for every X in S.
3. *König's theorem*: In a bipartite graph, the cardinality of a maximum matching equals the cardinality of a minimum vertex cover.

3.3.5 TREE DIAGRAMS

Let $T_{n,m}$ denote the m^{th} isomorphism class of trees of order n.





3.4 COMBINATORIAL DESIGN THEORY

Combinatorial design theory is the study of families of subsets with various prescribed regularity properties. An *incidence structure* is an ordered pair $(X, \mathcal{B})$:

1. $X = \{x_1, \ldots, x_v\}$ is a set of *points*.
2. $\mathcal{B} = \{B_1, \ldots, B_b\}$ is a set of *blocks* or *lines*; each $B_j \subseteq X$.
3. The *replication number* r_i of x_i is the number of blocks that contain x_i.
4. The size of B_j is k_j.

Counting the number of pairs (x, B) with $x \in B$ yields $\sum_{i=1}^{v} r_i = \sum_{j=1}^{b} k_j$. The *incidence matrix* of an incidence structure is the $v \times b$ matrix $A = (a_{ij})$ with $a_{ij} = 1$ if $x_i \in B_j$ and 0 otherwise.

3.4.1 *t*-DESIGNS

The incidence structure $(X, \mathcal{B})$ is called a t-(v, k, λ) *design* if

1. For all j, $k_j = k$ and $1 < k < v$, and
2. Any subset of t points is contained in exactly λ blocks.

A 1-design is equivalent to a $v \times b$ 0-1 matrix with constant row and column sums. Every t-(v, k, λ) design is also a ℓ-(v, k, λ_ℓ) design ($1 \leq \ell \leq t$), where

$$\lambda_\ell = \lambda \binom{v-\ell}{t-\ell} \bigg/ \binom{k-\ell}{t-\ell}. \tag{3.4.1}$$

A necessary condition for the existence of a t-(v, k, λ) design is that λ_ℓ must be an integer for all ℓ, $1 \leq \ell \leq t$. Another necessary condition is the generalized Fisher's inequality: if $t = 2s$ then $b \geq \binom{v}{s}$.

3.4.1.1 Related designs

The existence of a t-(v, k, λ) design also implies the existence of the following designs:

Complementary design
: Let $\mathcal{B}_C = \{X \backslash B \mid B \in \mathcal{B}\}$. Then the incidence structure $(X, \mathcal{B}_C)$ is a t-$(v, v-k, \lambda\binom{v-t}{k} \big/ \binom{v-t}{k-t})$ design (provided $v \geq k + t$).

Derived design
: Fix $x \in X$ and let $\mathcal{B}_D = \{B \backslash \{x\} \mid B \in \mathcal{B}$ with $x \in B\}$. Then the incidence structure $(X \backslash \{x\}, \mathcal{B}_D)$ is a $(t-1)$-$(v-1, k-1, \lambda)$ design.

Residual design
: Fix $x \in X$and let $\mathcal{B}_R = \{B \mid B \in \mathcal{B}$ with $x \notin B\}$. Then the incidence structure $(X \backslash \{x\}, \mathcal{B}_R)$ is a $(t-1)$-$(v-1, k-1, \lambda\binom{v-t}{k-t+1} \big/ \binom{v-t}{k-t})$ design.

3.4.1.2 Examples

- 2-(4,3,1)-design: $X = \{1, 2, 3, 4\}$ and $\mathcal{B} = \{1, 2, 3\}$, $\{2, 3, 4\}$, $\{3, 4, 1\}$, $\{4, 1, 2\}$ $v = 4$ points, $b = 4$ blocks, block size $k = 3$, replication number $r = 3$, $\lambda = 1$ (each $t = 2$ points appear together in one block)
- 2-(7,3,1)-design: $X = \{1, 2, 3, 4, 5, 6, 7\}$ and $\mathcal{B} = \{1, 2, 4\}$, $\{2, 3, 5\}$, $\{3, 4, 6\}$, $\{4, 5, 7\}$, $\{5, 6, 1\}$, $\{6, 7, 2\}$, $\{7, 1, 3\}$ $v = 7$ points, $b = 7$ blocks, block size $k = 3$, replication number $r = 3$, $\lambda = 1$ (each $t = 2$ points appear together in one block)
- 3-(8,4,1)-design: $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $\mathcal{B} = \{1, 2, 5, 6\}$, $\{3, 4, 7, 8\}$, $\{1, 3, 5, 7\}$, $\{2, 4, 6, 8\}$, $\{1, 4, 5, 8\}$, $\{2, 3, 6, 7\}$, $\{1, 2, 3, 4\}$, $\{5, 6, 7, 8\}$, $\{1, 2, 7, 8\}$, $\{3, 4, 5, 6\}$, $\{1, 3, 6, 8\}$, $\{2, 4, 5, 7\}$, $\{1, 4, 6, 7\}$, $\{2, 3, 5, 8\}$
  $v = 8$ points, $b = 14$ blocks, block size $k = 4$, replication number $r = 7$, $\lambda = 1$ (each $t = 3$ points appear together in one block)
- 2-(8,4,3)-design: $X = \{1, 2, 3, 4, 5, 6, 7, 8\}$ and $\mathcal{B} = \{1, 3, 7, 8\}$, $\{1, 2, 4, 8\}$, $\{2, 3, 5, 8\}$, $\{3, 4, 6, 8\}$, $\{4, 5, 7, 8\}$, $\{1, 5, 6, 8\}$, $\{2, 6, 7, 8\}$, $\{1, 2, 3, 6\}$, $\{1, 2, 5, 7\}$, $\{1, 3, 4, 5\}$, $\{1, 4, 6, 7\}$, $\{2, 3, 4, 7\}$, $\{2, 4, 5, 6\}$, $\{3, 5, 6, 7\}$ $v = 8$ points, $b = 14$ blocks, block size $k = 4$, replication number $r = 7$, $\lambda = 3$ (each $t = 2$ points appear together in three blocks)

3.4.2 BALANCED INCOMPLETE BLOCK DESIGNS (BIBDS)

Balanced incomplete block designs (BIBDs) are t-designs with $t = 2$, so that every pair of points is on the same number of blocks. The relevant parameters are v, b, r, k, and λ with

$$vr = bk \qquad \text{and} \qquad v(v-1)\lambda = bk(k-1). \tag{3.4.2}$$

If A is the $v \times b$ incidence matrix, then $AA^{\mathrm{T}} = (r - \lambda)I_v + \lambda J_v$, where I_n is the $n \times n$ identity matrix and J_n is the $n \times n$ matrix of all ones.

3.4.2.1 Symmetric designs

Fisher's inequality states that $b \geq v$. If $b = v$ (equivalently, $r = k$), then the BIBD is called a *symmetric design*, denoted as a (v, k, λ)-design. The incidence matrix for a symmetric design satisfies

$$J_v A = kJ_v = AJ_v \quad \text{and} \quad A^{\mathrm{T}} A = (k - \lambda)I_v + \lambda J_v, \tag{3.4.3}$$

that is, any two blocks intersect in λ points. The dualness of symmetric designs can be summarized by the following:

v points	$\longleftrightarrow$	v blocks,
k blocks on a point	$\longleftrightarrow$	k points in a block, and
Any two points on λ blocks	$\longleftrightarrow$	Any two blocks share λ points.

Some necessary conditions for symmetric designs are

1. If v is even, then $k - \lambda$ is a square integer.
2. *Bruck–Ryser–Chowla theorem*: If v is odd, then the following equation has integer solutions (not all zero):

$$x^2 = (k - \lambda)y^2 + (-1)^{(v-1)/2}\lambda z^2.$$

3.4.2.2 Existence table for BIBDs

Some of the most fruitful construction methods for BIBD are dealt with in separate sections, difference sets (page 174), finite geometry (page 176), Steiner triple systems (page 180), and Hadamard matrices (page 87). The table below gives all parameters for which BIBDs exist with $k \leq v/2$ and $b \leq 30$.

v	b	r	k	λ
6	10	5	3	2
6	20	10	3	4
6	30	15	3	6
7	7	3	3	1
7	14	6	3	2
7	21	9	3	3
7	28	12	3	4
8	14	7	4	3
8	28	14	4	6
9	12	4	3	1
9	18	8	4	3
9	24	8	3	2
10	15	6	4	2

v	b	r	k	λ
10	18	9	5	4
10	30	9	3	2
10	30	12	4	4
11	11	5	5	2
11	22	10	5	4
12	22	11	6	5
13	13	4	4	1
13	26	6	3	1
13	26	8	4	2
13	26	12	6	5
14	26	13	7	6
15	15	7	7	3

v	b	r	k	λ
15	30	14	7	6
16	16	6	6	2
16	20	5	4	1
16	24	9	6	3
16	30	15	8	7
19	19	9	9	4
21	21	5	5	1
21	30	10	7	3
23	23	11	11	5
25	25	9	9	3
25	30	6	5	1
27	27	13	13	6

3.4.3 DIFFERENCE SETS

Let G be a finite group of order v (see page 106). A subset D of size k is a (v, k, λ)-*difference set* in G if every non-identity element of G can be written λ times as a "difference" $d_1 d_2^{-1}$ with d_1 and d_2 in D. If G is the cyclic group $\mathbb{Z}_v$, then the difference set is a *cyclic difference set*. The *order* of a difference set is $n = k - \lambda$. For example, $\{1, 2, 4\}$ is a $(7, 3, 1)$ cyclic difference set of order 2.

The existence of a (v, k, λ)-difference set implies the existence of a (v, k, λ)-design. The points are the elements of G and the blocks are the translates of D: all sets $Dg = \{dg : d \in D\}$ for $g \in G$. Note that each translate Dg is itself a difference set.

EXAMPLES

1. Here are the 7 blocks for a $(7, 3, 1)$-design based on $D = \{1, 2, 4\}$:
   1 2 4   2 3 5   3 4 6   4 5 0   5 6 1   6 0 2   0 1 3
2. A $(16, 6, 2)$-difference set in $G = \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2 \oplus \mathbb{Z}_2$ is

   0000   0001   0010   0100   1000   1111

3. A $(21, 5, 1)$-difference set in $G = \langle a, b : a^3 = b^7 = 1, a^{-1}ba = a^4 \rangle$ is $\{a, a^2, b, b^2, b^4\}$.

3.4.3.1 Some families of cyclic difference sets

Paley: Let v be a prime congruent to 3 modulo 4. Then the non-zero squares in $\mathbb{Z}_v$ form a $(v, (v-1)/2, (v-3)/4)$-difference set. Example: $(v, k, \lambda) = (11, 5, 2)$.

Stanton–Sprott: Let $v = p(p+2)$, where p and $p+2$ are both primes. Then there is a $(v, (v-1)/2, (v-3)/4)$-difference set. Example: $(v, k, \lambda) = (35, 17, 8)$.

Biquadratic residues (I): If $v = 4a^2 + 1$ is a prime with a odd, then the non-zero fourth powers modulo v form a $(v, (v-1)/4, (v-5)/16)$-difference set. Example: $(v, k, \lambda) = (37, 9, 2)$.

Biquadratic residues (II): If $v = 4a^2 + 9$ is a prime with a odd, then zero and the fourth powers modulo v form a $(v, (v+3)/4, (v+3)/16)$-difference set. Example: $(v, k, \lambda) = (13, 4, 1)$.

Singer: If q is a prime power, then there exists a $\left(\frac{q^m-1}{q-1}, \frac{q^{m-1}-1}{q-1}, \frac{q^{m-2}-1}{q-1}\right)$-difference set for all $m \geq 3$.

3.4.3.2 Existence table of cyclic difference sets

This table gives all cyclic difference sets for $k \leq v/2$ and $v \leq 50$ up to equivalence by translation and multiplication by a number relatively prime to v.

v	k	λ	n	Difference set
7	3	1	2	1 2 4
11	5	2	3	1 3 4 5 9
13	4	1	3	0 1 3 9
15	7	3	4	0 1 2 4 5 8 10
19	9	4	5	1 4 5 6 7 9 11 16 17
21	5	1	4	3 6 7 12 14
23	11	5	6	1 2 3 4 6 8 9 12 13 16 18
31	6	1	5	1 5 11 24 25 27
31	15	7	8	1 2 3 4 6 8 12 15 16 17 23 24 27 29 30
				1 2 4 5 7 8 9 10 14 16 18 19 20 25 28
35	17	8	9	0 1 3 4 7 9 11 12 13 14 16 17 21 27 28 29 33
37	9	2	7	1 7 9 10 12 16 26 33 34
40	13	4	9	1 2 3 5 6 9 14 15 18 20 25 27 35
43	21	10	11	1 2 3 4 5 8 11 12 16 19 20 21 22 27 32 33 35 37 39 41 42
				1 4 6 9 10 11 13 14 15 16 17 21 23 24 25 31 35 36 38 40 41
47	23	11	12	1 2 3 4 6 7 8 9 12 14 16 17 18 21 24 25 27 28 32 34 36 37 42

3.4.4 FINITE GEOMETRY

3.4.4.1 Affine planes

A finite *affine plane* is a finite set of *points* together with subsets of points called *lines* that satisfy the axioms:

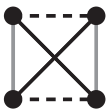
1. Any two points are on exactly one line.
2. *(Parallel postulate)* Given a point P and a line L not containing P, there is exactly one line through P that does not intersect L.
3. There are four points, no three of which are collinear.

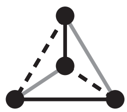
These axioms are sufficient to show that a finite affine plane is a BIBD (see page 173) with

$$v = n^2 \qquad b = n^2 + n \qquad r = n + 1 \qquad k = n \qquad \lambda = 1$$

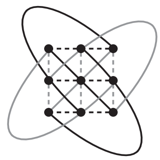
(n is called the *order* of the plane). The lines of an affine plane can be divided into $n+1$ parallel classes each containing n lines. A sufficient condition for affine planes to exist is for n to be a prime power.

Below are two views of the affine plane of order 2 showing the parallel classes.





Here is the affine plane of order 3 showing the parallel classes.



3.4.4.2 Projective planes

A finite *projective plane* is a finite set of points together with subsets of points called *lines* that satisfy the axioms:

1. Any two points are on exactly one line.
2. Any two lines intersect in exactly one point.
3. There are four points, no three of which are collinear.

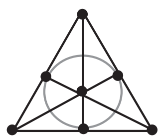
These axioms are sufficient to show that a finite affine plane is a symmetric design (see page 173) with

$$v = n^2 + n + 1 \qquad k = n + 1 \qquad \lambda = 1 \tag{3.4.4}$$

(n is the *order* of the plane). A sufficient condition for affine planes to exist is for n to be a prime power.

A projective plane of order n can be constructed from an affine plane of order n by adding a *line at infinity*. A line of $n + 1$ new points is added to the affine plane. For each parallel class, one distinct new point is added to each line. The construction works in reverse: removing any one line from a projective plane of order n and its points leaves an affine plane of order n.

To the right is the projective plane of order 2. The center circle functions as a line at infinity; removing it produces the affine plane of order 2.



3.4.5 FINITE FIELDS

Pertinent definitions for finite fields may be found in Section 2.5.8 on page 117.

3.4.5.1 Irreducible polynomials

Let $N_q(n)$ be the number of monic irreducible polynomials of degree n over GF(q). Then

$$q^n = \sum_{d|n} dN_q(d) \qquad \text{and} \qquad N_q(n) = \frac{1}{n}\sum_{d|n} \mu\left(\frac{n}{d}\right) q^d, \tag{3.4.5}$$

where $\mu(\cdot)$ is the number theoretic Möbius function (see page 36).

3.4.5.2 Table of binary irreducible polynomials

The table lists the non-zero coefficients of binary irreducible polynomials, e.g., 2 1 0 corresponds to $x^2+x^1+x^0 = x^2+x+1$. The *exponent* of an irreducible polynomial is the smallest L such that $f(x)$ divides $x^L - 1$. A "P" after the exponent indicates that the polynomial is primitive.

$f(x)$	Exponent
2 1 0	3 P
3 1 0	7 P
3 2 0	7 P
4 1 0	15 P
4 2 0	15 P
4 3 2 1 0	5
5 2 0	31 P
5 3 0	31 P
5 3 2 1 0	31 P
5 4 2 1 0	31 P
5 4 3 1 0	31 P
5 4 3 2 0	31 P
6 1 0	63 P
6 3 0	9
6 4 2 1 0	21
6 4 3 1 0	63 P
6 5 0	63 P
6 5 2 1 0	63 P
6 5 3 2 0	63 P
6 5 4 1 0	63
6 5 4 2 0	21 P
7 1 0	127 P
7 3 0	127 P

$f(x)$	Exponent
7 3 2 1 0	127 P
7 4 0	127 P
7 4 3 2 0	127 P
7 5 2 1 0	127 P
7 5 3 1 0	127 P
7 5 4 3 0	127 P
7 5 4 3 2 1 0	127 P
7 6 0	127 P
7 6 3 1 0	127 P
7 6 4 1 0	127 P
7 6 4 2 0	127 P
7 6 5 2 0	127 P
7 6 5 3 2 1 0	127 P
7 6 5 4 0	127 P
7 6 5 4 2 1 0	127 P
7 6 5 4 3 2 0	127 P
8 4 3 1 0	51
8 4 3 2 0	255 P
8 5 3 1 0	255 P
8 5 3 2 0	255 P
8 5 4 3 0	17
8 5 4 3 2 1 0	85
8 6 3 2 0	255 P

$f(x)$	Exponent
8 6 4 3 2 1 0	255 P
8 6 5 1 0	255 P
8 6 5 2 0	255 P
8 6 5 3 0	255 P
8 6 5 4 0	255 P
8 6 5 4 2 1 0	85
8 6 5 4 3 1 0	85
8 7 2 1 0	255 P
8 7 3 1 0	85
8 7 3 2 0	255 P
8 7 4 3 2 1 0	51
8 7 5 1 0	85
8 7 5 3 0	255 P
8 7 5 4 0	51
8 7 5 4 3 2 0	85
8 7 6 1 0	255 P
8 7 6 3 2 1 0	255 P
8 7 6 4 2 1 0	17
8 7 6 4 3 2 0	85
8 7 6 5 2 1 0	255 P
8 7 6 5 4 1 0	51
8 7 6 5 4 2 0	255 P
8 7 6 5 4 3 0	85

3.4.5.3 Table of binary primitive polynomials

Listed below[1] are primitive polynomials, with the least number of non-zero terms, of degree from 1 to 64. Only the exponents of the non-zero terms are listed. For example the entry "2 1 0" corresponds to $x^2 + x + 1$.

$f(x)$	$f(x)$	$f(x)$	$f(x)$	$f(x)$
1 0	14 5 3 1 0	27 5 2 1 0	40 5 4 3 0	53 6 2 1 0
2 1 0	15 1 0	28 3 0	41 3 0	54 6 5 4 3 2 0
3 1 0	16 5 3 2 0	29 2 0	42 5 4 3 2 1 0	55 6 2 1 0
4 1 0	17 3 0	30 6 4 1 0	43 6 4 3 0	56 7 4 2 0
5 2 0	18 5 2 1 0	31 3 0	44 6 5 2 0	57 5 3 2 0
6 1 0	19 5 2 1 0	32 7 5 3 2 1 0	45 4 3 1 0	58 6 5 1 0
7 1 0	20 3 0	33 6 4 1 0	46 8 5 3 2 1 0	59 6 5 4 3 1 0
8 4 3 2 0	21 2 0	34 7 6 5 2 1 0	47 5 0	60 1 0
9 4 0	22 1 0	35 2 0	48 7 5 4 2 1 0	61 5 2 1 0
10 3 0	23 5 0	36 6 5 4 2 1 0	49 6 5 4 0	62 6 5 3 0
11 2 0	24 4 3 1 0	37 5 4 3 2 1 0	50 4 3 2 0	63 1 0
12 6 4 1 0	25 3 0	38 6 5 1 0	51 6 3 1 0	64 4 3 1 0
13 4 3 1 0	26 6 2 1 0	39 4 0	52 3 0	

3.4.6 GRAY CODE

A *Gray code* is a sequence ordering such that a small change in the sequence number results in a small change in the sequence. The standard recursive construction for Gray codes is $G_n = 0G_{n-1}, 1G_{n-1}^{\text{reversed}}$.

EXAMPLES

1. The four 2-bit strings can be ordered so that adjacent bit strings differ in only 1 bit: $\{00, 01, 11, 10\}$.
2. The eight 3-bit strings can be ordered so that adjacent bit strings differ in only 1 bit: $\{000, 001, 011, 010, 110, 111, 101, 100\}$.
3. The sixteen 4-bit strings can be ordered so that adjacent bit strings differ in only 1 bit: $\{0000, 0001, 0011, 0010, 0110, 0111, 0101, 0100, 1100, 1101, 1111, 1110, 1010, 1011, 1001, 1000\}$.
4. The subsets of $\{a, b, c\}$ can be ordered so that adjacent subsets differ by only the insertion or deletion of a single element:

$$\phi, \quad \{a\}, \quad \{a,b\}, \quad \{b\}, \quad \{b,c\}, \quad \{a,b,c\}, \quad \{a,c\}, \quad \{c\}.$$

[1]Taken in part from "Primitive Polynomials (Mod 2)," E. J. Watson, *Math. Comp.*, **16**, 368–369, 1962.

3.4.7 LATIN SQUARES

A *Latin square* of size n is an $n \times n$ array $S = [s_{ij}]$ of n symbols such that every symbol appears exactly once in each row and column. Two Latin squares S and T are *orthogonal* if every pair of symbols occurs exactly once as a pair (s_{ij}, t_{ij}). Let $M(n)$ be the maximum size of a set of mutually orthogonal Latin squares (MOLS).

1. $M(n) \leq n - 1$.
2. $M(n) = n - 1$ if n is a prime power.
3. $M(n_1 n_2) \geq \min(M(n_1), M(n_2))$.
4. $M(6) = 1$ (i.e., there are no two MOLS of size 6).
5. $M(n) \geq 2$ for all $n \geq 3$ except $n = 6$. (Latin squares of all sizes exist.)

The existence of $n - 1$ MOLS of size n is equivalent to the existence of an affine plane of order n (see page 176).

3.4.7.1 Examples of mutually orthogonal Latin squares

These are complete sets of MOLS for $n = 3$, 4, and 5.

$n = 3$

0	1	2
1	2	0
2	0	1

0	1	2
2	0	1
1	2	0

$n = 4$

0	1	2	3
1	0	3	2
2	3	0	1
3	2	1	0

0	1	2	3
2	3	0	1
3	2	1	0
1	0	3	2

0	1	2	3
3	2	1	0
1	0	3	2
2	3	0	1

$n = 5$

0	1	2	3	4
1	2	3	4	0
2	3	4	0	1
3	4	0	1	2
4	0	1	2	3

0	1	2	3	4
2	3	4	0	1
4	0	1	2	3
1	2	3	4	0
3	4	0	1	2

0	1	2	3	4
3	4	0	1	2
1	2	3	4	0
4	0	1	2	3
2	3	4	0	1

0	1	2	3	4
4	0	1	2	3
3	4	0	1	2
2	3	4	0	1
1	2	3	4	0

These are two superimposed MOLS for $n = 7$ and $n = 8$.

$n = 7$

00	11	22	33	44	55	66
16	20	31	42	53	64	05
25	36	40	51	62	03	14
34	45	56	60	01	12	23
43	54	65	06	10	21	32
52	63	04	15	26	30	41
61	02	13	24	35	46	50

$n = 8$

00	11	22	33	44	55	66	77
12	03	30	21	56	47	74	65
24	35	06	17	60	71	42	53
33	22	11	00	77	66	55	44
46	57	64	75	02	13	20	31
57	46	75	64	13	02	31	20
65	74	47	56	21	30	03	12
71	60	53	42	35	24	17	06

3.4.8 STEINER TRIPLE SYSTEMS

A *Steiner triple system* (STS) is a 2-(v,3,1) design. In particular, STSs are BIBDs (see page 173). STSs exist if and only if $v \equiv 1$ or $3 \pmod 6$. The number of blocks in an STS is $b = v(v-1)/6$.

3.4.8.1 Some families of Steiner triple systems

$v = 2^m - 1$**:** Take as points all non-zero vectors over $\mathbb{Z}_2$ of length m. A block consists of any set of three distinct vectors $\{x, y, z\}$ such that $x + y + z = 0$.

$v = 3^m$**:** Take as points all vectors over $\mathbb{Z}_3$ of length m. A block consists of any set of three distinct vectors $\{x, y, z\}$ such that $x + y + z = 0$.

3.4.8.2 Resolvable Steiner triple systems

An STS is *resolvable* if the blocks can be divided into parallel classes such that each point occurs in exactly one block per class. A resolvable STS exists if and only if $v \equiv 3 \pmod 6$. For example, the affine plane of order 3 is a resolvable STS with $v = 9$ (see page 176).

A resolvable STS with $v = 15$ ($b = 35$) is known as the *Kirkman schoolgirl problem* and dates from 1850. Each column of 5 triples is a parallel class:

a b i	a c j	a d k	a e l	a f m	a g n	a h o
c d f	d e g	e f h	f g b	g h c	h b d	b c e
g j o	h k i	b l j	c m k	d n l	e o m	f i n
e k n	f l o	g m i	h n j	b o k	c i l	d j m
h l m	b m n	c n o	d o i	e i j	f j k	g k l

3.4.9 DESIGNS AND HADAMARD MATRICES

See page 87 for Hadamard matrices.

BIBDs: Let H be a Hadamard matrix of order $4t$, normalized so the first row and column are 1's. Remove the first row and column. Let A be the remaining $(4t-1) \times (4t-1)$ matrix. If J is the $(4t-1) \times (4t-1)$ matrix of all ones, then the incidence matrix

- for a 2-$(4t-1, 2t-1, t-1)$ design is given by $\frac{1}{2}(J + A)$.
- for a 2-$(4t-1, 2t, t)$ design is given by $\frac{1}{2}(J - A)$.

3-Designs: Let H be a Hadamard matrix of order $4t$. Choose any row and normalize it (scale the columns as needed) so that every entry is $+1$. Let the columns represent points. Let the sets of columns carrying $+1$s and -1s in all but the chosen row be blocks. This results in a 3-$(4t, 2t, t-1)$ design; it is an affine design. Different choices of row may, or may not, give isomorphic designs.

3.4.10 ROOK POLYNOMIALS

Consider an $n \times m$ board in which some certain squares are forbidden (shown white) and others are acceptable (shown dark). Let $r_k(B)$ be the number of ways to choose k acceptable (darkened) squares, no two of which lie in the same row and no two of which lie in the same column. (This scenario might arise from a job assignment.) Equivalently, in how many ways can k rooks be placed on a board's dark squares in such a way that no two can take each other? The expression

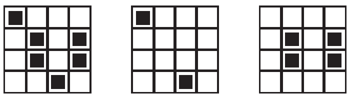
$$R(x, B) = r_0(B) + r_1(B)x + r_2(B)x^2 + \dots \tag{3.4.6}$$

is the board's *rook polynomial*.

Consider the 2×2 board $B_1 =$ . For this board there is 1 way to place no rooks, 2 ways to place one rook (use either darkened square), 1 way to place two rooks (use both darkened squares), and no way to place more than two rooks. Hence, $R(x, B_1) = 1 + 2x + x^2$.

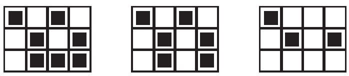
Consider the 2×2 board $B_2 =$ . For this board there is 1 way to place no rooks, 4 ways to place one rook (use any square), 2 ways to place two rooks (use diagonal squares), and no way to place more than two rooks. Hence, $R(x, B_2) = 1 + 4x + 2x^2$.

Suppose I is a set of darkened squares in a board B and B_I is the board obtained from B by lightening the darkened squares in B that are not in I. If the darkened squares in B are partitioned into two sets, I and J, such that no square in I lies in the same row or column as any square of J, then B_I and B_J *decompose* the board B. In this case: $R(x, B) = R(x, B_1)\ R(x, B_2)$. See the following example.



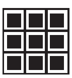
B $\qquad$ B_I $\qquad$ B_J

Suppose that s is any darkened square in a board B. Let B_s be obtained from B by lightening s. Let B'_s be obtained from B by lightening all squares in the same row and column as s. Then $R(x, B) = R(x, B_s) + x\, R(x, B'_s)$. (See the following example, where s is the darkened square in row 3, column 3.)

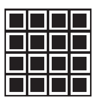


B $\qquad$ B_s $\qquad$ B'_s

EXAMPLES For a square checkerboard in which every square is darkened (B_n), the rook polynomial is $R(x, B_n) = n!\, x^n\, L_n\left(-\frac{1}{x}\right)$, where L_n is the n^{th} Laguerre polynomial.



$R(x, B_3) = 1 + 9x + 18x^2 + 6x^3$



$R(x, B_4) = 1 + 16x + 72x^2 + 96x^3 + 24x^4$

We find $R(x, B_5) = 1 + 25x + 200x^2 + 600x^3 + 600x^4 + 120x^5$.

3.4.11 BINARY SEQUENCES

3.4.11.1 Barker sequences

A *Barker sequence* $(s_1, \ldots, s_N)$ with $s_j = \pm 1$ has $\sum_{j=1}^{N-i} s_j s_{j+i} = \pm 1$ or 0, for $i = 1, \ldots, N-1$. The following table lists all known Barker sequences (up to reversal, multiplication by -1, and multiplying alternate values by -1).

Length	Barker sequence
2	$+1$ $+1$
3	$+1$ $+1$ -1
4	$+1$ $+1$ $+1$ -1
4	$+1$ $+1$ -1 $+1$
5	$+1$ $+1$ $+1$ -1 $+1$
7	$+1$ $+1$ $+1$ -1 -1 $+1$ -1
11	$+1$ $+1$ $+1$ -1 -1 -1 $+1$ -1 -1 $+1$ -1
13	$+1$ $+1$ $+1$ $+1$ $+1$ -1 -1 $+1$ $+1$ -1 $+1$ -1 $+1$

3.4.11.2 Periodic sequences

Let $\mathbf{s} = (s_0, s_1, \ldots, s_{N-1})$ be a periodic sequence with period N. A (left) *shift* of $\mathbf{s}$ is the sequence $(s_1, \ldots, s_{N-1}, s_0)$. For τ relatively prime to N, the *decimation* of $\mathbf{s}$ is the sequence $(s_0, s_\tau, s_{2\tau}, \ldots)$, which also has period N. The *periodic autocorrelation* is defined as the vector $(a_0, \ldots, a_{N-1})$, with

$$a_i = \sum_{j=0}^{N-1} s_j s_{j+i}, \qquad \text{(subscripts taken modulo } N). \tag{3.4.7}$$

An autocorrelation is *two-valued* if all values are equal except possibly for the 0^{th} term.

3.4.11.3 The m-sequences

A binary *m-sequence* of length $N = 2^r - 1$ is the sequence of period N defined by

$$(s_0, s_1, \ldots, s_{N-1}), \qquad s_i = \text{Tr}(\alpha^i), \tag{3.4.8}$$

where α is a primitive element of GF(2^r) and Tr is the trace function from GF(2^r) to GF(2).[2]

1. All m-sequences of a given length are equivalent under decimation.
2. Binary m-sequences have a two-valued autocorrelation (with the identification that $0 \leftrightarrow +1$ and $1 \leftrightarrow -1$).
3. All m-sequences possess the *span property*: all binary r-tuples occur in the sequence except the all-zeros r-tuple.

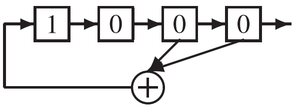
The existence of a binary sequence of length $2^n - 1$ with a two-valued autocorrelation is equivalent to the existence of a cyclic difference set with parameters $(2^n - 1, 2^{n-1} - 1, 2^{n-2} - 1)$.

[2] If $x \in F = GF(q^m)$ and $K = GF(q)$, the *trace function* $\text{Tr}_{F/K}(x)$ from F onto K is defined by $\text{Tr}_{F/K}(x) = x + x^q + \cdots + x^{q^{m-1}}$.

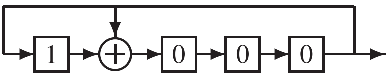
3.4.11.4 Shift registers

Below are examples of the two types of shift registers used to generate binary m-sequences. The generating polynomial in each case is $x^4 + x + 1$, the initial register loading is 1 0 0 0, and the generated sequence is $\{0, 0, 0, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, \ldots\}$.

Additive shift register



Multiplicative shift register



3.4.11.5 Binary sequences with two-valued autocorrelation

The following table lists all binary sequences with two-valued periodic autocorrelation of length $2^n - 1$ for $n = 3$ to 8 (up to shifts, decimations, and complementation). The table indicates the positions of 0; the remaining values are 1. An S indicates that the sequence has the span property.

n	Positions of 0
3 S	1 2 4
4 S	0 1 2 4 5 8 10
5 S	1 2 3 4 6 8 12 15 16 17 23 24 27 29 30
5	1 2 4 5 7 8 9 10 14 16 18 19 20 25 28
6 S	0 1 2 3 4 6 7 8 9 12 13 14 16 18 19 24 26 27 28 32 33 35 36 38 41 45 48 49 52 54 56
6	0 1 2 3 4 5 6 8 9 10 12 16 17 18 20 23 24 27 29 32 33 34 36 40 43 45 46 48 53 54 58
7	1 2 4 8 9 11 13 15 16 17 18 19 21 22 25 26 30 31 32 34 35 36 37 38 41 42 44 47 49 50 52 60 61 62 64 68 69 70 71 72 73 74 76 79 81 82 84 87 88 94 98 99 100 103 104 107 113 115 117 120 121 122 124
7	1 2 3 4 5 6 7 8 10 12 14 16 19 20 23 24 25 27 28 32 33 38 40 46 47 48 50 51 54 56 57 61 63 64 65 66 67 73 75 76 77 80 87 89 92 94 95 96 97 100 101 102 107 108 111 112 114 117 119 122 123 125 126
7 S	1 2 3 4 6 7 8 9 12 14 15 16 17 18 24 27 28 29 30 31 32 34 36 39 47 48 51 54 56 58 60 61 62 64 65 67 68 71 72 77 78 79 83 87 89 94 96 97 99 102 103 105 107 108 112 113 115 116 117 120 121 122 124
7	1 2 3 4 6 7 8 9 12 13 14 16 17 18 19 24 25 26 27 28 31 32 34 35 36 38 47 48 50 51 52 54 56 61 62 64 65 67 68 70 72 73 76 77 79 81 87 89 94 96 97 100 102 103 104 107 108 112 115 117 121 122 124
7	1 2 3 4 5 6 8 9 10 12 15 16 17 18 19 20 24 25 27 29 30 32 33 34 36 38 39 40 48 50 51 54 55 58 59 60 64 65 66 68 71 72 73 76 77 78 80 83 89 91 93 96 99 100 102 105 108 109 110 113 116 118 120
7	1 2 3 4 5 6 8 10 11 12 16 19 20 21 22 24 25 27 29 32 33 37 38 39 40 41 42 44 48 49 50 51 54 58 63 64 65 66 69 73 74 76 77 78 80 82 83 84 88 89 95 96 98 100 102 105 108 111 116 119 123 125 126
8 S	0 1 2 3 4 6 7 8 12 13 14 16 17 19 23 24 25 26 27 28 31 32 34 35 37 38 41 45 46 48 49 50 51 52 54 56 59 62 64 67 68 70 73 74 75 76 82 85 90 92 96 98 99 100 102 103 104 105 108 111 112 113 118 119 123 124 127 128 129 131 134 136 137 139 140 141 143 145 146 148 150 152 153 157 161 164 165 170 177 179 180 183 184 187 189 191 192 193 196 197 198 199 200 204 206 208 210 216 217 219 221 222 223 224 226 227 236 237 238 239 241 246 247 248 251 253 254
8	0 1 2 4 7 8 9 11 14 16 17 18 19 21 22 23 25 27 28 29 32 33 34 35 36 38 42 43 44 46 49 50 51 54 56 58 61 64 66 68 69 70 71 72 76 79 81 84 85 86 87 88 89 92 93 95 97 98 99 100 101 102 108 112 113 116 117 119 122 125 128 131 132 133 136 137 138 139 140 141 142 144 145 149 152 153 158 162 163 167 168 170 171 172 174 175 176 177 178 184 186 187 190 193 194 196 197 198 200 202 204 209 211 213 215 216 221 224 226 232 233 234 235 238 244 245 250
8	0 1 2 3 4 6 8 12 13 15 16 17 24 25 26 27 29 30 31 32 34 35 39 47 48 50 51 52 54 57 58 59 60 61 62 64 67 68 70 71 78 79 85 91 94 96 99 100 102 103 104 107 108 109 114 116 118 119 120 121 122 124 127 128 129 134 135 136 140 141 142 143 145 147 151 153 156 157 158 161 163 167 170 173 177 179 181 182 187 188 191 192 195 198 199 200 201 203 204 206 208 209 211 214 216 217 218 221 223 225 227 228 229 232 233 236 238 239 240 241 242 244 247 248 251 253 254
8	0 1 2 3 4 6 7 8 11 12 14 15 16 17 21 22 23 24 25 28 29 30 32 34 35 37 41 42 44 46 47 48 50 51 56 58 60 64 68 69 70 71 73 74 81 82 84 85 88 91 92 94 96 97 100 102 107 109 111 112 113 116 119 120 121 123 127 128 129 131 133 135 136 138 139 140 142 145 146 148 151 153 162 163 164 168 170 173 176 181 182 183 184 187 188 189 191 192 193 194 195 197 200 203 204 209 214 218 219 221 222 223 224 225 226 229 232 237 238 239 240 242 246 247 251 253 254

3.5 DIFFERENCE EQUATIONS

In many ways, difference equations are analogous to differential equations.

3.5.1 THE CALCULUS OF FINITE DIFFERENCES

1. $\Delta(f(x)) = f(x+h) - f(x)$ (forward difference).
2. $\Delta^2(f(x)) = f(x+2h) - 2f(x+h) + f(x)$.
3. $\Delta^n(f(x)) = \Delta\left(\Delta^{n-1}(f(x))\right) = \sum_{k=0}^{n}(-1)^k \binom{n}{k} f(x+(n-k)h)$.
4. $\Delta(cf(x)) = c\Delta(f(x))$.
5. $\Delta(f(x)+g(x)) = \Delta(f(x)) + \Delta(g(x))$.
6. $\Delta(f(x)g(x)) = g(x)\Delta(f(x)) + f(x)\Delta(g(x))$.
7. $\Delta\left(\frac{f(x)}{g(x)}\right) = \dfrac{g(x)\Delta(f(x)) - f(x)\Delta(g(x))}{g(x)g(x+h)}$, provided that $g(x)g(x+h) \neq 0$.
8. $\Delta^n(x^n) = n!\, h^n, \quad n = 0, 1, \ldots$.

3.5.2 EXISTENCE AND UNIQUENESS

A *difference equation of order* k has the form

$$x_{n+k} = f(x_n, x_{n+1}, \ldots, x_{n+(k-1)}, n) \tag{3.5.1}$$

where f is a given function and k is a positive integer. A *solution* to Equation (3.5.1) is a sequence of numbers $\{x_n\}_{n=0}^{\infty}$ which satisfies the equation. Any constant solution of Equation (3.5.1) is called an *equilibrium solution.*

A *linear* difference equation of order k is one that can written in the form

$$a_n^{(k)} x_{n+k} + a_n^{(k-1)} x_{n+(k-1)} + \cdots + a_n^{(1)} x_{n-1} + a_n^{(0)} x_n = g_n, \tag{3.5.2}$$

where k is a positive integer and the coefficients $a_n^{(0)}, \ldots, a_n^{(k)}$ along with $\{g_n\}$ are known. If the sequence g_n is identically zero, then Equation (3.5.2) is called *homogeneous*; otherwise, it is called *non-homogeneous*. If the coefficients $a_n^{(0)}, \ldots, a_n^{(k)}$ are constants (i.e., do not depend on n), Equation (3.5.2) is a *difference equation with constant coefficients*; otherwise it is a *difference equation with variable coefficients*.

THEOREM 3.5.1 *(Existence and uniqueness)*
Consider the initial value problem (IVP)

$$\begin{aligned} &x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = f_n, \\ &x_i = \alpha_i, \qquad i = 0, 1, \ldots, k-1, \end{aligned} \tag{3.5.3}$$

for $n = 0, 1, \ldots$, where $b_n^{(i)}$ and f_n are given sequences with $b_n^{(0)} \neq 0$ for all n and the $\{\alpha_i\}$ are given initial conditions. Then the above equations have exactly one solution.

3.5.3 LINEAR INDEPENDENCE: GENERAL SOLUTION

The sequences $x^{(1)}, x^{(2)}, \dots, x^{(k)}$ (sequence $x^{(i)}$ has the terms $\{x_1^{(i)}, x_2^{(i)}, x_2^{(i)}, \dots\}$) are *linearly dependent* if constants $c_1, c_2, \dots, c_k$ (not all of them zero) exist such that

$$\sum_{i=1}^{k} c_i x_n^{(i)} = 0 \quad \text{for } n = 0, 1, \dots. \tag{3.5.4}$$

Otherwise the sequences $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ are *linearly independent*.

The *Casoratian* of the k sequences $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ is the $k \times k$ determinant

$$C\left(x_n^{(1)}, x_n^{(2)}, \dots, x_n^{(k)}\right) = \begin{vmatrix} x_n^{(1)} & x_n^{(2)} & \cdots & x_n^{(k)} \\ x_{n+1}^{(1)} & x_{n+1}^{(2)} & \cdots & x_{n+1}^{(k)} \\ \cdots & \cdots & \cdots & \cdots \\ x_{n+k-1}^{(1)} & x_{n+k-1}^{(2)} & \cdots & x_{n+k-1}^{(k)} \end{vmatrix}. \tag{3.5.5}$$

THEOREM 3.5.2

The solutions $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ of the linear homogeneous difference equation,

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = 0, \qquad n = 0, 1, \dots, \tag{3.5.6}$$

are linearly independent if and only if their Casoratian is different from zero for $n = 0$.

Note that the solutions to Equation (3.5.6) form a k-dimensional vector space. The set $\{x^{(1)}, x^{(2)}, \dots, x^{(k)}\}$ is a *fundamental system* of solutions for Equation (3.5.6) if and only if the sequences $x^{(1)}$, $x^{(2)}$, ..., $x^{(k)}$ are linearly independent solutions of the homogeneous difference Equation (3.5.6).

THEOREM 3.5.3

Consider the non-homogeneous linear difference equation

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = d_n, \qquad n = 0, 1, \dots \tag{3.5.7}$$

where $b_n^{(i)}$ and d_n are given sequences. Let $x_n^{(h)}$ be the general solution of the corresponding homogeneous equation

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = 0, \qquad n = 0, 1, \dots,$$

and let $x_n^{(p)}$ be a particular solution of Equation (3.5.7). Then $x_n^{(p)} + x_n^{(h)}$ is the general solution of Equation (3.5.7).

THEOREM 3.5.4 *(Superposition principle)*

Let $x^{(1)}$ and $x^{(2)}$ be solutions of the non-homogeneous linear difference equations

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = \alpha_n, \quad n = 0, 1, \dots,$$

and

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = \beta_n, \quad n = 0, 1, \ldots,$$

respectively, where $b^{(i)}$ *and* $\{\alpha_n\}$ *and* $\{\beta_n\}$ *are given sequences. Then* $x^{(1)} + x^{(2)}$ *is a solution of the equation*

$$x_{n+k} + b_n^{(k-1)} x_{n+(k-1)} + \cdots + b_n^{(1)} x_{n+1} + b_n^{(0)} x_n = \alpha_n + \beta_n, \quad n = 0, 1, \ldots.$$

3.5.4 HOMOGENEOUS EQUATIONS WITH CONSTANT COEFFICIENTS

The results given below for second-order linear difference equations extend naturally to higher order equations.

Consider the second-order linear homogeneous difference equation,

$$\alpha_2 x_{n+2} + \alpha_1 x_{n+1} + \alpha_0 x_n = 0, \tag{3.5.8}$$

where the $\{\alpha_i\}$ are real constant coefficients with $\alpha_2 \alpha_0 \neq 0$. The *characteristic equation* corresponding to Equation (3.5.8) is defined as the quadratic equation

$$\alpha_2 \lambda^2 + \alpha_1 \lambda + \alpha_0 = 0. \tag{3.5.9}$$

The solutions λ_1, λ_2 of the characteristic equation are the *eigenvalues* or the *characteristic roots* of Equation (3.5.8).

THEOREM 3.5.5

Let λ_1 *and* λ_2 *be the eigenvalues of Equation (3.5.8). Then the general solution of Equation (3.5.8) is given as described below with arbitrary constants* c_1 *and* c_2.

Case 1: $\lambda_1 \neq \lambda_2$ *with* $\lambda_1, \lambda_2 \in \mathbb{R}$ *(real and distinct roots).*
The general solution is given by $x_n = c_1 \lambda_1^n + c_2 \lambda_2^n$.

Case 2: $\lambda_1 = \lambda_2 \in \mathbb{R}$ *(real and equal roots)* .
The general solution is given by $x_n = c_1 \lambda_1^n + c_2 n \lambda_1^n$.

Case 3: $\lambda_1 = \overline{\lambda_2}$ *(complex conjugate roots).*
Suppose that $\lambda_1 = re^{i\phi}$. *The general solution is given by*

$$x_n = c_1 r^n \cos(n\phi) + c_2 r^n \sin(n\phi).$$

The constants $\{c_1, c_2\}$ are determined from the initial conditions.

EXAMPLE The unique solution of the initial value problem

$$\begin{aligned} F_{n+2} &= F_{n+1} + F_n, \qquad n = 0, 1, \ldots, \\ F_0 &= 0, \qquad F_1 = 1, \end{aligned} \tag{3.5.10}$$

is the Fibonacci sequence. The equation $\lambda^2 = \lambda + 1$ has the real and distinct roots $\lambda_{1,2} = \frac{1 \pm \sqrt{5}}{2}$. Using Theorem 3.5.5 the solution is

$$F_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right], \qquad n = 0, 1, \ldots. \tag{3.5.11}$$

3.5.5 NON-HOMOGENEOUS EQUATIONS

THEOREM 3.5.6 *(Variation of parameters)*
Consider the difference equation, $x_{n+2} + \alpha_n x_{n+1} + \beta_n x_n = \gamma_n$, *where* $\{\alpha_n\}$, $\{\beta_n\}$, *and* $\{\gamma_n\}$ *are given sequences with* $\beta_n \neq 0$. *Let* $x^{(1)}$ *and* $x^{(2)}$ *be two linearly independent solutions of the homogeneous equation corresponding to this equation. A particular solution* $x^{(p)}$ *has the component values* $x_n^{(p)} = x_n^{(1)} v_n^{(1)} + x_n^{(2)} v_n^{(2)}$ *where the sequences* $v^{(1)}$ *and* $v^{(2)}$ *satisfy the following system of equations:*

$$\begin{aligned} x_{n+1}^{(1)}\left(v_{n+1}^{(1)} - v_n^{(1)}\right) + x_{n+1}^{(2)}\left(v_{n+1}^{(2)} - v_n^{(2)}\right) &= 0, \quad \textit{and} \\ x_{n+2}^{(1)}\left(v_{n+1}^{(1)} - v_n^{(1)}\right) + x_{n+2}^{(2)}\left(v_{n+1}^{(2)} - v_n^{(2)}\right) &= \gamma_n. \end{aligned} \tag{3.5.12}$$

3.5.6 GENERATING FUNCTIONS AND Z TRANSFORMS

Generating functions can be used to solve initial value problems of difference equations in the same way that Laplace transforms are used to solve initial value problems of differential equations.

The *generating function* of the sequence $\{x_n\}$, denoted by $G[x_n]$, is defined by the infinite series

$$G[x_n] = \sum_{n=0}^{\infty} x_n s^n \tag{3.5.13}$$

provided that the series converges for $|s| < r$, for some positive number r. The following are useful properties of the generating function:

1. *Linearity*: $G[c_1 x_n + c_2 y_n] = c_1 G[x_n] + c_2 G[y_n]$.
2. *Translation invariance*: $G[x_{n+k}] = \dfrac{1}{s^k}\left(G[x_n] - \sum_{n=0}^{k-1} x_n s^n\right)$.
3. *Uniqueness*: $G[x_n] = G[y_n] \iff x_n = y_n \quad \text{for} \quad n = 0, 1, \ldots,$

The Z-*transform* of a sequence $\{x_n\}$ is denoted by $\mathcal{Z}[x_n]$ and is defined by the infinite series,

$$\mathcal{Z}[x_n] = \sum_{n=0}^{\infty} \frac{x_n}{z^n}, \tag{3.5.14}$$

provided that the series converges for $|z| > r$, for some positive number r. (The Z-transform is also in Section 6.40 on page 512.)

Comparing the definitions for the generating function and the Z-transform one can see that they are connected because Equation (3.5.14) can be obtained from Equation (3.5.13) by setting $s = z^{-1}$.

3.5.6.1 Generating functions for some common sequences

$\{x_n\}$	$G[x_n]$
1	$\dfrac{1}{1-s}$
a^n	$\dfrac{1}{1-as}$
na^n	$\dfrac{as}{(1-as)^2}$
$n^p a^n$	$\left(s\dfrac{d}{ds}\right)^p \dfrac{1}{1-as}$
n	$\dfrac{s}{(1-s)^2}$
$n+1$	$\dfrac{1}{(1-s)^2}$
n^p	$\left(s\dfrac{d}{ds}\right)^p \dfrac{1}{1-s}$

$\{x_n\}$	$G[x_n]$
$\sin(\beta n)$	$\dfrac{s\sin\beta}{1-2s\cos\beta+s^2}$
$\cos(\beta n)$	$\dfrac{1-s\cos\beta}{1-2s\cos\beta+s^2}$
$a^n\sin(\beta n)$	$\dfrac{as\sin\beta}{1-2as\cos\beta+a^2s^2}$
$a^n\cos(\beta n)$	$\dfrac{1-as\cos\beta}{1-2as\cos\beta+a^2s^2}$
x_{n+1}	$\dfrac{1}{s}\left(G[x_n]-x_0\right)$
x_{n+2}	$\dfrac{1}{s^2}\left(G[x_n]-x_0-sx_1\right)$
x_{n+k}	$\dfrac{1}{s^k}\left(G[x_n]-\sum_{n=0}^{k-1} x_n s^n\right)$

3.5.7 CLOSED-FORM SOLUTIONS FOR SPECIAL EQUATIONS

In general, it is difficult to find a closed-form solution for a difference equation which is not linear of order one or linear of any order with constant coefficients. A few special difference equations which possess closed-form solutions are presented below.

3.5.7.1 First-order equation

The general solution of the first-order linear difference equation with variable coefficients,

$$x_{n+1} - \alpha_n x_n = \beta_n, \qquad n = 0, 1, \ldots, \tag{3.5.15}$$

is given by

$$x_n = \left(\prod_{k=0}^{n-1} \alpha_k\right) x_0 + \sum_{m=0}^{n-2} \left(\prod_{k=m+1}^{n-1} \alpha_k\right) \beta_m + \beta_{n-1}, \qquad n = 0, 1, \ldots, \tag{3.5.16}$$

where x_0 is an arbitrary constant.

3.5.7.2 Riccati equation

Consider the non-linear first-order equation,

$$x_{n+1} = \frac{\alpha_n x_n + \beta_n}{\gamma_n x_n + \delta_n}, \qquad n = 0, 1, \ldots, \tag{3.5.17}$$

where α_n, β_n, γ_n, δ_n are given sequences of real numbers with

$$\gamma_n \neq 0 \quad \text{and} \quad \begin{vmatrix} \alpha_n & \beta_n \\ \gamma_n & \delta_n \end{vmatrix} \neq 0, \qquad n = 0, 1, \ldots. \tag{3.5.18}$$

The following statements are true:

1. The change of variables,

$$\frac{u_{n+1}}{u_n} = \gamma_n x_n + \delta_n, \qquad n = 0, 1, \ldots, \qquad u_0 = 1, \tag{3.5.19}$$

   reduces Equation (3.5.17) to the linear second-order equation,

$$\begin{aligned} u_{n+2} &= A_n u_{n+1} + B_n u_n, \qquad n = 0, 1, \ldots, \\ u_0 &= 1, \\ u_1 &= \gamma_0 x_0 + \delta_0, \end{aligned} \tag{3.5.20}$$

   where $A_n = \delta_{n+1} + \alpha_n \dfrac{\gamma_{n+1}}{\gamma_n}$, and $B_n = (\beta_n \gamma_n - \alpha_n \delta_n) \dfrac{\gamma_{n+1}}{\gamma_n}$.

2. Let $x^{(p)}$ be a particular solution of Equation (3.5.17). The change of variables,

$$v_n = \frac{1}{x_n - x_n^{(p)}}, \qquad n = 0, 1, \ldots, \tag{3.5.21}$$

   reduces Equation (3.5.17) to the linear first-order equation,

$$v_{n+1} + C_n v_n + D_n = 0, \qquad n = 0, 1, \ldots, \tag{3.5.22}$$

   where $C_n = \dfrac{\left(\gamma_n x_n^{(p)} + \delta_n\right)^2}{\beta_n \gamma_n - \alpha_n \delta_n}$, and $D_n = \dfrac{\gamma_n \left(\gamma_n x_n^{(p)} + \delta_n\right)}{\beta_n \gamma_n - \alpha_n \delta_n}$.

3. Let $x^{(1)}$ and $x^{(2)}$ be two particular solutions of Equation (3.5.17) with $x_n^{(1)} \neq x_n^{(2)}$ for $n = 0, 1, \ldots$. Then the change of variables,

$$w_n = \frac{1}{x_n - x_n^{(1)}} + \frac{1}{x_n^{(1)} - x_n^{(2)}}, \qquad n = 0, 1, \ldots, \tag{3.5.23}$$

   reduces Equation (3.5.17) to the linear homogeneous first-order equation,

$$w_{n+1} + E_n w_n = 0, \qquad n = 0, 1, \ldots, \tag{3.5.24}$$

   where $E_n = \dfrac{\left(\gamma_n x_n^{(1)} + \delta_n\right)^2}{\beta_n \gamma_n - \alpha_n \delta_n}$.

3.5.7.3 Logistic equation

Consider the initial value problem

$$x_{n+1} = rx_n\left(1 - \frac{x_n}{k}\right), \qquad n = 0, 1, \ldots, \tag{3.5.25}$$
$$x_0 = \alpha, \qquad \text{with } \alpha \in [0, k],$$

where r and k are positive numbers with $r \leq 4$. The following are true:

1. When $r = k = 4$, Equation (3.5.25) reduces to

$$x_{n+1} = 4x_n - x_n^2. \tag{3.5.26}$$

   If $\alpha = 4\sin^2(\theta)$ with $\theta \in [0, \frac{\pi}{2}]$, then Equation (3.5.26) has the closed-form solution

$$x_{n+1} = 4\sin^2(2^{n+1}\theta), \qquad n = 0, 1, \ldots, \tag{3.5.27}$$
$$x_0 = 4\sin^2(\theta), \qquad \text{with } \theta \in \left[0, \frac{\pi}{2}\right].$$

2. When $r = 4$ and $k = 1$, Equation (3.5.25) reduces to

$$x_{n+1} = 4x_n - 4x_n^2. \tag{3.5.28}$$

   If $\alpha = \sin^2(\theta)$ with $\theta \in \left[0, \frac{\pi}{2}\right]$, then Equation (3.5.28) has the closed-form solution

$$x_{n+1} = \sin^2(2^{n+1}\theta), \qquad n = 0, 1, \ldots, \tag{3.5.29}$$
$$x_0 = \sin^2(\theta), \qquad \text{with } \theta \in \left[0, \frac{\pi}{2}\right].$$

Chapter 4

Geometry

4.1 EUCLIDEAN GEOMETRY

Euclidean geometry is based on 5 axioms:

1. One can draw a straight line from any point to any point.
2. One can extend a finite straight line continuously in a straight line.
3. One can describe a circle with any center and radius.
4. All right angles are equal to one another.
5. If a straight line falling on two straight lines make the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side where the angles are less than the two right angles. (Parallel Postulate)

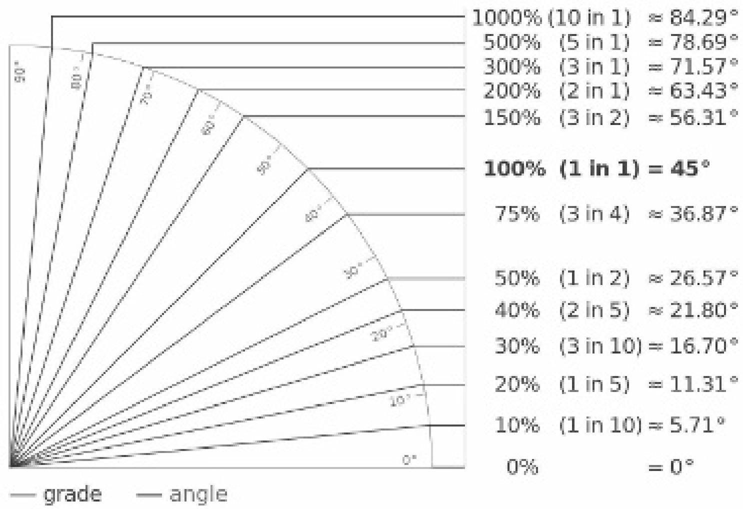
A logically equivalent formulation of the parallel postulate is Playfair's postulate:

- (5') Through a point not on a given straight line, at most one line can be drawn that never meets the given line.

The essential difference between Euclidean and non-Euclidean geometry is the nature of parallel lines. Changing the parallel postulate results in other geometries:

- (5; for *hyperbolic geometry*) Through a point not on a given straight line, infinitely many lines can be drawn that never meet the given line. For example, the surface of a hyperboloid is an example of hyperbolic geometry.
- (5; for *elliptic geometry*) Through a point not on a given straight line, no lines can be drawn that never meet the given line. For example, the surface of a sphere is an example of elliptic geometry.

4.2 GRADES AND DEGREES



File:Grades degrees.svg, CC-BY-SA-3.0, Wikimedia Commons, retrieved 1 April 2017.

4.3 COORDINATE SYSTEMS IN THE PLANE

4.3.1 CONVENTION

When we talk about "the point with coordinates (x, y)" or "the curve with equation $y = f(x)$," we always mean Cartesian coordinates. If a formula involves other coordinates, this fact will be stated explicitly.

4.3.2 SUBSTITUTIONS AND TRANSFORMATIONS

Formulas for changes in coordinate systems can lead to confusion because (for example) moving the coordinate axes *up* has the same effect on equations as moving objects *down* while the axes stay fixed. (To read the next paragraph, you can move your eyes down or slide the page up.)

To avoid confusion, we will carefully distinguish between transformations of the plane and substitutions, as explained below. Similar considerations will apply to transformations and substitutions in three dimensions (Section 4.13).

4.3.2.1 Substitutions

A *substitution*, or *change of coordinates*, relates the coordinates of a point in one coordinate system to those of *the same point in a different coordinate system*. Usually one coordinate system has the superscript $'$ and the other does not, and we write

$$\begin{cases} x = F_x(x', y'), \\ y = F_y(x', y'), \end{cases} \qquad \text{or} \qquad (x, y) = F(x', y') \tag{4.3.1}$$

(where subscripts and primes are not derivatives, they are coordinates). This means: given the equation of an object in the unprimed coordinate system, one obtains the equation of the *same* object in the primed coordinate system by substituting $F_x(x', y')$ for x and $F_y(x', y')$ for y in the equation. For instance, suppose the primed coordinate system is obtained from the unprimed system by moving the x axis up a distance d. Then $x = x'$ and $y = y' + d$. The circle with equations $x^2 + y^2 = 1$ in the unprimed system has equations ${x'}^2 + (y' + d)^2 = 1$ in the primed system. Thus, transforming an implicit equation in (x, y) into one in (x', y') is immediate.

The point $P = (a, b)$ in the unprimed system, with equation $x = a$, $y = b$, has equation $F_x(x', y') = a$, $F_y(x', y') = b$ in the new system. To get the primed coordinates explicitly, one must solve for x' and y' (in the example just given we have $x' = a$, $y' + d = b$, which yields $x' = a$, $y' = b - d$). Therefore, if possible, we give the *inverse equations*

$$\begin{cases} x' = G_{x'}(x, y), \\ y' = G_{y'}(x, y) \end{cases} \qquad \text{or} \qquad (x', y') = G(x, y),$$

which are equivalent to Equation (4.3.1) if $G(F(x', y')) = (x', y')$ and $F(G(x, y)) = (x, y)$. Then to go from the unprimed to the primed system, one merely inserts the known values of x and y into these equations. This is also the best strategy when dealing with a curve expressed parametrically, that is, $x = x(t)$, $y = y(t)$.

4.3.2.2 Transformations

A *transformation* associates with each point (x, y) *a different point in the same coordinate system*; we denote this by

$$(x, y) \mapsto F(x, y), \tag{4.3.2}$$

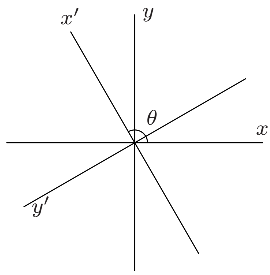
where F is a map from the plane to itself (a two-component function of two variables). For example, translating down by a distance d is accomplished by $(x, y) \mapsto (x, y - d)$ (see Section 4.4). Thus, the action of the transformation on a point whose coordinates are known (or on a curve expressed parametrically) can be immediately computed.

If, on the other hand, we have an object (say a curve) defined *implicitly* by the equation $C(x, y) = 0$, finding the equation of the transformed object requires using the *inverse transformation*

$$(x, y) \mapsto G(x, y) \tag{4.3.3}$$

defined by $G(F(x, y)) = (x, y)$ and $F(G(x, y)) = (x, y)$. The equation of the transformed object is $C(G(x, y)) = 0$. For instance, if C is the circle with equation $x^2 + y^2 = 1$ and we are translating down by a distance d, the inverse transformation is $(x, y) \mapsto (x, y + d)$ (translating up), and the equation of the translated circle is $x^2 + (y + d)^2 = 1$. Compare to the example following Equation (4.3.1).

FIGURE 4.1
Change of coordinates by a rotation.



4.3.2.3 Using transformations to perform changes of coordinates

Usually, we will not give formulas of the form (4.3.1) for changes between two coordinate systems of the same type, because they can be immediately derived from the corresponding formulas (4.3.2) for transformations, which are given in Section 4.4. We give two examples for clarity.

Let the two Cartesian coordinate systems (x, y) and (x', y') be related as follows: They have the same origin, and the positive x'-axis is obtained from the positive x-axis by a (counterclockwise) rotation through an angle θ (Figure 4.1). If a point has coordinates (x, y) in the unprimed system, its coordinates (x', y') in the primed system are the same as the coordinates in the unprimed system of a point that undergoes the *inverse rotation*, that is, a rotation by an angle $\alpha = -\theta$. According to Equation (4.4.2) (page 200), this transformation acts as follows:

$$(x, y) \mapsto \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} (x, y) = (x\cos\theta + y\sin\theta,\ -x\sin\theta + y\cos\theta). \quad (4.3.4)$$

Therefore the right-hand side of Equation (4.3.4) is (x', y'), and the desired substitution is

$$\begin{aligned} x' &= x\cos\theta + y\sin\theta, \\ y' &= -x\sin\theta + y\cos\theta. \end{aligned} \quad (4.3.5)$$

Switching the roles of the primed and unprimed systems, we get the equivalent substitution

$$\begin{aligned} x &= x'\cos\theta - y'\sin\theta, \\ y &= x'\sin\theta + y'\cos\theta \end{aligned} \quad (4.3.6)$$

(because the x-axis is obtained from the x'-axis by a rotation through an angle $-\theta$).

Similarly, let the two Cartesian coordinate systems (x, y) and (x', y') differ by a translation: x is parallel to x' and y to y', and the origin of the second system coincides with the point (x_0, y_0) of the first system. The coordinates (x, y) and (x', y') of a point are related by

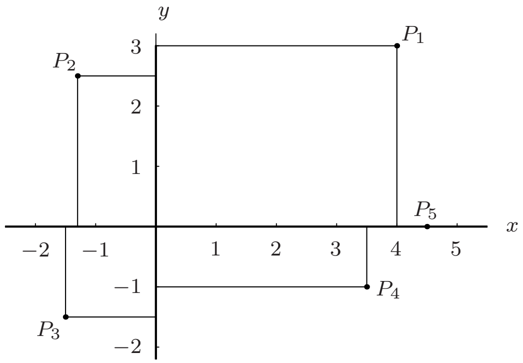
$$\begin{aligned} x &= x' + x_0, & x' &= x - x_0, \\ y &= y' + y_0, & y' &= y - y_0. \end{aligned} \quad (4.3.7)$$

4.3.3 CARTESIAN COORDINATES IN THE PLANE

In *Cartesian coordinates* (or *rectangular coordinates*), the “address” of a point P is given by two real numbers indicating the positions of the perpendicular projections from the point to two fixed, perpendicular, graduated lines, called the *axes*. If one coordinate is denoted x and the other y, the axes are called the x*-axis* and the y*-axis*, and we write $P = (x, y)$. Usually the x-axis is horizontal, with x increasing to the right, and the y-axis is vertical, with y increasing vertically up. The point $x = 0$, $y = 0$, where the axes intersect, is the *origin*. See Figure 4.2.

FIGURE 4.2

In Cartesian coordinates, $P_1 = (4,3)$*,* $P_2 = (-1.3, 2.5)$*,* $P_3 = (-1.5, -1.5)$*,* $P_4 = (3.5, -1)$*, and* $P_5 = (4.5, 0)$*. The axes divide the plane into four quadrants.* P_1 *is in the first quadrant,* P_2 *in the second,* P_3 *in the third, and* P_4 *in the fourth.* P_5 *is on the positive* x*-axis.*

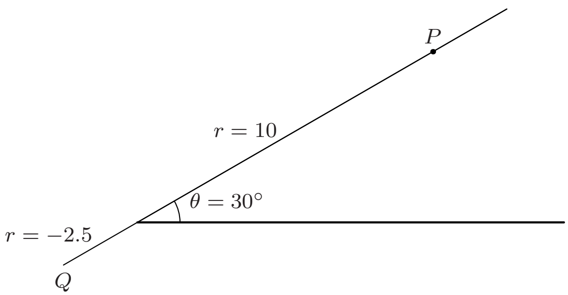


4.3.4 POLAR COORDINATES IN THE PLANE

In *polar coordinates* a point P is also characterized by two numbers: the distance $r \geq 0$ to a fixed *pole* or *origin* O, and the angle θ that the ray OP makes with a fixed ray originating at O, which is generally drawn pointing to the right (this is called the *initial ray*). The angle θ is defined only up to a multiple of $360°$ or 2π radians. In addition, it is sometimes convenient to relax the condition $r > 0$ and allow r to be a signed distance, so (r, θ) and $(-r,\ \theta + 180°)$ represent the same point (Figure 4.3).

FIGURE 4.3

Among the possible sets of polar coordinates for P *are* $(10, 30°)$*,* $(10, 390°)$ *and* $(10,\ -330°)$*. Among the sets of polar coordinates for Q are* $(2.5,\ 210°)$ *and* $(-2.5, 30°)$*.*



4.3.4.1 Relations between Cartesian and polar coordinates

Consider a system of polar coordinates and a system of Cartesian coordinates with the same origin. Assume that the initial ray of the polar coordinate system coincides with the positive x-axis, and that the ray $\theta = 90°$ coincides with the positive y-axis. Then the polar coordinates (r, θ) with $r > 0$ and the Cartesian coordinates (x, y) of the same point are related as follows (x and y are assumed positive for the θ definition):

$$\begin{cases} x = r\cos\theta, \\ y = r\sin\theta, \end{cases} \qquad \begin{cases} r = \sqrt{x^2+y^2}, \\ \theta = \tan^{-1}\dfrac{y}{x}, \end{cases} \qquad \begin{cases} \sin\theta = \dfrac{y}{\sqrt{x^2+y^2}}, \\ \cos\theta = \dfrac{x}{\sqrt{x^2+y^2}}. \end{cases}$$

4.3.5 HOMOGENEOUS COORDINATES IN THE PLANE

A triple of real numbers $(x : y : t)$, with $t \neq 0$, is a set of *homogeneous coordinates* for the point P with Cartesian coordinates $(x/t,\ y/t)$. Thus the same point has many sets of homogeneous coordinates: $(x : y : t)$ and $(x' : y' : t')$ represent the same point if and only if there is some real number α such that $x' = \alpha x, y' = \alpha y, z' = \alpha z$.

When we think of the same triple of numbers as the Cartesian coordinates of a point in three-dimensional space (page 249), we write it as (x, y, t) instead of $(x : y : t)$. The connection between the point in space with Cartesian coordinates (x, y, t) and the point in the plane with homogeneous coordinates $(x : y : t)$ becomes apparent when we consider the plane $t = 1$ in space, with Cartesian coordinates given by the first two coordinates x, y of space (Figure 4.4). The point (x, y, t) in space can be connected to the origin by a line L that intersects the plane $t = 1$ in the point with Cartesian coordinates $(x/t,\ y/t)$ or homogeneous coordinates $(x : y : t)$.

Homogeneous coordinates are useful for several reasons. One the most important is that they allow one to unify all symmetries of the plane (as well as other transformations) under a single umbrella. All of these transformations can be regarded as linear maps in the space of triples $(x : y : t)$, and so can be expressed in terms of matrix multiplications (see page 201).

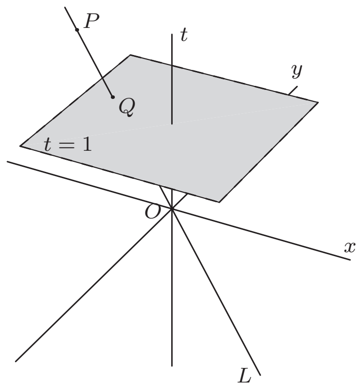
If we consider triples $(x : y : t)$ such that at least one of x, y, t is non-zero, we can name not only the points in the plane but also points "at infinity." Thus, $(x : y : 0)$ represents the point at infinity in the direction of the ray emanating from the origin going through the point (x, y).

4.3.6 OBLIQUE COORDINATES IN THE PLANE

The following generalization of Cartesian coordinates is sometimes useful. Consider two *axes* (graduated lines), intersecting at the *origin* but not necessarily perpendicularly. Let the angle between them be ω. In this system of *oblique coordinates*, a point P is given by two real numbers indicating the positions of the projections from the point to each axis, in the direction of the other axis (see Figure 4.5). The first

FIGURE 4.4

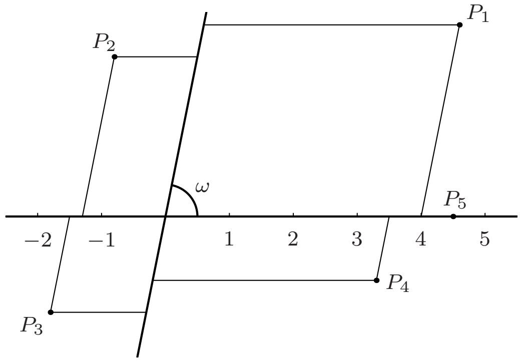
The point P with spatial coordinates (x, y, t) projects to the point Q with spatial coordinates $(x/t,\ y/t,\ 1)$. The plane Cartesian coordinates of Q are $(x/t,\ y/t)$, and $(x : y : t)$ is one set of homogeneous coordinates for Q. Any point on the line L (except for the origin O) would also project to P'.



axis (x-axis) is generally drawn horizontally. The case $\omega = 90°$ yields a Cartesian coordinate system.

FIGURE 4.5

In oblique coordinates, $P_1 = (4, 3)$, $P_2 = (-1.3, 2.5)$, $P_3 = (-1.5, -1.5)$, $P_4 = (3.5, -1)$, and $P_5 = (4.5, 0)$. Compare to Figure 4.2.



4.3.6.1 Relations between two oblique coordinate systems

Let the two oblique coordinate systems (x, y) and (x', y'), with angles ω and ω', share the same origin, and suppose the positive x'-axis makes an angle θ with the positive x-axis. The coordinates (x, y) and (x', y') of a point in the two systems are related by

$$x = \frac{x' \sin(\omega - \theta) + y' \sin(\omega - \omega' - \theta)}{\sin \omega}, \\ y = \frac{x' \sin \theta + y' \sin(\omega' + \theta)}{\sin \omega}. \tag{4.3.8}$$

This formula also covers passing from a Cartesian system to an oblique system and vice versa, by taking $\omega = 90°$ or $\omega' = 90°$.

The relation between two oblique coordinate systems that differ by a translation is the same as for Cartesian systems. See Equation (4.3.7).

4.4 PLANE SYMMETRIES OR ISOMETRIES

A transformation of the plane (invertible map of the plane to itself) that preserves distances is called an *isometry* of the plane. Every isometry of the plane is of one of the following types:

1. The *identity* (which leaves every point fixed)
2. A *translation* by a vector **v**
3. A *rotation* through an angle α around a point P
4. A *reflection* in a line L
5. A *glide-reflection* in a line L with displacement d

4.4.1 SYMMETRIES: CARTESIAN COORDINATES

In the formulas below, a multiplication between a matrix and a pair of coordinates should be carried out regarding the pair as a column vector (i.e., a matrix with two rows and one column). Thus $\left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right](x, y) = (ax + by,\ cx + dy)$.

1. *Translation* by (x_0, y_0):
$$(x, y) \mapsto (x + x_0,\ y + y_0). \tag{4.4.1}$$
2. *Rotation* through α (counterclockwise) around the origin:
$$(x, y) \mapsto \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} (x, y). \tag{4.4.2}$$
3. *Rotation* through α (counterclockwise) around an arbitrary point (x_0, y_0):
$$(x, y) \mapsto (x_0, y_0) + \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix} (x - x_0,\ y - y_0). \tag{4.4.3}$$
4. *Reflection:*
$$\begin{array}{rl} \text{in the } x\text{-axis:} & (x, y) \mapsto (x,\ -y), \\ \text{in the } y\text{-axis:} & (x, y) \mapsto (-x,\ y), \\ \text{in the diagonal } x = y\text{:} & (x, y) \mapsto (y, x). \end{array} \tag{4.4.4}$$
5. *Reflection* in a line with equation $ax + by + c = 0$:
$$(x, y) \mapsto \frac{1}{a^2 + b^2} \left(\begin{bmatrix} b^2 - a^2 & -2ab \\ -2ab & a^2 - b^2 \end{bmatrix} (x, y) - (2ac, 2bc) \right). \tag{4.4.5}$$

6. *Reflection* in a line going through (x_0, y_0) and making an angle α with the x-axis:
$$(x, y) \mapsto (x_0, y_0) + \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{bmatrix} (x - x_0,\ y - y_0). \qquad (4.4.6)$$

7. *Glide-reflection* in a line L with displacement d: Apply first a reflection in L, then a translation by a vector of length d in the direction of L, that is, by the vector
$$\frac{1}{a^2 + b^2}(\pm ad, \mp bd) \qquad (4.4.7)$$
if L has equation $ax + by + c = 0$.

4.4.2 SYMMETRIES: HOMOGENEOUS COORDINATES

All isometries of the plane can be expressed in homogeneous coordinates in terms of multiplication by a matrix. This fact is useful in implementing these transformations on a computer. It also means that the successive application of transformations reduces to matrix multiplication. The corresponding matrices are as follows:

1. *Translation* by (x_0, y_0):
$$T_{(x_0,y_0)} = \begin{bmatrix} 1 & 0 & x_0 \\ 0 & 1 & y_0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (4.4.8)$$

2. *Rotation* through α around the origin:
$$R_\alpha = \begin{bmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (4.4.9)$$

3. *Reflection* in a line going through the origin and making an angle α with the x-axis:
$$M_\alpha = \begin{bmatrix} \cos 2\alpha & \sin 2\alpha & 0 \\ \sin 2\alpha & -\cos 2\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix}. \qquad (4.4.10)$$

From this, one can deduce all other transformations.

EXAMPLE To find the matrix for a rotation through α around an arbitrary point $P = (x_0, y_0)$, we apply a translation by $-(x_0, y_0)$ to move P to the origin, a rotation through α around the origin, and then a translation by (x_0, y_0):
$$T_{(x_0,y_0)} R_\alpha T_{-(x_0,y_0)} = \begin{bmatrix} \cos\alpha & -\sin\alpha & x_0 - x_0\cos\alpha + y_0\sin\alpha \\ \sin\alpha & \cos\alpha & y_0 - y_0\cos\alpha - x_0\sin\alpha \\ 0 & 0 & 1 \end{bmatrix} \qquad (4.4.11)$$
(Note the order of the multiplication.)

4.4.3 SYMMETRIES: POLAR COORDINATES

1. *Rotation around the origin* through an angle α: $(r, \theta) \mapsto (r,\ \theta + \alpha)$.
2. *Reflection in a line through the origin* and making an angle α with the positive x-axis: $(r, \theta) \mapsto (r,\ 2\alpha - \theta)$.

4.4.4 CRYSTALLOGRAPHIC GROUPS

A group of symmetries of the plane that is doubly infinite is a *wallpaper group*, or *crystallographic group*. There are 17 types of such groups, corresponding to 17 essentially distinct ways to tile the plane in a doubly periodical pattern. (There are also 230 three-dimensional crystallographic groups.)

4.4.4.1 Crystallographic group classification

To classify an image representing a crystallographic group, answer the following sequence of questions starting with: "What is the minimal rotational invariance?".

- None
  Is there a reflection?
  - No.
    Is there a glide-reflection?
    * No: p1
    * Yes: pg
  - Yes.
    Is there a glide-reflection in an axis that is not a reflection axis?
    * No: pm
    * Yes: cm
- 2-fold (180° rotation);
  Is there a reflection?
  - No.
    Is there a glide-reflection?
    * No: p2
    * Yes: pgg
  - Yes.
    Are there reflections in two directions?
    * No: pmg
    * Yes: Are all rotation centers on reflection axes?
      · No: cmm
      · Yes: pmm
- 3-fold (120° rotation);
  Is there a reflection?
  - No: p3
  - Yes.
    Are all centers of threefold rotations on reflection axes?
    * No: p31m
    * Yes: p3m1
- 4-fold (90° rotation);
  Is there a reflection?
  - No: p4
  - Yes.
    Are there four reflection axes?
    * No: p4g
    * Yes: p4m
- 6-fold (60° rotation);
  Is there a reflection?
  - No: p6
  - Yes: p6m

4.4.4.2 Crystallographic groups descriptions

The simplest crystallographic group involves translations only (page 204, top left). The others involve, in addition to translations, one or more of the other types of symmetries (rotations, reflections, glide-reflections). The *Conway notation* for crystallographic groups is based on the types of non-translational symmetries occurring in the "simplest description" of the group:

1. $^\circ$ indicates a translations only,
2. * indicates a reflection (mirror symmetry),
3. $^\times$ a glide-reflection,
4. a number n indicates a rotational symmetry of order n (rotation by $360^\circ/n$).

In addition, if a number n comes after the *, the center of the corresponding rotation lies on mirror lines, so that the symmetry there is actually dihedral of order $2n$.

Thus the group ** has two inequivalent lines of mirror symmetry; the group 333 has three inequivalent centers of order-3 rotation; the group 22^* has two inequivalent centers of order-2 rotation as well as mirror lines; and ^{*632} has points of dihedral symmetry of order 12 $(= 2 \times 6)$, 6, and 4.

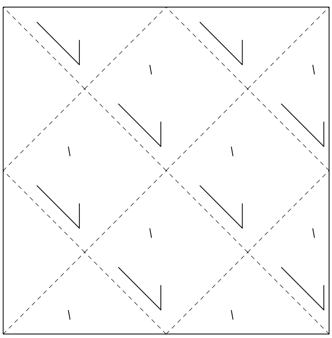
The table on page 204 gives the groups in the Conway notation and in the notation traditional in crystallography. It also gives the quotient space of the plane by the action of the group. The entry "4,4,2 turnover" means the surface of a triangular puff pastry with corner angles $45^\circ (= 180^\circ/4)$, 45° and 90°. The entry "4,4,2 turnover slit along 2,4" means the same surface, slit along the edge joining a 45° vertex to the 90° vertex. Open edges are silvered (mirror lines); such edges occur exactly for those groups whose Conway notation includes a *.

The last column of the table gives the dimension of the space of inequivalent groups of the given type (equivalent groups are those that can be obtained from one another by proportional scaling or rigid motion). For instance, there is a group of type $^\circ$ for every shape parallelogram, and there are two degrees of freedom for the choice of such a shape (say the ratio and angle between sides). Thus, the $^\circ$ group is based on a square fundamental domain, while for the $^\circ$ group a fundamental parallelogram would have the shape of two juxtaposed equilateral triangles. These two groups are inequivalent, although they are of the same type.

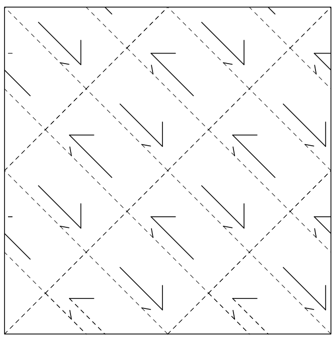
Conway	Cryst	Quotient space	Dim
∘	p1	Torus	2
××	pg	Klein bottle	1
**	pm	Cylinder	1
×*	cm	Möbius strip	1
22×	pgg	Non-orientable football	1
22*	pmg	Open pillowcase	1
2222	p2	Closed pillowcase	2
2*22	cmm	4,4,2 turnover, slit along 4,4	1
*2222	pmm	Square	1
442	p4	4,4,2 turnover	0
4*2	p4g	4,4,2 turnover, slit along 4,2	0
*442	p4m	4,4,2 triangle	0
333	p3	3,3,3 turnover	0
*333	p3m1	3,3,3 triangle	0
3*3	p31m	6,3,2 turnover, slit along 3,2	0
632	p6	6,3,2 turnover	0
*632	p6m	6,3,2 triangle	0

The following figures show wallpaper patterns for each of the 17 types of crystallographic groups (two patterns are shown for the ∘, or translations-only, type). Thin lines bound *unit cells*, or *fundamental domains*. When solid, they represent lines of mirror symmetry, and are fully determined. When dashed, they represent arbitrary boundaries, which can be shifted so as to give different fundamental domains. Dots at the intersections of thin lines represent centers of rotational symmetry.

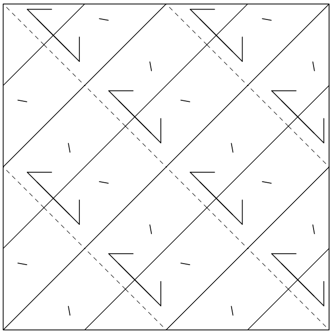
Some of the relationships between the types are made obvious by the patterns. However, there are more relationships than can be indicated in a single set of pictures. For instance, there is a group of type ×× hiding in any group of type 3*3.



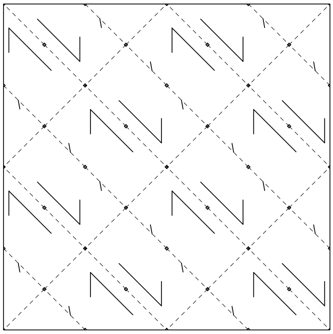
∘ p1



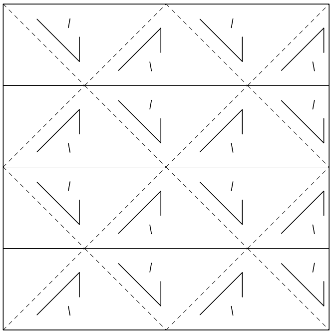
×× pg



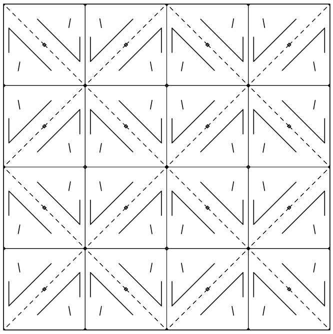
** pm



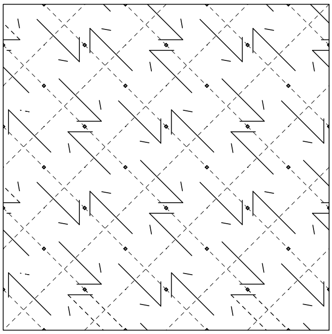
2222 p2



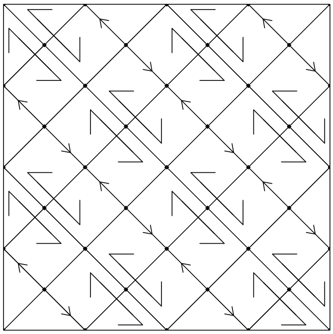
×* cm



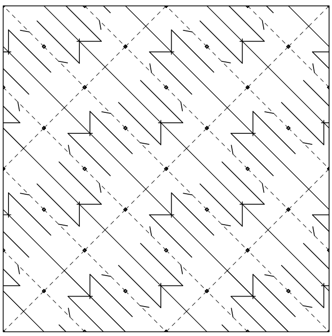
2*22 cmm



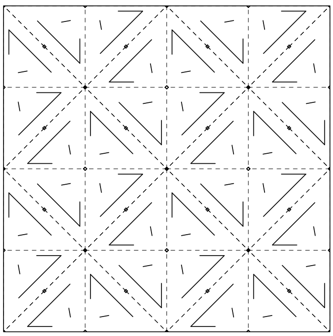
22× pgg



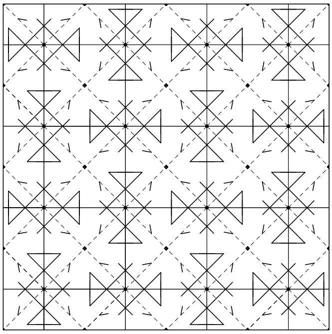
*2222 pmm



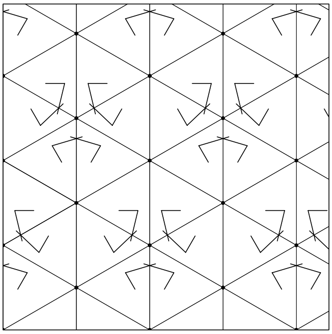
22* pmg



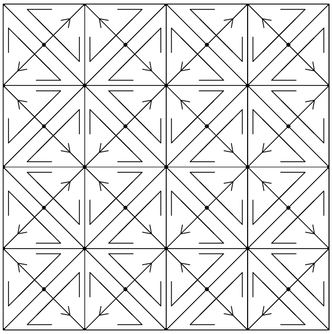
442 p4



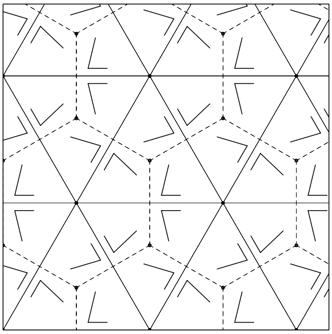
4*2 p4g



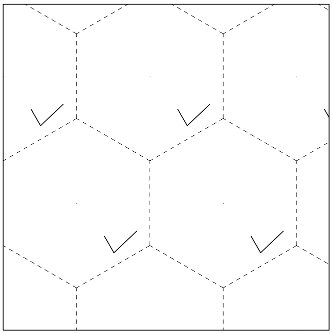
*333 p3m1



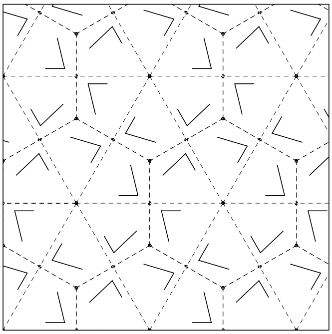
*442 p4m



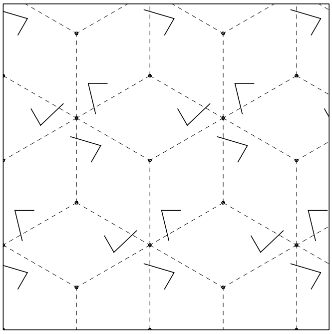
3*3 p31m



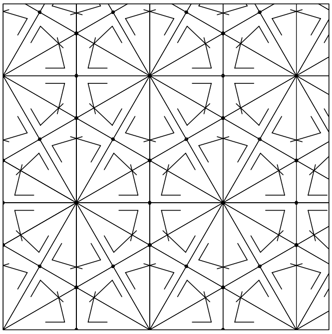
∘ p1



632 p6



333 p3



*632 p6m

4.5 OTHER TRANSFORMATIONS OF THE PLANE

4.5.1 SIMILARITIES

A transformation of the plane that preserves shapes is called a *similarity*. Every similarity of the plane is obtained by composing a *proportional scaling transformation* (also known as a *homothety*) with an isometry. A proportional scaling transformation centered at the origin has the form

$$(x, y) \mapsto (ax, ay), \tag{4.5.1}$$

where $a \neq 0$ is the *scaling factor* (a real number). The corresponding matrix in *homogeneous coordinates* is

$$H_a = \begin{bmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.5.2}$$

In *polar coordinates*, the transformation is $(r, \theta) \mapsto (ar, \theta)$.

4.5.2 AFFINE TRANSFORMATIONS

A transformation that preserves lines and parallelism (maps parallel lines to parallel lines) is an *affine transformation*. There are two important particular cases of such transformations:

A *non-proportional scaling transformation* centered at the origin has the form $(x, y) \mapsto (ax, by)$, where $a, b \neq 0$ are the *scaling factors* (real numbers). The corresponding matrix in *homogeneous coordinates* is

$$H_{a,b} = \begin{bmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.5.3}$$

A *shear* preserving horizontal lines has the form $(x, y) \mapsto (x + ry, y)$, where r is the *shearing factor* (see Figure 4.6). The corresponding matrix in *homogeneous coordinates* is

$$S_r = \begin{bmatrix} 1 & r & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{4.5.4}$$

Every affine transformation is obtained by composing a scaling transformation with an isometry, or a shear with a homothety and an isometry.

FIGURE 4.6
A shear with factor $r = \frac{1}{2}$.

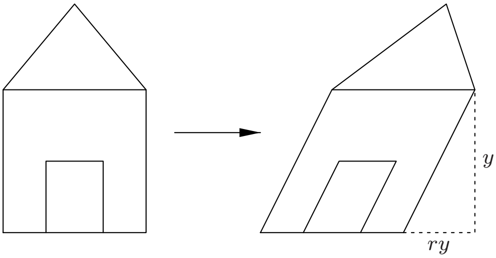
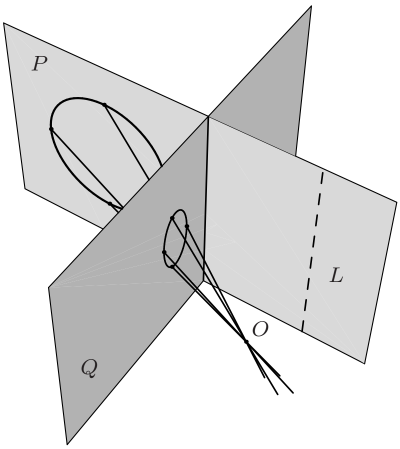


FIGURE 4.7
A perspective transformation with center O, mapping the plane P to the plane Q. The transformation is not defined on the line L, where P intersects the plane parallel to Q and going through O.



4.5.3 PROJECTIVE TRANSFORMATIONS

A transformation that maps lines to lines (but does not necessarily preserve parallelism) is a *projective transformation*. Any plane projective transformation can be expressed by an invertible 3×3 matrix in homogeneous coordinates; conversely, any invertible 3×3 matrix defines a projective transformation of the plane. Projective transformations (if not affine) are not defined on all of the plane but only on the complement of a line (the missing line is "mapped to infinity").

A common example of a projective transformation is given by a *perspective transformation* (Figure 4.7). Strictly speaking this gives a transformation from one plane to another, but, if we identify the two planes by (for example) fixing a Cartesian system in each, we get a projective transformation from the plane to itself.

4.6 LINES

The (Cartesian) equation of a *straight line* is linear in the coordinates x and y:

$$ax + by + c = 0. \tag{4.6.1}$$

The *slope* of this line is $-a/b$, the *intersection with the x-axis* (or *x-intercept*) is $x = -c/a$, and the *intersection with the y-axis* (or y-intercept) is $y = -c/b$. If $a = 0$, the line is parallel to the x-axis, and if $b = 0$, then the line is parallel to the y-axis.

(In an *oblique coordinate system*, everything in the preceding paragraph remains true, except for the value of the slope.)

When $a^2 + b^2 = 1$ and $c \le 0$ in the equation $ax + by + c = 0$, the equation is said to be in *normal form*. In this case $-c$ is the *distance of the line to the origin*, and ω (with $\cos\omega = a$ and $\sin\omega = b$) is the angle that the perpendicular dropped to the line from the origin makes with the positive x-axis (Figure 4.8, with $p = -c$).

To reduce an arbitrary equation $ax + by + c = 0$ to normal form, divide by $\pm\sqrt{a^2 + b^2}$, where the sign of the radical is chosen opposite the sign of c when $c \ne 0$ and the same as the sign of b when $c = 0$.

4.6.1 LINES WITH PRESCRIBED PROPERTIES

1. Line of slope m intersecting the x-axis at $x = x_0$: $y = m(x - x_0)$.
2. Line of slope m intersecting the y-axis at $y = y_0$: $y = mx + y_0$.
3. Line intersecting the x-axis at $x = x_0$ and the y-axis at $y = y_0$:

$$\frac{x}{x_0} + \frac{y}{y_0} = 1. \tag{4.6.2}$$

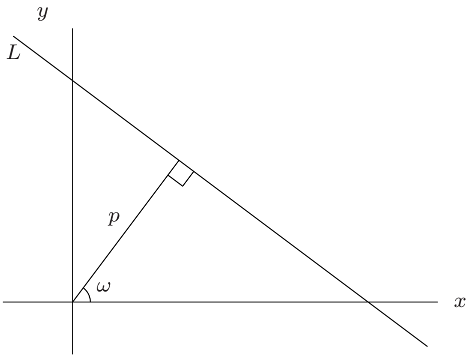
   (This formula remains true in *oblique coordinates*.)
4. Line of slope m passing through (x_0, y_0): $y - y_0 = m(x - x_0)$.
5. Line passing through points (x_0, y_0) and (x_1, y_1):

$$\frac{y - y_1}{x - x_1} = \frac{y_0 - y_1}{x_0 - x_1} \qquad \text{or} \qquad \begin{vmatrix} x & y & 1 \\ x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \end{vmatrix} = 0. \tag{4.6.3}$$

   (These formulas remain true in *oblique coordinates*.)
6. *Slope* of line going through points (x_0, y_0) and (x_1, y_1): $\dfrac{y_1 - y_0}{x_1 - x_0}$.
7. Line passing through points with *polar coordinates* (r_0, θ_0) and (r_1, θ_1):

$$r(r_0 \sin(\theta - \theta_0) - r_1 \sin(\theta - \theta_1)) = r_0 r_1 \sin(\theta_1 - \theta_0). \tag{4.6.4}$$

FIGURE 4.8
The normal form of the line L is $x\cos\omega + y\sin\omega = p$.



4.6.2 DISTANCES

The *distance* between two points in the plane is the *length of the line segment* joining the two points. If the points have *Cartesian coordinates* (x_0, y_0) and (x_1, y_1), this distance is

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2}. \tag{4.6.5}$$

If the points have *polar coordinates* (r_0, θ_0) and (r_1, θ_1), this distance is

$$\sqrt{r_0^2 + r_1^2 - 2r_0 r_1 \cos(\theta_0 - \theta_1)}. \tag{4.6.6}$$

If the points have *oblique coordinates* (x_0, y_0) and (x_1, y_1), this distance is

$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + 2(x_1 - x_0)(y_1 - y_0)\cos\omega}, \tag{4.6.7}$$

where ω is the angle between the axes (Figure 4.5).

The point $k\%$ of the way from $P_0 = (x_0, y_0)$ to $P_1 = (x_1, y_1)$ is

$$\left(\frac{kx_1 + (100 - k)x_0}{100}, \frac{ky_1 + (100 - k)y_0}{100}\right). \tag{4.6.8}$$

(The same formula also works in oblique coordinates.) This point divides the segment P_0P_1 in the ratio $k : (100 - k)$. As a particular case, the *midpoint* of P_0P_1 is given by $\left(\frac{1}{2}(x_0 + x_1), \frac{1}{2}(y_0 + y_1)\right)$.

The *distance* from the point (x_0, y_0) to the line $ax + by + c = 0$ is

$$\left|\frac{ax_0 + by_0 + c}{\sqrt{a^2 + b^2}}\right|. \tag{4.6.9}$$

4.6.3 ANGLES

The *angle* between two lines $a_0x + b_0y + c_0 = 0$ and $a_1x + b_1y + c_1 = 0$ is

$$\tan^{-1}\left(\frac{b_1}{a_1}\right) - \tan^{-1}\left(\frac{b_0}{a_0}\right) = \tan^{-1}\left(\frac{a_0b_1 - a_1b_0}{a_0a_1 + b_0b_1}\right). \tag{4.6.10}$$

In particular, the two lines are *parallel* when $a_0b_1 = a_1b_0$, and *perpendicular* when $a_0a_1 = -b_0b_1$.

The *angle* between two lines of slopes m_0 and m_1 is $\tan^{-1}(m_1) - \tan^{-1}(m_0)$ (or $\tan^{-1}((m_1 - m_0)/(1 + m_0m_1))$). In particular, the two lines are *parallel* when $m_0 = m_1$ and *perpendicular* when $m_0m_1 = -1$.

4.6.4 CONCURRENCE AND COLLINEARITY

Three lines $a_0x + b_0y + c_0 = 0$, $a_1x + b_1y + c_1 = 0$, and $a_2x + b_2y + c_2 = 0$ are *concurrent* (i.e., intersect at a single point) if and only if

$$\begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0. \tag{4.6.11}$$

(This remains true in *oblique coordinates*.)

Three points (x_0, y_0), (x_1, y_1) and (x_2, y_2) are *collinear* (i.e., all three points are on a straight line) if and only if

$$\begin{vmatrix} x_0 & y_0 & 1 \\ x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \end{vmatrix} = 0. \tag{4.6.12}$$

(This remains true in *oblique coordinates*.)

Three points with polar coordinates (r_0, θ_0), (r_1, θ_1) and (r_2, θ_2) are collinear if and only if

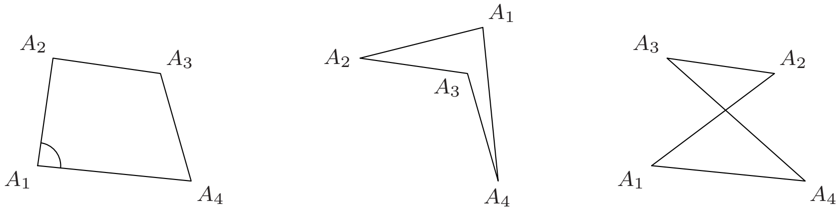
$$r_1r_2\sin(\theta_2 - \theta_1) + r_0r_1\sin(\theta_1 - \theta_0) + r_2r_0\sin(\theta_0 - \theta_2) = 0. \tag{4.6.13}$$

4.7 POLYGONS

Given $k \geq 3$ points $A_1, \ldots, A_k$ in the plane, in a certain order, we obtain a *k-sided polygon* or *k-gon* by connecting each point to the next, and the last to the first, with a line segment. The points A_i are the *vertices* and the segments A_iA_{i+1} are the *sides* or *edges* of the polygon. When $k = 3$ we have a *triangle*, when $k = 4$ we have a *quadrangle* or *quadrilateral*, and so on (see page 218 for names of regular polygons). Here we assume that all polygons are *simple*: no consecutive edges are on the same line and no two edges intersect (except that consecutive edges intersect at the common vertex).

FIGURE 4.9

Two simple quadrilaterals (left and middle) and one that is not simple (right). We will treat only simple polygons.



When we refer to the *angle* at a vertex A_k we have in mind the interior angle (as marked in the leftmost polygon in Figure 4.9). We denote this angle by the same symbol as the vertex. The complement of A_k is the *exterior angle* at that vertex; geometrically, it is the angle between one side and the extension of the adjacent side. *In any k-gon, the sum of the angles equals $2(k-2)$ right angles, or $2(k-2)\times 90°$*; for example, the sum of the angles of a triangle is $180°$.

The *area* of a polygon whose vertices A_i have coordinates (x_i, y_i), for $1 \le i \le k$, is the absolute value of

$$\begin{aligned}\text{area} &= \tfrac{1}{2}(x_1y_2 - x_2y_1) + \cdots + \tfrac{1}{2}(x_{k-1}y_k - x_ky_{k-1}) + \tfrac{1}{2}(x_ky_1 - x_1y_k),\\ &= \frac{1}{2}\sum_{i=1}^{k}(x_iy_{i+1} - x_{i+1}y_i),\end{aligned} \tag{4.7.1}$$

where in the summation we take $x_{k+1} = x_1$ and $y_{k+1} = y_1$. For a triangle

$$\text{area} = \tfrac{1}{2}(x_1y_2 - x_2y_1 + x_2y_3 - x_3y_2 + x_3y_1 - x_1y_3) = \frac{1}{2}\begin{vmatrix} x_1 & y_1 & 1\\ x_2 & y_2 & 1\\ x_3 & y_3 & 1\end{vmatrix}. \tag{4.7.2}$$

In *oblique coordinates* with angle ω between the axes, the area is as given above, multiplied by $\sin\omega$.

If the vertices have *polar coordinates* (r_i, θ_i), for $1 \le i \le k$, the area is the absolute value of

$$\text{area} = \frac{1}{2}\sum_{i=1}^{k} r_ir_{i+1}\sin(\theta_{i+1} - \theta_i), \tag{4.7.3}$$

where we take $r_{k+1} = r_1$ and $\theta_{k+1} = \theta_1$.

4.7.1 TRIANGLES

Because the angles of a triangle add up to $180°$, at least two of them must be acute (less than $90°$). In an *acute triangle* all angles are acute. A *right triangle* has one right angle, and an *obtuse triangle* has one obtuse angle.

The *altitude* corresponding to a side is the perpendicular dropped to the line containing that side from the opposite vertex. The *bisector* of a vertex is the line that divides the angle at that vertex into two equal parts. The *median* is the segment joining a vertex to the midpoint of the opposite side. See Figure 4.10.

Every triangle also has an *inscribed circle* tangent to its sides and interior to the triangle (in other words, any three non-concurrent lines determine a circle). The center of this circle is the point of intersection of the bisectors. We denote the radius of the inscribed circle by r.

Every triangle has a *circumscribed circle* going through its vertices; in other words, any three non-collinear points determine a circle. The point of intersection of the medians is the center of mass of the triangle (considered as an area in the plane). We denote the radius of the circumscribed circle by R.

Introduce the following notations for an *arbitrary triangle* of vertices A, B, C and sides a, b, c (see Figure 4.10). Let h_c, t_c, and m_c be the lengths of the altitude, bisector and median originating in vertex C, let r and R be as usual the radii of the inscribed and circumscribed circles, and let s be the semi-perimeter: $s = \frac{1}{2}(a+b+c)$.

A triangle is *equilateral* if all of its sides have the same length, or, equivalently, if all of its angles are the same (and equal to $60°$). It is *isosceles* if two sides are the same, or, equivalently, if two angles are the same. Otherwise it is *scalene*.

For an *equilateral triangle* of side a we have

$$\text{area} = \tfrac{1}{4}a^2\sqrt{3}, \quad r = \tfrac{1}{6}a\sqrt{3}, \quad R = \tfrac{1}{3}a\sqrt{3}, \quad h = \tfrac{1}{2}a\sqrt{3}, \tag{4.7.4}$$

where h is any altitude. The altitude, the bisector, and the median for each vertex coincide.

For an *isosceles triangle*, the altitude for the unequal side is also the corresponding bisector and median, but this is not true for the other two altitudes. Many formulas for an isosceles triangle of sides a, a, c can be immediately derived from those for a right triangle whose legs are a and $\frac{1}{2}c$ (see Figure 4.11, left).

For a *right triangle*, the *hypotenuse* is the longest side and (opposite the right angle); the *legs* are the two shorter sides (adjacent to the right angle). The altitude for each leg equals the other leg. In Figure 4.11 (right), h denotes the altitude for the hypotenuse, while m and n denote the segments into which this altitude divides the hypotenuse.

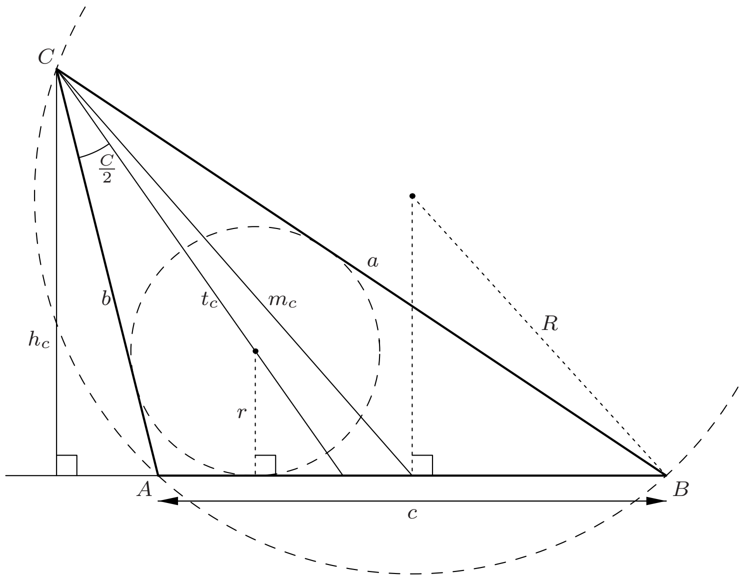
The following formulas apply to a right triangle:

$$\begin{aligned}
A + B &= 90°, & c^2 &= a^2 + b^2 \qquad \textit{(Pythagoras),}\\
r &= \frac{ab}{a+b+c}, & R &= \tfrac{1}{2}c,\\
a &= c\sin A = c\cos B, & b &= c\sin B = c\cos A,\\
mc &= b^2, & nc &= a^2,\\
\text{area} &= \tfrac{1}{2}ab, & hc &= ab, \qquad mn = h^2.
\end{aligned}$$

The hypotenuse is a diameter of the circumscribed circle. The median joining the midpoint of the hypotenuse (the center of the circumscribed circle) to the right angle makes angles $2A$ and $2B$ with the hypotenuse.

FIGURE 4.10

Notations for an arbitrary triangle of sides a, b, c and vertices A, B, C. The altitude corresponding to C is h_c, the median is m_c, the bisector is t_c. The radius of the circumscribed circle is R, that of the inscribed circle is r.



$$A + B + C = 180^\circ$$

$$c^2 = a^2 + b^2 - 2ab\cos C \qquad \text{(\textit{law of cosines})},$$

$$a = b\cos C + c\cos B,$$

$$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C} \qquad \text{(\textit{law of sines})},$$

$$\text{area} = \tfrac{1}{2}h_c c = \tfrac{1}{2}ab\sin C = \frac{c^2 \sin A \sin B}{2\sin C} = rs = \frac{abc}{4R},$$

$$= \sqrt{s(s-a)(s-b)(s-c)} \qquad \text{(\textit{Heron formula})},$$

$$r = c\sin(\tfrac{1}{2}A)\sin(\tfrac{1}{2}B)\sec(\tfrac{1}{2}C) = \frac{ab\sin C}{2s} = (s-c)\tan(\tfrac{1}{2}C),$$

$$= \left(\frac{1}{h_a} + \frac{1}{h_b} + \frac{1}{h_c}\right)^{-1},$$

$$R = \frac{c}{2\sin C} = \frac{abc}{4\,\text{area}},$$

$$h_c = a\sin B = b\sin A = \frac{2\,\text{area}}{c},$$

$$t_c = \frac{2ab}{a+b}\cos\tfrac{1}{2}C = \sqrt{ab\left(1 - \frac{c^2}{(a+b)^2}\right)}, \quad \text{and}$$

$$m_c = \sqrt{\tfrac{1}{2}a^2 + \tfrac{1}{2}b^2 - \tfrac{1}{4}c^2}.$$

FIGURE 4.11

Left: an isosceles triangle can be divided into two congruent right triangles. Right: notations for a right triangle.

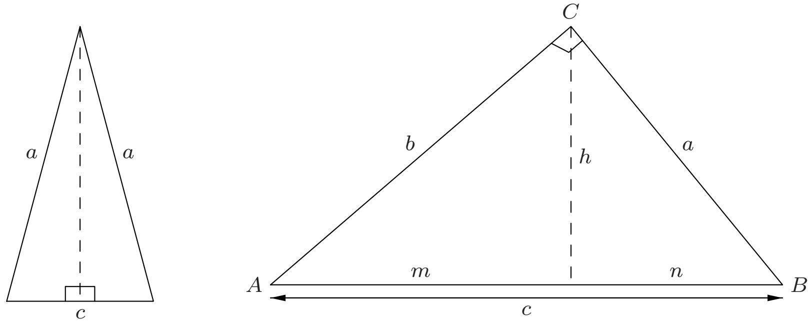
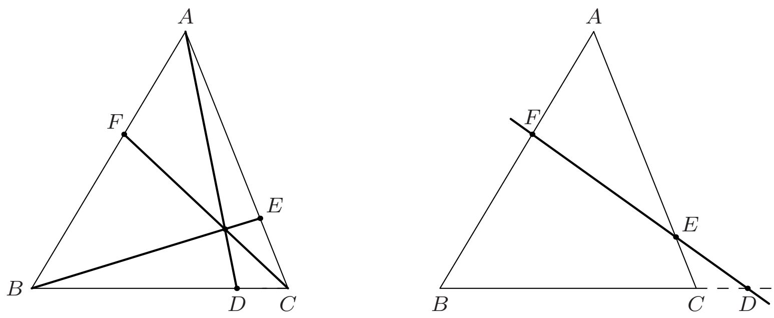


FIGURE 4.12

Left: Ceva's theorem. Right: Menelaus's theorem.



Additional facts about triangles:

1. In any triangle, the longest side is opposite the largest angle, and the shortest side is opposite the smallest angle. This follows from the law of sines.
2. *Ceva's theorem* (see Figure 4.12, left): In a triangle ABC, let D, E, and F be points on the lines BC, CA, and AB, respectively. Then the lines AD, BE, and CF are concurrent if and only if the signed distances BD, CE, $\ldots$ satisfy

$$BD \cdot CE \cdot AF = DC \cdot EA \cdot FB. \tag{4.7.5}$$

   This is so in three important particular cases: when the three lines are the medians, when they are the bisectors, and when they are the altitudes.
3. *Menelaus's theorem* (see Figure 4.12, right): In a triangle ABC, let D, E, and F be points on the lines BC, CA, and AB, respectively. Then D, E, and F are collinear if and only if the signed distances BD, CE, $\ldots$ satisfy

$$BD \cdot CE \cdot AF = -DC \cdot EA \cdot FB. \tag{4.7.6}$$

4. Each side of a triangle is less than the sum of the other two. For any three lengths such that each is less than the sum of the other two, there is a triangle with these side lengths.
5. Determining if a point is inside a triangle.
   Given a triangle's vertices $\{P_0, P_1, P_2\}$ and the test point P_3. Place P_0 at the origin by subtracting its coordinates from each of the others. Then compute (here $P_i = (x_i, y_i)$)

$$\begin{aligned} a &= x_1y_2 - x_2y_1, \\ b &= x_1y_3 - x_3y_1, \\ c &= x_2y_3 - x_3y_2. \end{aligned} \tag{4.7.7}$$

   The point P_3 is inside the triangle $\{P_0, P_1, P_2\}$ if and only if

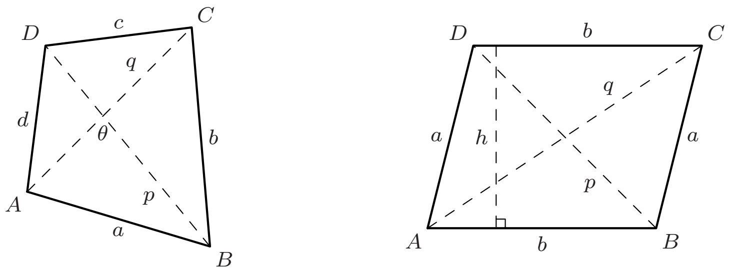
$$ab > 0 \quad \text{and} \quad ac < 0 \quad \text{and} \quad a(a - b + c) > 0. \tag{4.7.8}$$

4.7.2 QUADRILATERALS

The following formulas give the area of a *general quadrilateral* (see Figure 4.13, left, for the notation).

$$\begin{aligned} \text{area} &= \tfrac{1}{2}pq\sin\theta = \tfrac{1}{4}(b^2 + d^2 - a^2 - c^2)\tan\theta \\ &= \tfrac{1}{4}\sqrt{4p^2q^2 - (b^2 + d^2 - a^2 - c^2)^2} \\ &= \sqrt{(s-a)(s-b)(s-c)(s-d) - abcd\cos^2\left[\tfrac{1}{2}(A + C)\right]}. \end{aligned} \tag{4.7.9}$$

FIGURE 4.13
Left: notation for a general quadrilateral; in addition $s = \frac{1}{2}(a + b + c + d)$. *Right: a parallelogram.*



Often, however, it is easiest to compute the area by dividing the quadrilateral into triangles. One can also divide the quadrilateral into triangles to compute one side given the other sides and angles, etc.

More formulas can be given for *special cases* of quadrilaterals. In a *parallelogram*, opposite sides are parallel and the diagonals intersect in the middle (Figure 4.13, right). It follows that opposite sides have the same length and that two consecutive angles add up to $180°$. In the notation of the figure, we have

$$\begin{aligned} &A = C, \quad B = D, && A + B = 180°, \\ &h = a \sin A = a \sin B, && \text{area} = bh, \\ &p = \sqrt{a^2 + b^2 - 2ab\cos A}, && q = \sqrt{a^2 + b^2 - 2ab\cos B}. \end{aligned}$$

(All this follows from the triangle formulas applied to the triangles ABD and ABC.) From the last terms we obtain the "parallelogram law": $p^2 + q^2 = a^2 + b^2$ or "the sums of the squares of the diagonals equals the sum of the squares of the sides".

Two particular cases of parallelograms are

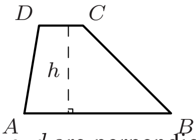
1. The *rectangle* ▭, where all angles equal $90°$. The diagonals of a rectangle have the same length. The general formulas for parallelograms reduce to

$$h = a, \quad \text{area} = ab, \quad \text{and} \quad p = q = \sqrt{a^2 + b^2}. \tag{4.7.10}$$

2. The *rhombus* or *diamond* ◇, where adjacent sides have the same length ($a = b$). The diagonals of a rhombus are perpendicular. In addition to the general formulas for parallelograms, we have $\text{area} = \frac{1}{2}pq$ and $p^2 + q^2 = 4a^2$.

The *square* or regular quadrilateral is both a rectangle and a rhombus.

A quadrilateral is a *trapezoid* if two sides are parallel. In the notation of the figure on the right we have



$$A + D = B + C = 180°, \quad \text{area} = \tfrac{1}{2}(AB + CD)h.$$

The diagonals of a quadrilateral with consecutive sides a, b, c, d are perpendicular if and only if $a^2 + c^2 = b^2 + d^2$.

A quadrilateral is *cyclic* if it can be inscribed in a circle, that is, if its four vertices belong to a single, circumscribed, circle. This is possible if and only if the sum of opposite angles is $180°$. If R is the radius of the circumscribed circle, we have (in the notation of Figure 4.13, left)

$$\begin{aligned} \text{area} &= \sqrt{(s-a)(s-b)(s-c)(s-d)} = \tfrac{1}{2}(ac + bd)\sin\theta, \\ &= \frac{\sqrt{(ac+bd)(ad+bc)(ab+cd)}}{4R} \quad (\textit{Brahmagupta}), \\ p &= \sqrt{\frac{(ac+bd)(ab+cd)}{(ad+bc)}}, \\ R &= \frac{1}{4}\sqrt{\frac{(ac+bd)(ad+bc)(ab+cd)}{(s-a)(s-b)(s-c)(s-d)}}, \\ \sin\theta &= \frac{2\,\text{area}}{ac+bd}, \\ pq &= ac + bd \quad (\textit{Ptolemy}). \end{aligned}$$

A quadrilateral is *circumscribable* if it has an inscribed circle (that is, a circle tangent to all four sides). Its area is rs, where r is the radius of the inscribed circle and s is as above. A quadrilateral is circumscribable if and only if $a+c=b+d$.

For a quadrilateral that is both cyclic and circumscribable, we have the following additional equalities, where m is the distance between the centers of the inscribed and circumscribed circles:

$$a+c=b+d, \qquad \text{area} = \sqrt{abcd} = rs,$$

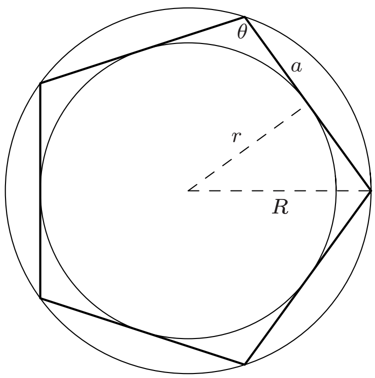
$$R = \frac{1}{4}\sqrt{\frac{(ac+bd)(ad+bc)(ab+cd)}{abcd}}, \qquad \frac{1}{r^2} = \frac{1}{(R-m)^2} + \frac{1}{(R+m)^2}.$$

4.7.3 REGULAR POLYGONS

A polygon is *regular* if all its sides are equal and all its angles are equal. Either condition implies the other in the case of a triangle, but not in general. (A rhombus has equal sides but not necessarily equal angles, and a rectangle has equal angles but not necessarily equal sides.)

For a k-sided regular polygon of side a, let θ be the angle at any vertex, and r and R the radii of the inscribed and circumscribed circles (r is called the *apothem*). As usual, let $s = \frac{1}{2}ka$ be the half-perimeter. Then

$$\begin{aligned}
\theta &= \left(\frac{k-2}{k}\right)180^\circ, \\
a &= 2r\tan\frac{180^\circ}{k} = 2R\sin\frac{180^\circ}{k}, \\
\text{area} &= \tfrac{1}{4}ka^2\cot\frac{180^\circ}{k} = kr^2\tan\frac{180^\circ}{k} \\
&= \frac{s^2}{k}\cot\frac{360^\circ}{k} = rs, \\
r &= \tfrac{1}{2}a\cot\frac{180^\circ}{k}, \\
R &= \tfrac{1}{2}a\csc\frac{180^\circ}{k}.
\end{aligned}$$



Name	k	area	r	R
Equilateral triangle	3	$0.43301\,a^2$	$0.28868\,a$	$0.57735\,a$
Square	4	a^2	$0.50000\,a$	$0.70711\,a$
Regular pentagon	5	$1.72048\,a^2$	$0.68819\,a$	$0.85065\,a$
Regular hexagon	6	$2.59808\,a^2$	$0.86603\,a$	a
Regular heptagon	7	$3.63391\,a^2$	$1.03826\,a$	$1.15238\,a$
Regular octagon	8	$4.82843\,a^2$	$1.20711\,a$	$1.30656\,a$
Regular nonagon	9	$6.18182\,a^2$	$1.37374\,a$	$1.46190\,a$
Regular decagon	10	$7.69421\,a^2$	$1.53884\,a$	$1.61803\,a$
Regular undecagon	11	$9.36564\,a^2$	$1.70284\,a$	$1.77473\,a$
Regular dodecagon	12	$11.19625\,a^2$	$1.86603\,a$	$1.93185\,a$

If a_k denotes the side of a k-sided regular polygon inscribed in a circle of radius R

$$a_{2k} = \sqrt{2R^2 - R\sqrt{4R^2 - a_k^2}}. \tag{4.7.11}$$

If A_k denotes the side of a k-sided regular polygon circumscribed about the same circle,

$$A_{2k} = \frac{2RA_k}{2R + \sqrt{4R^2 + A_k^2}}. \tag{4.7.12}$$

In particular,

$$A_{2k} = \frac{a_k A_k}{a_k + A_k}, \quad a_{2k} = \sqrt{\frac{a_k A_{2k}}{2}}. \tag{4.7.13}$$

The areas s_k, s_{2k}, S_k and S_{2k} of the same polygons satisfy

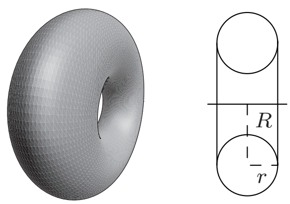
$$s_{2k} = \sqrt{s_k S_k}, \quad S_{2k} = \frac{2s_{2k}S_k}{s_{2k} + S_k}. \tag{4.7.14}$$

4.8 SURFACES OF REVOLUTION: THE TORUS

A *surface of revolution* is formed by the rotation of a planar curve C about an axis in the plane of the curve and not cutting the curve. The *Pappus–Guldinus theorem* says that:

1. The *area of the surface of revolution* on a curve C is equal to the product of the length of C and the length of the path traced by the centroid of C (which is 2π times the distance from this centroid to the axis of revolution).
2. The *volume bounded by the surface of revolution* on a simple closed curve C is equal to the product of the area bounded by C and the length of the path traced by the centroid of the area bounded by C.

When C is a circle, the surface obtained is a *circular torus* or *torus of revolution*. Let r be the radius of the revolving circle and let R be the distance from its center to the axis of rotation. The *area* of the torus is $4\pi^2 Rr$, and its *volume* is $2\pi^2 Rr^2$.



4.9 QUADRICS

A surface defined by an algebraic equation of degree two is called a *quadric*. Spheres, circular cylinders, and circular cones are quadrics. By means of a rigid motion, any quadric can be transformed into a quadric having one of the following equations (where $a, b, c \neq 0$):

1.	Real ellipsoid:	$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$
2.	Imaginary ellipsoid:	$x^2/a^2 + y^2/b^2 + z^2/c^2 = -1$
3.	Hyperboloid of one sheet:	$x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$
4.	Hyperboloid of two sheets:	$x^2/a^2 + y^2/b^2 - z^2/c^2 = -1$
5.	Real quadric cone:	$x^2/a^2 + y^2/b^2 - z^2/c^2 = 0$
6.	Imaginary quadric cone:	$x^2/a^2 + y^2/b^2 + z^2/c^2 = 0$
7.	Elliptic paraboloid:	$x^2/a^2 + y^2/b^2 + 2z = 0$
8.	Hyperbolic paraboloid:	$x^2/a^2 - y^2/b^2 + 2z = 0$
9.	Real elliptic cylinder:	$x^2/a^2 + y^2/b^2 = 1$
10.	Imaginary elliptic cylinder:	$x^2/a^2 + y^2/b^2 = -1$
11.	Hyperbolic cylinder:	$x^2/a^2 - y^2/b^2 = 1$
12.	Real intersecting planes:	$x^2/a^2 - y^2/b^2 = 0$
13.	Imaginary intersecting planes:	$x^2/a^2 + y^2/b^2 = 0$
14.	Parabolic cylinder:	$x^2 + 2y = 0$
15.	Real parallel planes:	$x^2 = 1$
16.	Imaginary parallel planes:	$x^2 = -1$
17.	Coincident planes:	$x^2 = 0$

Surfaces with Equations 9–17 are cylinders over the plane curves of the same equation (Section 4.20). Equations 2, 6, 10, and 16 have no real solutions, so that they do not describe surfaces in real three-dimensional space. A surface with Equation 5 can be regarded as a cone (Section 4.21) over a conic C (any ellipse, parabola or hyperbola can be taken as the directrix; there is a two-parameter family of essentially distinct cones over it, determined by the position of the vertex with respect to C). The surfaces with Equations 1, 3, 4, 7, and 8 are shown in Figure 4.14.

The surfaces with Equations 1–6 are *central quadrics*; in the form given, the center is at the origin. The quantities a, b, c are the *semi-axes*.

The *volume of the ellipsoid* with semi-axes a, b, c is $\frac{4}{3}\pi abc$. When two of the semi-axes are the same, we can also write the *area of the ellipsoid* in closed-form. Suppose $b = c$, so the ellipsoid $x^2/a^2 + (y^2 + z^2)/b^2 = 1$ is the surface of revolution obtained by rotating the ellipse $x^2/a^2 + y^2/b^2 = 1$ around the x-axis. Its area is

$$2\pi b^2 + \frac{2\pi a^2 b}{\sqrt{a^2 - b^2}} \sin^{-1} \frac{\sqrt{a^2 - b^2}}{a} = 2\pi b^2 + \frac{\pi a^2 b}{\sqrt{b^2 - a^2}} \log \frac{b + \sqrt{b^2 - a^2}}{b - \sqrt{b^2 - a^2}}. \quad (4.9.1)$$

The two quantities are equal, but only one avoids complex numbers, depending on whether $a > b$ or $a < b$. When $a > b$, we have a *prolate spheroid*, that is, an ellipse rotated around its major axis; when $a < b$ we have an *oblate spheroid*, which is an ellipse rotated around its minor axis.

Given a general quadratic equation in three variables,

$$ax^2 + by^2 + cz^2 + 2fyz + 2gzx + 2hxy + 2px + 2qy + 2rz + d = 0, \quad (4.9.2)$$

one can determine the type of conic by consulting the table:

ρ_3	ρ_4	Δ	k signs	K signs	Type of quadric
3	4	< 0			Real ellipsoid
3	4	> 0	same		Imaginary ellipsoid
3	4	> 0	opposite		Hyperboloid of one sheet
3	4	< 0	opposite		Hyperboloid of two sheets
3	3		opposite		Real quadric cone
3	3		same		Imaginary quadric cone
2	4	< 0	same		Elliptic paraboloid
2	4	> 0	opposite		Hyperbolic paraboloid
2	3		same	opposite	Real elliptic cylinder
2	3		same	same	Imaginary elliptic cylinder
2	3		opposite		Hyperbolic cylinder
2	2		opposite		Real intersecting planes
2	2		same		Imaginary intersecting planes
1	3				Parabolic cylinder
1	2			opposite	Real parallel planes
1	2			same	Imaginary parallel planes
1	1				Coincident planes

The columns have the following meaning. Let

$$e = \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \quad \text{and} \quad E = \begin{bmatrix} a & h & g & p \\ h & b & f & q \\ g & f & c & r \\ p & q & r & d \end{bmatrix}. \tag{4.9.3}$$

Let ρ_3 and ρ_4 be the ranks of e and E, and let Δ be the determinant of E. The column "k signs" refers to the non-zero eigenvalues of e, that is, the roots of

$$\begin{vmatrix} a-x & h & g \\ h & b-x & f \\ g & f & c-x \end{vmatrix} = 0; \tag{4.9.4}$$

if all non-zero eigenvalues have the same sign, choose "same," otherwise "opposite." Similarly, "K signs" refers to the sign of the non-zero eigenvalues of E.

4.9.1 SPHERES

The set of points in space whose distance to a fixed point (the *center*) is a fixed positive number (the *radius*) is a *sphere*. A circle of radius r and center (x_0, y_0, z_0) is defined by the equation

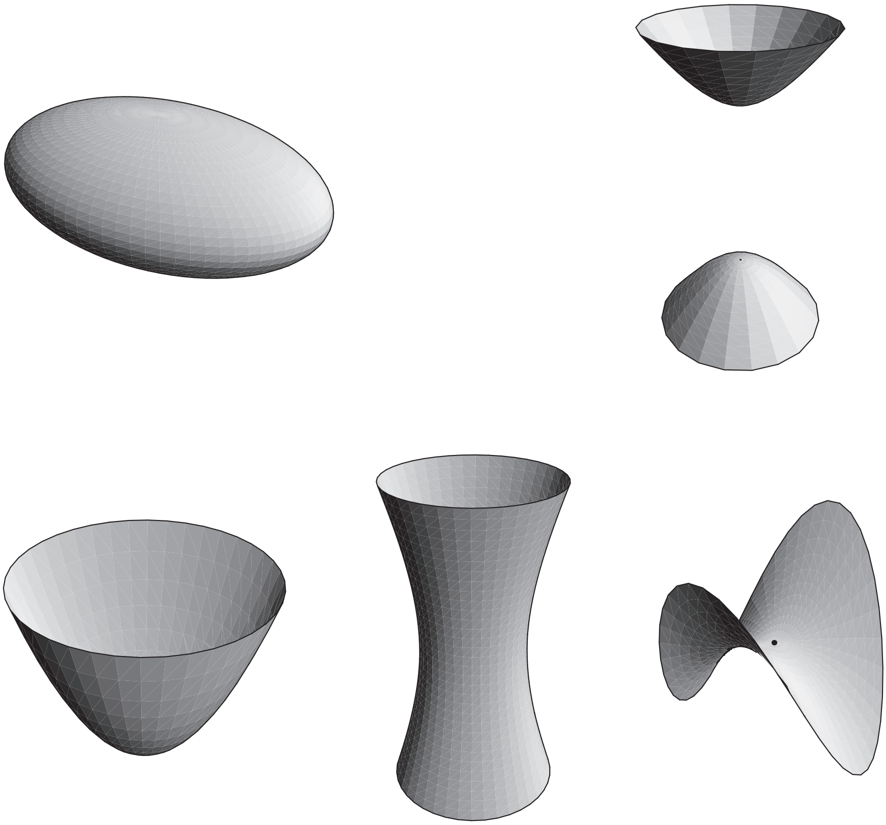
$$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = r^2, \tag{4.9.5}$$

or

$$x^2 + y^2 + z^2 - 2xx_0 - 2yy_0 - 2zz_0 + x_0^2 + y_0^2 + z_0^2 - r^2 = 0. \tag{4.9.6}$$

FIGURE 4.14

The five non-degenerate real quadrics. Top left: ellipsoid. Top right: hyperboloid of two sheets (one facing up and one facing down). Bottom left: elliptic paraboloid. Bottom middle: hyperboloid of one sheet. Bottom right: hyperbolic paraboloid.



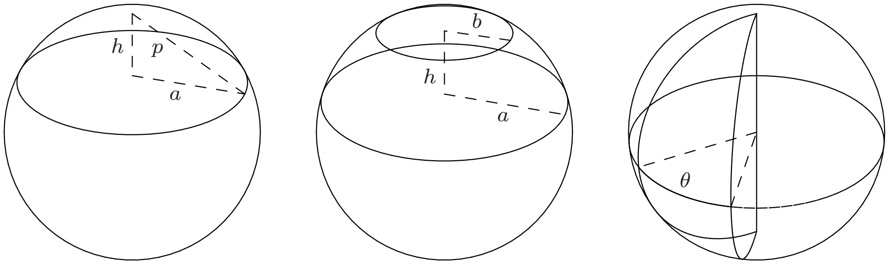
Conversely, an equation of the form

$$x^2 + y^2 + z^2 + 2dx + 2ey + 2fz + g = 0 \tag{4.9.7}$$

defines a sphere if $d^2 + e^2 + f^2 > g$; the center is $(-d, -e, -f)$ and the radius is $\sqrt{d^2 + e^2 + f^2 - g}$.

1. Four points not in the same plane determine a unique sphere. If the points have coordinates (x_1, y_1, z_1), (x_2, y_2, z_2), (x_3, y_3, z_3) and (x_4, x_4, z_4), the equation

FIGURE 4.15
Left: a spherical cap. Middle: a spherical zone (of two bases). Right: a spherical segment.



of the sphere is

$$\begin{vmatrix} x^2+y^2+z^2 & x & y & z & 1 \\ x_1^2+y_1^2+z_1^2 & x_1 & y_1 & z_1 & 1 \\ x_2^2+y_2^2+z_2^2 & x_2 & y_2 & z_2 & 1 \\ x_3^2+y_3^2+z_3^2 & x_3 & y_3 & z_3 & 1 \\ x_4^2+y_4^2+z_4^2 & x_4 & y_4 & z_4 & 1 \end{vmatrix} = 0. \qquad (4.9.8)$$

2. Given two points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$, there is a unique sphere whose diameter is P_1P_2; its equation is

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) + (z - z_1)(z - z_2) = 0. \qquad (4.9.9)$$

3. The *area* of a sphere of radius r is $4\pi r^2$, and the *volume* is $\frac{4}{3}\pi r^3$.
4. The *area of a spherical polygon* (that is, of a polygon on the sphere whose sides are arcs of great circles) is

$$S = \left(\sum_{i=1}^{n} \theta_i - (n-2)\pi\right) r^2, \qquad (4.9.10)$$

where r is the radius of the sphere, n is the number of vertices, and θ_i are the internal angles of the polygons in radians. In particular, the sum of the angles of a spherical triangle is always greater than $\pi = 180°$, and the excess is proportional to the area.

4.9.1.1 Spherical cap

Let the radius be r (Figure 4.15, left). The *area* of the curved region is $2\pi rh = \pi p^2$. The *volume* of the cap is $\frac{1}{3}\pi h^2(3r - h) = \frac{1}{6}\pi h(3a^2 + h^2)$.

4.9.1.2 Spherical zone (of two bases)

Let the radius be r (Figure 4.15, middle). The *area* of the curved region (called a *spherical zone*) is $2\pi rh$. The *volume* of the zone is $\frac{1}{6}\pi h(3a^2 + 3b^2 + h^2)$.

4.9.1.3 Spherical segment and lune

Let the radius be r (Figure 4.15, right). The *area* of the curved region (called a *spherical segment* or *lune*) is $2r^2\theta$, the angle being measured in radians. The *volume* of the segment is $\frac{2}{3}r^3\theta$.

4.9.1.4 Volume and area of spheres

If the volume of an n-dimensional sphere of radius r is $V_n(r)$ and its surface area is $S_n(r)$, then

$$\begin{aligned} V_n(r) &= \frac{2\pi r^2}{n} V_{n-2}(r) = \frac{2\pi^{n/2} r^n}{n\Gamma\left(\frac{n}{2}\right)} = \frac{\pi^{n/2} r^n}{\left(\frac{n}{2}\right)!}, \\ S_n(r) &= \frac{n}{r} V_n(r) = \frac{d}{dr}[V_n(r)]. \end{aligned} \tag{4.9.11}$$

Hence, the area of a circle is $V_2 = \pi r^2 \approx 3.1416r^2$, the volume of a 3-dimensional sphere is $V_3 = \frac{4}{3}\pi r^3 \approx 4.1888r^3$, the volume of a 4-dimensional sphere is $V_4 = \frac{1}{2}\pi^2 r^4 \approx 4.9348r^4$, the circumference of a circle is $S_2 = 2\pi r$, and the surface area of a sphere is $S_3 = 4\pi r^2$.

For large values of n,

$$V_n(r) \approx \frac{n^{-(n+1)/2}}{\sqrt{\pi}} (2\pi e)^{n/2} r^n. \tag{4.9.12}$$

4.10 SPHERICAL GEOMETRY & TRIGONOMETRY

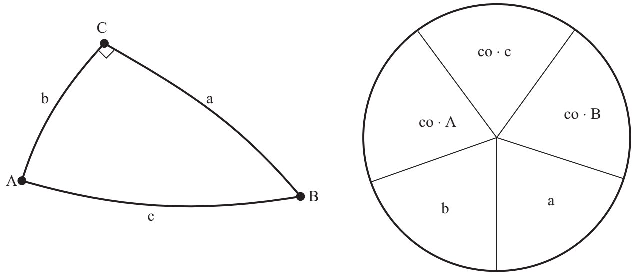
The angles in a spherical triangle do not have to add up to 180 degrees. It is possible for a spherical triangle to have 3 right angles.

4.10.1 RIGHT SPHERICAL TRIANGLES

Let a, b, and c be the sides of a right spherical triangle with opposite angles A, B, and C, respectively, where each side is measured by the angle subtended at the center of the sphere. Assume that $C = \pi/2 = 90°$ (see Figure 4.16, left). Then,

$$\sin a = \tan b \cot B = \sin A \sin c, \qquad \cos A = \tan b \cot c = \cos a \sin B,$$
$$\sin b = \tan a \cot A = \sin B \sin c, \qquad \cos B = \tan a \cot c = \cos b \sin A,$$
$$\cos c = \cos A \cot B = \cos a \cos b.$$

FIGURE 4.16
Right spherical triangle (left) and diagram for Napier's rule (right).



4.10.1.1 Napier's rules of circular parts

Arrange the five quantities a, b, co-A (this is the complement of A), co-c, co-B of a right spherical triangle with right angle at C, in cyclic order as pictured in Figure 4.16, right. If any one of these quantities is designated a *middle part*, then two of the other parts are *adjacent* to it, and the remaining two parts are *opposite* to it. The formulas above for a right spherical triangle may be recalled by the following two rules:

1. The sine of any middle part is equal to the product of the *tangents* of the two *adjacent* parts.
2. The sine of any middle part is equal to the product of the *cosines* of the two *opposite* parts.

4.10.1.2 Rules for determining quadrant

1. A leg and the angle opposite to it are always of the same quadrant.
2. If the hypotenuse is less than $90°$, the legs are of the same quadrant.
3. If the hypotenuse is greater than $90°$, the legs are not in different quadrants.

4.10.2 OBLIQUE SPHERICAL TRIANGLES

In the following:

- a, b, c represent the sides of any spherical triangle.
- A, B, C represent the corresponding opposite angles.
- a', b', c', A', B', C' are the corresponding parts of the *polar triangle*.[1]

[1]Given 3 vertices of a spherical triangle, the spherical triangle formed by connecting those 3 points with great circles is called the "polar triangle."

- $s = (a + b + c)/2$.
- $S = (A + B + C)/2$.
- Δ is the area of the spherical triangle.
- E is the *spherical excess* of the triangle.
- R is the radius of the sphere upon which the triangle lies.

$$0^\circ < a + b + c < 360^\circ, \qquad 180^\circ < A + B + C < 540^\circ,$$
$$E = A + B + C - 180^\circ, \qquad \Delta = \pi R^2 E/180.$$
$$\tan \frac{1}{4} E = \sqrt{\tan \frac{s}{2} \tan \frac{1}{2}(s-a) \tan \frac{1}{2}(s-b) \tan \frac{1}{2}(s-c)}.$$

$$A = 180^\circ - a', \qquad B = 180^\circ - b', \qquad C = 180^\circ - c',$$
$$a = 180^\circ - A', \qquad b = 180^\circ - B', \qquad c = 180^\circ - C'.$$

4.10.2.1 Spherical law of sines

$$\frac{\sin a}{\sin A} = \frac{\sin b}{\sin B} = \frac{\sin c}{\sin C}.$$

4.10.2.2 Spherical law of cosines for sides

$$\cos a = \cos b \cos c + \sin b \sin c \cos A,$$
$$\cos b = \cos c \cos a + \sin c \sin a \cos B,$$
$$\cos c = \cos a \cos b + \sin a \sin b \cos C.$$

4.10.2.3 Spherical law of cosines for angles

$$\cos A = -\cos B \cos C + \sin B \sin C \cos a,$$
$$\cos B = -\cos C \cos A + \sin C \sin A \cos b,$$
$$\cos C = -\cos A \cos B + \sin A \sin B \cos c.$$

4.10.2.4 Spherical law of tangents

$$\frac{\tan \frac{1}{2}(B-C)}{\tan \frac{1}{2}(B+C)} = \frac{\tan \frac{1}{2}(b-c)}{\tan \frac{1}{2}(b+c)}, \qquad \frac{\tan \frac{1}{2}(C-A)}{\tan \frac{1}{2}(C+A)} = \frac{\tan \frac{1}{2}(c-a)}{\tan \frac{1}{2}(c+a)},$$
$$\frac{\tan \frac{1}{2}(A-B)}{\tan \frac{1}{2}(A+B)} = \frac{\tan \frac{1}{2}(a-b)}{\tan \frac{1}{2}(a+b)}.$$

4.10.2.5 Spherical half angle formulas

Define $k^2 = \dfrac{\sin(s-a)\sin(s-b)\sin(s-c)}{\sin s}$. Then

$$\begin{aligned}\tan\left(\frac{A}{2}\right) &= \frac{k}{\sin(s-a)},\\ \tan\left(\frac{B}{2}\right) &= \frac{k}{\sin(s-b)},\\ \tan\left(\frac{C}{2}\right) &= \frac{k}{\sin(s-c)}.\end{aligned} \tag{4.10.1}$$

4.10.2.6 Spherical half side formulas

Define $K^2 = (\tan R)^2 = \dfrac{-\cos S}{\cos(S-A)\cos(S-B)\cos(S-C)}$. Then

$$\begin{aligned}\tan(a/2) &= K\cos(S-A),\\ \tan(b/2) &= K\cos(S-B),\\ \tan(c/2) &= K\cos(S-C).\end{aligned} \tag{4.10.2}$$

4.10.2.7 Gauss' formulas

$$\frac{\sin\frac{1}{2}(a-b)}{\sin\frac{1}{2}c} = \frac{\sin\frac{1}{2}(A-B)}{\cos\frac{1}{2}C}, \qquad \frac{\cos\frac{1}{2}(a-b)}{\cos\frac{1}{2}c} = \frac{\sin\frac{1}{2}(A+B)}{\cos\frac{1}{2}C},$$

$$\frac{\sin\frac{1}{2}(a+b)}{\sin\frac{1}{2}c} = \frac{\cos\frac{1}{2}(A-B)}{\sin\frac{1}{2}C}, \qquad \frac{\cos\frac{1}{2}(a+b)}{\cos\frac{1}{2}c} = \frac{\cos\frac{1}{2}(A+B)}{\sin\frac{1}{2}C}.$$

4.10.2.8 Napier's analogs

$$\frac{\sin\frac{1}{2}(A-B)}{\sin\frac{1}{2}(A+B)} = \frac{\tan\frac{1}{2}(a-b)}{\tan\frac{1}{2}c}, \qquad \frac{\sin\frac{1}{2}(a-b)}{\sin\frac{1}{2}(a+b)} = \frac{\tan\frac{1}{2}(A-B)}{\cot\frac{1}{2}C},$$

$$\frac{\cos\frac{1}{2}(A-B)}{\cos\frac{1}{2}(A+B)} = \frac{\tan\frac{1}{2}(a+b)}{\tan\frac{1}{2}c}, \qquad \frac{\cos\frac{1}{2}(a-b)}{\cos\frac{1}{2}(a+b)} = \frac{\tan\frac{1}{2}(A+B)}{\cot\frac{1}{2}C}.$$

4.10.2.9 Rules for determining quadrant

1. If $A > B > C$, then $a > b > c$.
2. A side (angle) which differs by more than 90° from another side (angle) is in the same quadrant as its opposite angle (side).
3. Half the sum of any two sides and half the sum of the opposite angles are in the same quadrant.

4.10.2.10 Summary of solution of oblique spherical triangles

Given	Solution	Check
Three sides	Half-angle formulas	Law of sines
Three angles	Half-side formulas	Law of sines
Two sides and included angle	Napier's analogies (to find sum and difference of unknown angles); then law of sines (to find remaining side).	Gauss' formulas
Two angles and included side	Napier's analogies (to find sum and difference of unknown sides); then law of sines (to find remaining angle).	Gauss' formulas
Two sides and an opposite angle	Law of sines (to find an angle); then Napier's analogies (to find remaining angle and side). Note the number of solutions.	Gauss' formulas
Two angles and an opposite side	Law of sines (to find a side); then Napier's analogies (to find remaining side and angle). Note the number of solutions.	Gauss' formulas

4.10.2.11 Finding the distance between two points on the earth

To find the distance between two points on the surface of a spherical earth, let point P_1 have a (latitude, longitude) of (ϕ_1, θ_1) and point P_2 have a (latitude, longitude) of (ϕ_2, θ_2). Two different computational methods are as follows:

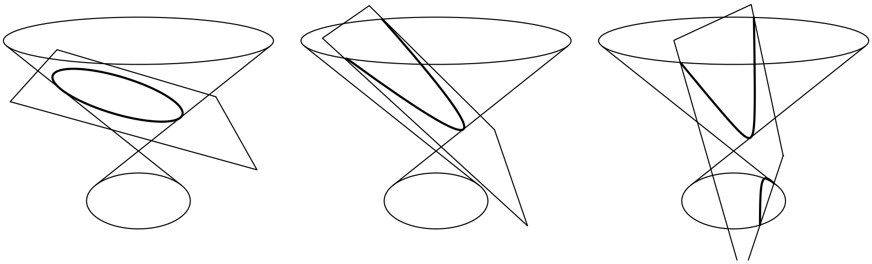
1. Let A be the North pole and let B and C be the points P_1 and P_2. Then the spherical law of cosines gives the desired distance, a, on a sphere of unit radius:
$$\cos(a) = \cos(b)\cos(c) + \sin(b)\sin(c)\cos(A)$$
where the angle A is the difference in longitudes, and b and c are the angles of the points from the pole (i.e., $90°-$ latitude). Scale by $R_\oplus$ (the radius of the earth) to get the final answer.
2. In (x, y, z) space (with $+z$ being the North pole) points P_1 and P_2 are represented as vectors from the center of the earth
$$\begin{aligned} \mathbf{v}_1 &= \begin{bmatrix} R_\oplus\cos(\phi_1)\cos(\theta_1) & R_\oplus\cos(\phi_1)\sin(\theta_1) & R_\oplus\sin(\phi_1) \end{bmatrix}, \\ \mathbf{v}_2 &= \begin{bmatrix} R_\oplus\cos(\phi_2)\cos(\theta_2) & R_\oplus\cos(\phi_2)\sin(\theta_2) & R_\oplus\sin(\phi_2) \end{bmatrix}. \end{aligned} \tag{4.10.3}$$
The angle between these vectors, α, is given by
$$\cos\alpha = \frac{\mathbf{v}_1\cdot\mathbf{v}_2}{|\mathbf{v}_1|\,|\mathbf{v}_2|} = \frac{\mathbf{v}_1\cdot\mathbf{v}_2}{R_\oplus^2} = \cos(\phi_1)\cos(\phi_2)\cos(\theta_1-\theta_2)+\sin(\phi_1)\sin(\phi_2)$$
or $\cos\alpha = 2\tan^{-1}\sqrt{b/(1-b)}$ where $b = \cos(\phi_1)\cos(\phi_2)\sin^2\left(\frac{\theta_1-\theta_2}{2}\right) + \sin^2\left(\frac{\phi_1-\phi_2}{2}\right)$. The great circle distance between P_1 and P_2 is then $R_\oplus\alpha$.

EXAMPLE The angle between $P_{\text{New York}}$ with $(\phi_1 = 40.78°, \theta_1 = 73.97°)$ and P_{Beijing} with $(\phi_2 = 39.93°, \theta_2 = 243.58°)$ is $\alpha = 98.8°$. Using $R_\oplus = 6367$ the great circle distance between New York and Beijing is about 11,000 km.

4.11 CONICS

A *conic* (or *conic section*) is a plane curve that can be obtained by intersecting a right circular cone (page 265) with a plane that does not go through the vertex of the cone. There are three possibilities, depending on the relative positions of the cone and the plane (Figure 4.17). If no line of the cone is parallel to the plane, then the intersection is a closed curve, called an *ellipse*. If one line of the cone is parallel to the plane, the intersection is an open curve whose two ends are asymptotically parallel; this is called a *parabola*. Finally, there may be two lines in the cone parallel to the plane; the curve in this case has two open segments, and is called a *hyperbola*.

FIGURE 4.17
A section of a cone by a plane can yield an ellipse (left), a parabola (middle) or a hyperbola (right).



4.11.1 ALTERNATIVE CHARACTERIZATION

Assume given a point F in the plane, a line d not going through F, and a positive real number e. The set of points P such that the distance PF is e times the distance from P to d (measured along a perpendicular) is a conic. We call F the *focus*, d the *directrix*, and e the *eccentricity* of the conic. If $e < 1$ we have an ellipse, if $e = 1$ a parabola, and if $e > 1$ a hyperbola (Figure 4.18). This construction gives all conics except the circle, which is a particular case of the ellipse according to the earlier definition (we can recover it by taking the limit $e \to 0$).

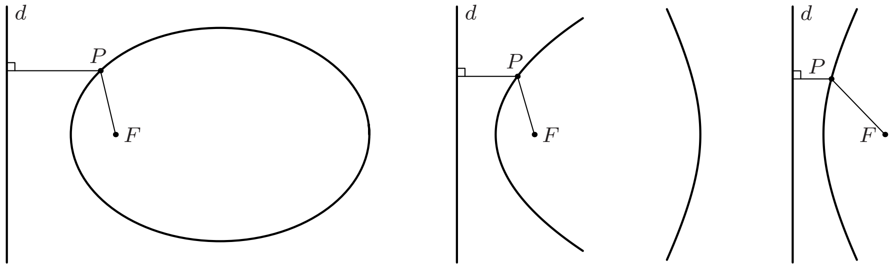
For any conic, a line perpendicular to d and passing through F is an axis of symmetry. The ellipse and the hyperbola have an additional axis of symmetry, perpendicular to the first, so that there is an alternate focus and directrix, F' and d', obtained as the reflection of F and d with respect to this axis. (By contrast, the focus and directrix are uniquely defined for a parabola.)

The simplest analytic form for the ellipse and hyperbola is obtained when the two symmetry axes coincide with the coordinate axes. The ellipse in Figure 4.19 has equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \qquad (4.11.1)$$

FIGURE 4.18

Definition of conics by means of the ratio (eccentricity) between the distance to a point and the distance to a line. On the left, $e = .7$; in the middle, $e = 1$; on the right, $e = 2$.



with $b < a$. The x-axis is the *major axis*, and the y-axis is the *minor axis*. These names are also applied to the segments, determined on the axes by the ellipse, and to the lengths of these segments: $2a$ for the major axis and $2b$ for the minor. The *vertices* are the intersections of the major axis with the ellipse and have coordinates $(a, 0)$ and $(-a, 0)$. The distance from the center to either *focus* is $\sqrt{a^2 - b^2}$, and the sum of the distances from a point in the ellipse to the foci is $2a$. The *latera recta* (in the singular, *latus rectum*) are the chords perpendicular to the major axis and going through the foci; their length is $2b^2/a$. The *eccentricity* is $\sqrt{a^2 - b^2}/a$. All ellipses of the same eccentricity are similar; in other words, the shape of an ellipse depends only on the ratio b/a. The distance from the center to either *directrix* is $a^2/\sqrt{a^2 - b^2}$.

The hyperbola in Figure 4.20 has equation

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1. \tag{4.11.2}$$

FIGURE 4.19

Ellipse with major semiaxis a and minor semiaxis b. Here $b/a = 0.6$.

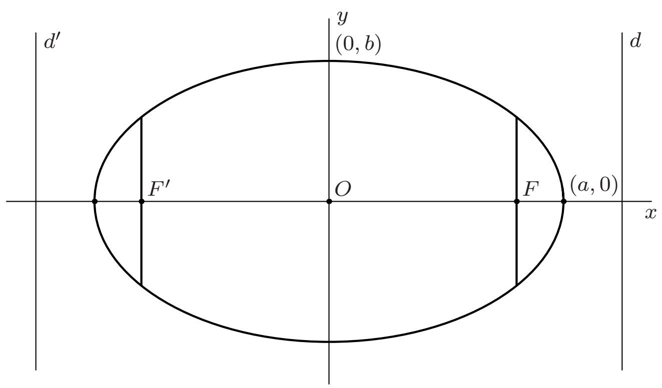
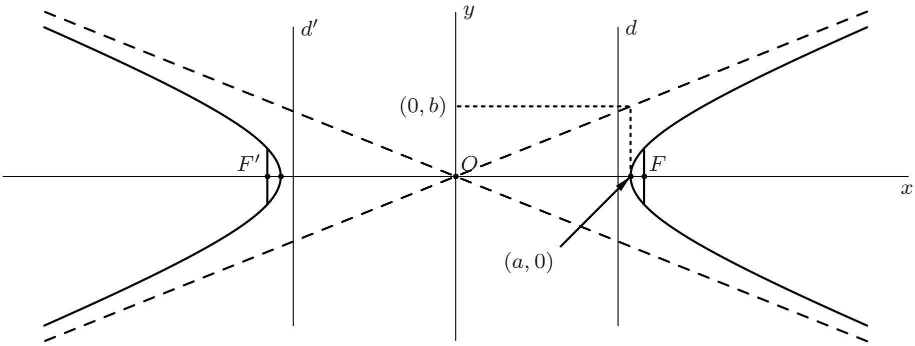


FIGURE 4.20

Hyperbola with transverse semiaxis a and conjugate semiaxis b. Here $b/a = 0.4$.



The x-axis is the *transverse axis*, and the y-axis is the *conjugate axis*. The *vertices* are the intersections of the transverse axis with the hyperbola and have coordinates $(a, 0)$ and $(-a, 0)$. The segment thus determined, or its length $2a$, is also called the transverse axis, while the length $2b$ is also called the conjugate axis. The distance from the center to either *focus* is $\sqrt{a^2 + b^2}$, and the difference between the distances from a point in the hyperbola to the foci is $2a$. The *latera recta* are the chords perpendicular to the transverse axis and going through the foci; their length is $2b^2/a$. The *eccentricity* is $\sqrt{a^2 + b^2}/a$. The distance from the center to either *directrix* is $a^2/\sqrt{a^2 + b^2}$. The legs of the hyperbola approach the *asymptotes*, lines of slope $\pm b/a$ that cross at the center.

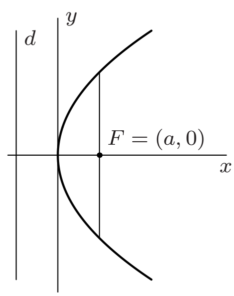
All hyperbolas of the same eccentricity are similar; in other words, the shape of a hyperbola depends only on the ratio b/a. Unlike the case of the ellipse (where the major axis, containing the foci, is always longer than the minor axis), the two axes of a hyperbola can have arbitrary lengths. When they have the same length, so that $a = b$, the asymptotes are perpendicular, and $e = \sqrt{2}$, the hyperbola is called *rectangular*.

The simplest analytic form for the parabola is obtained when the axis of symmetry coincides with one coordinate axis, and the *vertex* (the intersection of the axis with the curve) is at the origin.

The equation of the parabola on the right is

$$y^2 = 4ax, \qquad (4.11.3)$$

where a is the distance from the vertex to the focus, or, which is the same, from the vertex to the directrix. The *latus rectum* is the chord perpendicular to the axis and going through the focus; its length is $4a$. All parabolas are similar: they can be made identical by scaling, translation, and rotation.



4.11.2 THE GENERAL QUADRATIC EQUATION

The analytic equation for a conic in arbitrary position is the following:

$$Ax^2 + By^2 + Cxy + Dx + Ey + F = 0, \tag{4.11.4}$$

where at least one of A, B, C is nonzero. To reduce this to one of the forms given previously, perform the following steps (note that the decisions are based on the most recent values of the coefficients, taken after all the transformations so far):

1. If $C \neq 0$, simultaneously perform the substitutions $x \mapsto qx + y$ and $y \mapsto qy - x$, where
$$q = \sqrt{\left(\frac{B-A}{C}\right)^2 + 1} \; + \frac{B-A}{C}. \tag{4.11.5}$$
Now $C = 0$. (This step corresponds to rotating and scaling about the origin.)
2. If $B = 0$, interchange x and y. Now $B \neq 0$.
3. If $E \neq 0$, perform the substitution $y \mapsto y - \frac{1}{2}(E/B)$. (This corresponds to translating in the y direction.) Now $E = 0$.
4. If $A = 0$:
   (a) If $D \neq 0$, perform the substitution $x \mapsto x - (F/D)$ (translation in the x direction), and divide the equation by B to get Equation (4.11.3). The conic is a *parabola*.
   (b) If $D = 0$, the equation gives a *degenerate conic*. If $F = 0$, we have the line $y = 0$ with multiplicity two. If $F < 0$, we have two parallel lines $y = \pm\sqrt{F/B}$. If $F > 0$ we have two imaginary lines; the equation has no solution within the real numbers.
5. If $A \neq 0$:
   (a) If $D \neq 0$, perform the substitution $x \mapsto x - \frac{1}{2}(D/A)$. Now $D = 0$. (This corresponds to translating in the x direction.)
   (b) If $F \neq 0$, divide the equation by F to get a form with $F = 1$.
      i. If A and B have opposite signs, the conic is a *hyperbola*; to get to Equation (4.11.2), interchange x and y, if necessary, so that A is positive; then make $a = 1/\sqrt{A}$ and $b = 1/\sqrt{B}$.
      ii. If A and B are both positive, the conic is an *ellipse*; to get to Equation (4.11.1), interchange x and y, if necessary, so that $A \leq B$, then make $a = 1/\sqrt{A}$ and $b = 1/\sqrt{B}$. The *circle* is the particular case $a = b$.
      iii. If A and B are both negative, we have an *imaginary ellipse*; the equation has no solution in real numbers.
   (c) If $F = 0$, the equation again represents a *degenerate conic*: when A and B have different signs, we have a pair of lines $y = \pm\sqrt{-B/A}x$, and, when they have the same sign, we get a point (the origin).

EXAMPLE Here is an example for clarity. Suppose the original equation is

$$4x^2 + y^2 - 4xy + 3x - 4y + 1 = 0. \tag{4.11.6}$$

In step 1 we apply the substitutions $x \mapsto 2x + y$ and $y \mapsto 2y - x$. This gives $25x^2 + 10x - 5y + 1 = 0$. Next we interchange x and y (step 2) and get $25y^2 + 10y - 5x + 1 = 0$. Replacing y by $y - \frac{1}{5}$ in step 3, we get $25y^2 - 5x = 0$. Finally, in step 4a we divide the equation by 25, thus giving it the form of Equation (4.11.3) with $a = \frac{1}{20}$. We have reduced the conic to a parabola with vertex at the origin and focus at $(\frac{1}{20}, 0)$. To locate the features of the original curve, we work our way back along the chain of substitutions (recall the convention about substitutions and transformations from Section 4.3.2):

Substitution		$y \mapsto y - \frac{1}{5}$	$\substack{x \mapsto y \\ y \mapsto x}$	$\substack{x \mapsto 2x+y \\ y \mapsto 2y-x}$
Vertex	$(0, 0)$	$(0, -\frac{1}{5})$	$(-\frac{1}{5}, 0)$	$(-\frac{2}{5}, -\frac{1}{5})$
Focus	$(\frac{1}{20}, 0)$	$(\frac{1}{20}, -\frac{1}{5})$	$(-\frac{1}{5}, \frac{1}{20})$	$(-\frac{7}{20}, \frac{6}{20})$

We conclude that the original curve, Equation (4.11.6), is a parabola with vertex $(-\frac{2}{5}, -\frac{1}{5})$ and focus $(-\frac{7}{20}, \frac{6}{20})$.

An alternative analysis of Equation (4.11.4) consists in forming the quantities

$$\Delta = \begin{vmatrix} A & \frac{1}{2}C & \frac{1}{2}D \\ \frac{1}{2}C & B & \frac{1}{2}E \\ \frac{1}{2}D & \frac{1}{2}E & F \end{vmatrix} \qquad J = \begin{vmatrix} A & \frac{1}{2}C \\ \frac{1}{2}C & B \end{vmatrix}$$
$$K = \begin{vmatrix} A & \frac{1}{2}D \\ \frac{1}{2}D & F \end{vmatrix} + \begin{vmatrix} B & \frac{1}{2}E \\ \frac{1}{2}E & F \end{vmatrix} \qquad I = A + B \tag{4.11.7}$$

and finding the appropriate case in the following table, where an entry in parentheses indicates that the equation has no solution in real numbers:

Δ	J	Δ/I	K	Type of conic
$\neq 0$	< 0	any	any	Hyperbola
$\neq 0$	0	any	any	Parabola
$\neq 0$	> 0	< 0	any	Ellipse
$\neq 0$	> 0	> 0	any	(Imaginary ellipse)
0	< 0	any	any	Intersecting lines
0	> 0	any	any	Point
0	0	any	< 0	Distinct parallel lines
0	0	any	> 0	(Imaginary parallel lines)
0	0	any	0	Coincident lines

For the central conics (the ellipse, the hyperbola, intersecting lines, and the point), the center (x_0, y_0) is the solution of the system of equations

$$2Ax + Cy + D = 0,$$
$$Cx + 2By + E = 0,$$

namely

$$(x_0, y_0) = \left(\frac{2BD - CE}{C^2 - 4AB}, \frac{2AE - CD}{C^2 - 4AB} \right), \tag{4.11.8}$$

and the axes have slopes q and $-1/q$, where q is given by Equation (4.11.5). (The value $-1/q$ can be obtained from Equation (4.11.5) by simply placing a minus sign before the radical.) The length of the semiaxis with slope q is

$$\sqrt{\frac{|\Delta|}{|Jr|}}, \quad \text{where } r = \frac{1}{2}(A + B + \sqrt{(B-A)^2 + C^2}); \tag{4.11.9}$$

note that r is one of the eigenvalues of the matrix of which J is the determinant. To obtain the other semiaxis, take the other eigenvalue (change the sign of the radical in the expression of r just given).

EXAMPLE Consider the equation $3x^2 + 4xy - 2y^2 + 3x - 2y + 7 = 0$. We have

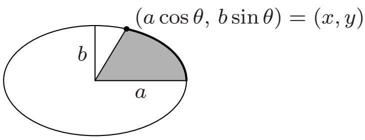
$$\Delta = \begin{vmatrix} 6 & 4 & 3 \\ 4 & -4 & -2 \\ 3 & -2 & 14 \end{vmatrix} = -596 \neq 0$$

$$J = \begin{vmatrix} 6 & 4 \\ 4 & -4 \end{vmatrix} = -40 < 0$$

We conclude that this is a hyperbola.

4.11.3 ADDITIONAL PROPERTIES OF ELLIPSES

Let C be the ellipse with equation $x^2/a^2 + y^2/b^2 = 1$, with $a > b$, and let $F, F' = (\pm\sqrt{a^2 - b^2}, 0)$ be its foci (see Figure 4.19).



1. A *parametric representation* for C is given by $(a\cos\theta,\ b\sin\theta)$. The area of the shaded sector on the right is $\frac{1}{2}ab\tan^{-1}\left(\frac{a}{b}\tan\theta\right)$. The *length* of the arc from $(a, 0)$ to the point $(a\cos\theta,\ b\sin\theta)$ is

$$a\int_0^\theta \sqrt{1 - e^2\cos^2\phi}\, d\phi = a\left(E\left(\frac{\pi}{2}, e\right) - E\left(\frac{\pi}{2} - \theta, e\right)\right),$$

where e is the eccentricity and E is an elliptic integral (see page 470). Setting $\theta = 2\pi$ results in

$$\text{area } = \pi ab, \qquad \text{perimeter } = 4a\, E(\pi/2, e). \tag{4.11.10}$$

Note the approximation: perimeter $\approx \pi\left[3(a+b) - \sqrt{(3a+b)(a+3b)}\right]$

2. Given an ellipse in the form $Ax^2 + Bxy + Cy^2 = 1$, form the matrix $D = \left[\begin{smallmatrix} A & B/2 \\ B/2 & C \end{smallmatrix}\right]$. Let the eigenvalues of D be $\{\lambda_1, \lambda_2\}$ and let $\{\mathbf{v}_1, \mathbf{v}_2\}$ be the corresponding unit eigenvectors (choose them orthogonal if $\lambda_1 = \lambda_2$). Then the major and minor semiaxes are given by $\mathbf{v}_i = \frac{\mathbf{u}_i}{\sqrt{\lambda_i}}$ and

   (a) The area of the ellipse is $\dfrac{\pi}{\sqrt{\lambda_1\lambda_2}} = \dfrac{2\pi}{\sqrt{4AC - B^2}}$.
   (b) The ellipse has the parametric representation $\mathbf{x}(t) = \cos(t)\mathbf{v}_1 + \sin(t)\mathbf{v}_2$.
   (c) The rectangle with vertices $(\pm\mathbf{v}_1, \pm\mathbf{v}_2)$ is tangent to the ellipse.

3. A *rational parametric representation* for C is given by $\left(a\,\dfrac{1-t^2}{1+t^2}, \dfrac{2bt}{1+t^2}\right)$.
4. The *polar equation* for C in the usual polar coordinate system is

$$r = \frac{ab}{\sqrt{a^2\sin^2\theta + b^2\cos^2\theta}}. \tag{4.11.11}$$

   With respect to a coordinate system with origin at a focus, the equation is

$$r = \frac{l}{1 \pm e\cos\theta}, \tag{4.11.12}$$

   where $l = b^2/a$ is half the latus rectum. (Use the $+$ sign for the focus with positive x-coordinate and the $-$ sign for the focus with negative x-coordinate.)
5. Let P be any point of C. The *sum of the distances* PF and PF' is constant and equal to $2a$.
6. Let P be any point of C. Then the rays PF and PF' make the same angle with the tangent to C at P. Thus any light ray originating at F and reflected in the ellipse will go through F'.
7. Let T be any line tangent to C. The product of the distances from F and F' to T is constant and equals b^2.
8. *Lahire's theorem*: Let D and D' be fixed lines in the plane, and consider a third moving line on which three points P, P' and P'' are marked. If we constrain P to lie in D and P' to lie in D', then P'' describes an ellipse.

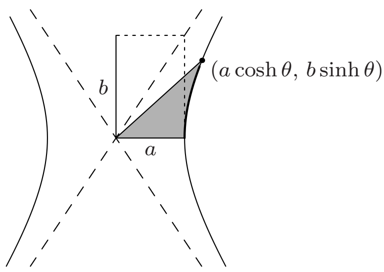
4.11.4 ADDITIONAL PROPERTIES OF HYPERBOLAS

Let C be the hyperbola with equation $x^2/a^2 - y^2/b^2 = 1$, and let

$$F, F' = (\pm\sqrt{a^2+b^2}, 0) \tag{4.11.13}$$

be its foci (see Figure 4.20). The *conjugate hyperbola* of C is the hyperbola C' with equation $-x^2/a^2 + y^2/b^2 = 1$. It has the same asymptotes as C, the same axes (transverse and conjugate axes being interchanged), and its eccentricity e' is related to that of C by $e'^{-2} + e^{-2} = 1$.

1. A *parametric representation* for C is given by $(a\sec\theta,\, b\tan\theta)$. A parametric representation, which gives one branch only, is $(a\cosh\theta,\, b\sinh\theta)$. The area of the shaded sector on the right is



$$\tfrac{1}{2}ab\theta = \tfrac{1}{2}ab\cosh^{-1}(x/a)$$
$$= \tfrac{1}{2}ab\log\frac{x+\sqrt{x^2-a^2}}{a}.$$

where $x = a\cosh\theta$. The *length* of the arc from $(a,0)$ to the point $(a\cosh\theta,\, b\sinh\theta)$ is given by the elliptic integral

$$a\int_0^\theta \sqrt{e^2\cosh^2\phi - 1}\,d\phi = -bi\,E\Big(\theta i,\, \frac{ea}{b}\Big) = a\int_1^x \sqrt{\frac{e^2\xi^2-a^2}{\xi^2-a^2}}\,d\xi,$$

where e is the eccentricity and $i=\sqrt{-1}$.

2. A *rational parametric representation* for C is given by

$$\Big(a\,\frac{1+t^2}{1-t^2},\, \frac{2bt}{1-t^2}\Big). \tag{4.11.14}$$

3. The *polar equation* for C in the usual polar coordinate system is

$$r = \frac{ab}{\sqrt{a^2\sin^2\theta - b^2\cos^2\theta}}. \tag{4.11.15}$$

With respect to a system with origin at a focus, the equation is

$$r = \frac{l}{1\pm e\cos\theta}, \tag{4.11.16}$$

where $l = b^2/a$ is half the latus rectum. (Use the $-$ sign for the focus with positive x-coordinate and the $+$ sign for the focus with negative x-coordinate.)

4. Let P be any point of C. The unsigned *difference between the distances* PF and PF' is constant and equal to $2a$.
5. Let P be any point of C. Then the rays PF and PF' make the same angle with the tangent to C at P. Thus any light ray originating at F and reflected in the hyperbola will appear to emanate from F'.
6. Let T be any line tangent to C. The product of the distances from F and F' to T is constant and equals b^2.
7. Let P be any point of C. The area of the parallelogram formed by the asymptotes and the parallels to the asymptotes going through P is constant and equals $\tfrac{1}{2}ab$.
8. Let L be any line in the plane. If L intersects C at P and P' and intersects the asymptotes at Q and Q', the distances PQ and $P'Q'$ are the same. If L is tangent to C we have $P=P'$, so that the point of tangency bisects the segment QQ'.

4.11.5 ADDITIONAL PROPERTIES OF PARABOLAS

Let C be the parabola with equation $y^2 = 4ax$, and let $F = (a, 0)$ be its focus.

1. Let $P = (x, y)$ and $P' = (x', y')$ be points on C. The area bounded by the chord PP' and the corresponding arc of the parabola is
$$\frac{|y' - y|^3}{24a}. \tag{4.11.17}$$
It equals four-thirds of the area of the triangle PQP', where Q is the point on C whose tangent is parallel to the chord PP' (formula due to *Archimedes*).
2. The *length* of the arc from $(0, 0)$ to the point (x, y) is
$$\frac{y}{4}\sqrt{4 + \frac{y^2}{a^2}} + a \sinh^{-1}\left(\frac{y}{2a}\right) = \frac{y}{4}\sqrt{4 + \frac{y^2}{a^2}} + a \log \frac{y + \sqrt{y^2 + 4a^2}}{2a}. \tag{4.11.18}$$
3. The *polar equation* for C in the usual polar coordinate system is
$$r = \frac{4a\cos\theta}{\sin^2\theta}. \tag{4.11.19}$$
With respect to a coordinate system with origin at F, the equation is
$$r = \frac{l}{1 - \cos\theta}, \tag{4.11.20}$$
where $l = 2a$ is half the latus rectum.
4. Let P be any point of C. Then the ray PF and the horizontal line through P make the same angle with the tangent to C at P. Thus light rays parallel to the axis and reflected in the parabola converge onto F (principle of the *parabolic reflector*).

4.11.6 CIRCLES

The set of points in a plane whose distance to a fixed point (the *center*) is a fixed positive number (the *radius*) is a *circle*. A circle of radius r and center (x_0, y_0) is described by the equation
$$(x - x_0)^2 + (y - y_0)^2 = r^2, \tag{4.11.21}$$
or
$$x^2 + y^2 - 2xx_0 - 2yy_0 + x_0^2 + y_0^2 - r^2 = 0. \tag{4.11.22}$$
Conversely, an equation of the form
$$x^2 + y^2 + 2dx + 2ey + f = 0 \tag{4.11.23}$$
defines a circle if $d^2 + e^2 > f$; the center is $(-d, -e)$ and the radius is $\sqrt{d^2 + e^2 - f}$.

Three points not on the same line determine a unique circle. If the points have coordinates (x_1, y_1), (x_2, y_2) and (x_3, y_3), then the equation of the circle is

$$\begin{vmatrix} x^2+y^2 & x & y & 1 \\ x_1^2+y_1^2 & x_1 & y_1 & 1 \\ x_2^2+y_2^2 & x_2 & y_2 & 1 \\ x_3^2+y_3^2 & x_3 & y_3 & 1 \end{vmatrix} = 0. \tag{4.11.24}$$

A *chord* of a circle is a line segment between two of its points (Figure 4.21). A *diameter* is a chord that goes through the center, or the length of such a chord (therefore the diameter is twice the radius). Given two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, there is a unique circle whose diameter is P_1P_2; its equation is

$$(x - x_1)(x - x_2) + (y - y_1)(y - y_2) = 0. \tag{4.11.25}$$

The *length* or *circumference* of a circle of radius r is $2\pi r$, and the *area* is πr^2. The length of the *arc of circle* subtended by an angle θ, shown as s in Figure 4.21, is $r\theta$. (All angles are measured in radians.) Other relations between the radius, the arc length, the chord, and the areas of the corresponding *sector* and *segment* are, in the notation of Figure 4.21,

$$\begin{aligned} d &= R\cos\tfrac{1}{2}\theta = \tfrac{1}{2}c\cot\tfrac{1}{2}\theta = \tfrac{1}{2}\sqrt{4R^2 - c^2}, \\ c &= 2R\sin\tfrac{1}{2}\theta = 2d\tan\tfrac{1}{2}\theta = 2\sqrt{R^2 - d^2} = \sqrt{4h(2R - h)}, \\ \theta &= \frac{s}{R} = 2\cos^{-1}\frac{d}{R} = 2\tan^{-1}\frac{c}{2d} = 2\sin^{-1}\frac{c}{2R}, \\ \text{area of sector} &= \tfrac{1}{2}Rs = \tfrac{1}{2}R^2\theta, \\ \text{area of segment} &= \tfrac{1}{2}R^2(\theta - \sin\theta) = \tfrac{1}{2}(Rs - cd) = R^2\cos^{-1}\frac{d}{R} - d\sqrt{R^2 - d^2} \\ &= R^2\cos^{-1}\frac{R-h}{R} - (R - h)\sqrt{2Rh - h^2}. \end{aligned}$$

FIGURE 4.21

The arc of a circle subtended by the angle θ is s; the chord is c; the sector is the whole slice of the pie; the segment is the cap bounded by the arc and the chord (that is, the slice minus the triangle).

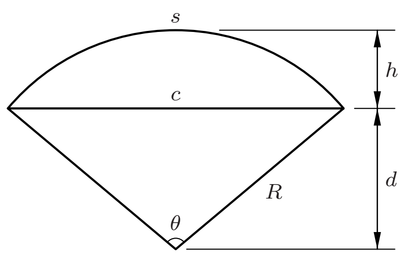
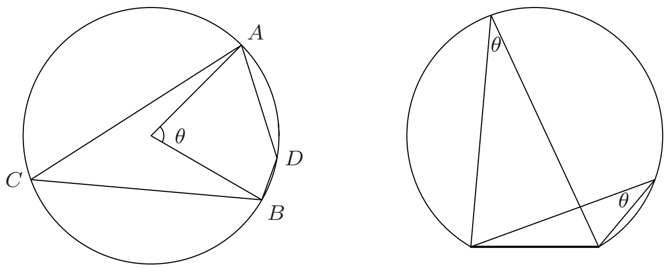


FIGURE 4.22

Left: the angle ACB equals $\frac{1}{2}\theta$ for any C in the long arc AB; ADB equals $180^\circ - \frac{1}{2}\theta$ for any D in the short arc AB. Right: the locus of points, from which the segment AB subtends a fixed angle θ, is an arc of the circle.



Other properties of circles:

1. If the central angle AOB equals θ, the angle ACB, where C is any point on the circle, equals $\frac{1}{2}\theta$ or $180^\circ - \frac{1}{2}\theta$ (Figure 4.22, left). Conversely, given a segment AB, the set of points that "see" AB under a fixed angle is an arc of a circle (Figure 4.22, right). In particular, the set of points that see AB under a right angle is a circle with diameter AB.
2. Let P_1, P_2, P_3, P_4 be points in the plane, and let d_{ij}, for $1 \leq i, j \leq 4$, be the distance between P_i and P_j. A necessary and sufficient condition for all of the points to lie on the same circle (or line) is that one of the following equalities be satisfied:
$$d_{12}d_{34} \pm d_{13}d_{24} \pm d_{14}d_{23} = 0. \tag{4.11.26}$$
This is equivalent to Ptolemy's formula for cyclic quadrilaterals (page 217).
3. In *oblique coordinates* with angle ω, a circle of center (x_0, y_0) and radius r is described by the equation
$$(x - x_0)^2 + (y - y_0)^2 + 2(x - x_0)(y - y_0)\cos\omega = r^2. \tag{4.11.27}$$
4. In *polar coordinates*, the equation for a circle centered at the pole and having radius a is $r = a$. The equation for a circle of radius a passing through the pole and with center at the point $(r, \theta) = (a, \theta_0)$ is $r = 2a\cos(\theta - \theta_0)$. The equation for a circle of radius a and with center at the point $(r, \theta) = (r_0, \theta_0)$
$$r^2 - 2r_0 r\cos(\theta - \theta_0) + r_0^2 - a^2 = 0. \tag{4.11.28}$$
5. If a line intersects a circle of center O at points A and B, the segments OA and OB make equal angles with the line. In particular, a tangent line is perpendicular to the radius that goes through the point of tangency.
6. Fix a circle and a point P in the plane, and consider a line through P that intersects the circle at A and B (with $A = B$ for a tangent). Then the product of the distances $PA \cdot PB$ is the same for all such lines. It is called the *power* of P with respect to the circle.

4.12 SPECIAL PLANE CURVES

4.12.1 ALGEBRAIC CURVES

Curves that can be given in implicit form as $f(x, y) = 0$, where f is a polynomial, are called *algebraic*. The degree of f is called the degree or *order* of the curve. Thus, conics (page 229) are algebraic curves of degree two. Curves of degree three already have a great variety of shapes, and only a few common ones will be given here.

The simplest case is the curve which is a graph of a polynomial of degree three: $y = ax^3 + bx^2 + cx + d$, with $a \neq 0$. This curve is a (general) *cubic parabola* (Figure 4.23), symmetric with respect to the point B where $x = -b/3a$.

The equation of a *semi-cubic parabola* (Figure 4.24, left) is $y^2 = kx^3$; by proportional scaling one can take $k = 1$. This curve should not be confused with the *cissoid of Diocles* (Figure 4.24, middle), whose equation is $(a - x)y^2 = x^3$ with $a \neq 0$.

FIGURE 4.23
The general cubic parabola for $a > 0$. For $a < 0$, reflect in a horizontal line.

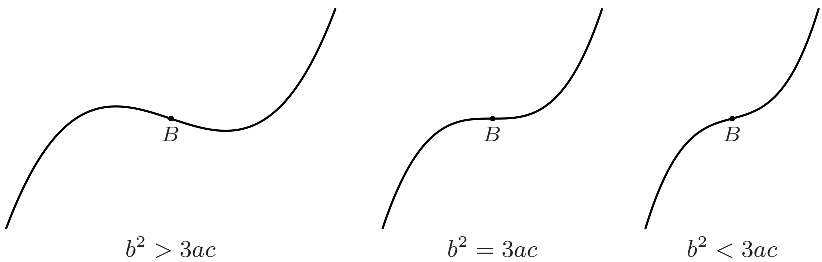
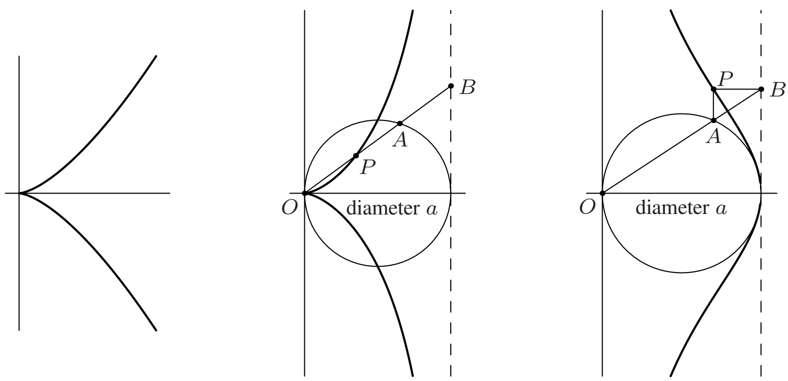


FIGURE 4.24
The semi-cubic parabola, the cissoid of Diocles, and the witch of Agnesi.

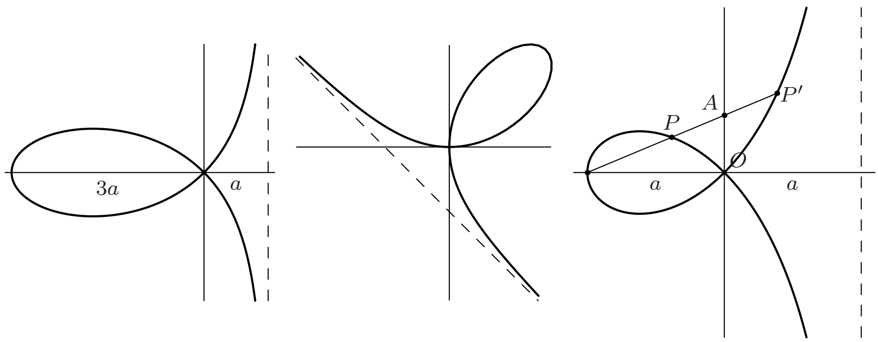


The latter is asymptotic to the line $x = a$, whereas the semi-cubic parabola has no asymptotes. The cissoid's points are characterized by the equality $OP = AB$ in Figure 4.24, middle. One can take $a = 1$ by proportional scaling.

More generally, any curve of degree three with equation $(x - x_0)y^2 = f(x)$, where f is a polynomial, is symmetric with respect to the x-axis and asymptotic to the line $x = x_0$. In addition to the cissoid, the following particular cases are important:

1. The *witch of Agnesi* has equation $xy^2 = a^2(a - x)$, with $a \neq 0$, and is characterized by the geometric construction shown in Figure 4.24, right. This construction provides the parametric representation $x = a\cos^2\theta$, $y = a\tan\theta$. Once more, proportional scaling reduces to the case $a = 1$.
2. The *folium of Descartes* (Figure 4.25, left) is described by equation $(x - a)y^2 = -x^2(\frac{1}{3}x + a)$, with $a \neq 0$ (reducible to $a = 1$ by proportional scaling). By rotating 135° (right) we get the alternative and more familiar equation $x^3 + y^3 = cxy$, where $c = \frac{1}{3}\sqrt{2}a$. The folium of Descartes is a *rational curve*, that is, it is parametrically represented by rational functions. In the tilted position, the equation is $x = ct/(1 + t^3)$, $y = ct^2/(1 + t^3)$ (so that $t = y/x$).
3. The *strophoid's* equation is $(x - a)y^2 = -x^2(x + a)$, with $a \neq 0$ (reducible to $a = 1$ by proportional scaling). It satisfies the property $AP = AP' = OA$ in Figure 4.25, right; this means that POP' is a right angle. The strophoid's polar representation is $r = -a\cos 2\theta \sec\theta$, and the rational parametric representation is $x = a(t^2 - 1)/(t^2 + 1)$, $y = at(t^2 - 1)/(t^2 + 1)$ (so that $t = y/x$).

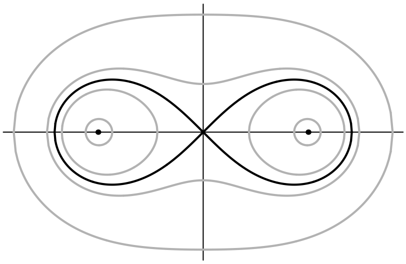
FIGURE 4.25
The folium of Descartes in two positions, and the strophoid.



Among the important curves of degree four are the following:

1. A *Cassini's oval* is characterized by the following condition: Given two *foci* F and F', a distance $2a$ apart, a point P belongs to the curve if the product of the distances PF and PF' is a constant k^2. If the foci are on the x-axis and equidistant from the origin, the curve's equation is $(x^2 + y^2 + a^2)^2 - 4a^2x^2 = k^4$. Changes in a correspond to rescaling, while the value of k/a controls

FIGURE 4.26
Cassini's ovals for $k = 0.5a$, $0.9a$, a, $1.1a$ and $1.5a$ (from the inside to the outside). The foci (dots) are at $x = a$ and $x = -a$. The black curve, $k = a$, is also called Bernoulli's lemniscate.



the shape: the curve has one smooth segment and one with a self-intersection, or two segments depending on whether k is greater than, equal to, or smaller than a (Figure 4.26). The case $k = a$ is also known as the *lemniscate* (of Jakob Bernoulli); the equation reduces to $(x^2 + y^2)^2 = a^2(x^2 - y^2)$, and upon a $45°$ rotation to $(x^2 + y^2)^2 = 2a^2xy$. Each Cassini's oval is the section of a torus of revolution by a plane parallel to the axis of revolution.

2. A *conchoid of Nichomedes* is the set of points such that the signed distance AP in Figure 4.27, left, equals a fixed real number k (the line L and the origin O being fixed). If L is the line $x = a$, the conchoid's polar equation is $r = a\sec\theta + k$. Once more, a is a scaling parameter, and the value of k/a controls the shape: when $k > -a$ the curve is smooth, when $k = -a$ there is a cusp, and when $k < -a$ there is a self-intersection. The curves for k and $-k$ can also be considered two leaves of the same conchoid, with Cartesian equation $(x - a)^2(x^2 + y^2) = k^2x^2$.

3. A *limaçon of Pascal* is the set of points such that the distance AP in Figure 4.28, left, equals a fixed positive number k measured on either side (the circle C and the origin O being fixed). If C has diameter a and center at $(0, \frac{1}{2}a)$, the limaçon's polar equation is $r = a\cos\theta + k$, and its Cartesian equation is

$$(x^2 + y^2 - ax)^2 = k^2(x^2 + y^2). \tag{4.12.1}$$

The value of k/a controls the shape, and there are two particularly interesting cases. For $k = a$, we get a *cardioid* (see also page 245). For $a = \frac{1}{2}k$, we get a curve that can be used to *trisect* an arbitrary angle α. If we draw a line L through the center of the circle C making an angle α with the positive x-axis, and if we call P the intersection of L with the limaçon $a = \frac{1}{2}k$, the line from O to P makes an angle with L equal to $\frac{1}{3}\alpha$.

Hypocycloids and epicycloids with rational ratios (see next section) are also algebraic curves, generally of higher degree.

FIGURE 4.27
Defining property of the conchoid of Nichomedes (left), and curves for $k = \pm 0.5a$, $k = \pm a$, *and* $k = \pm 1.5a$ *(right).*

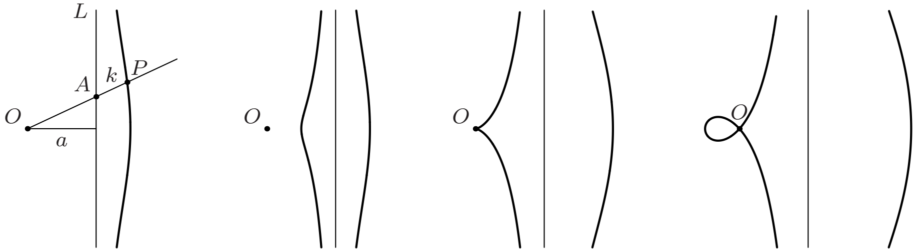


FIGURE 4.28
Defining property of the limaçon of Pascal (left), and curves for $k = 1.5a$, $k = a$, *and* $k = 0.5a$ *(right). The middle curve is the cardioid; the one on the right a trisectrix.*

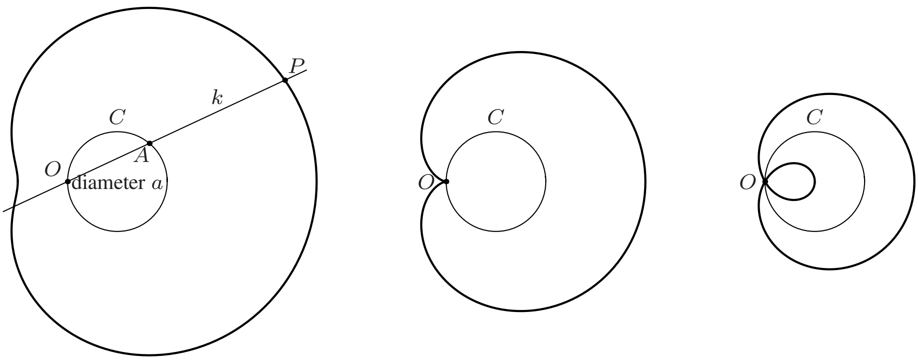
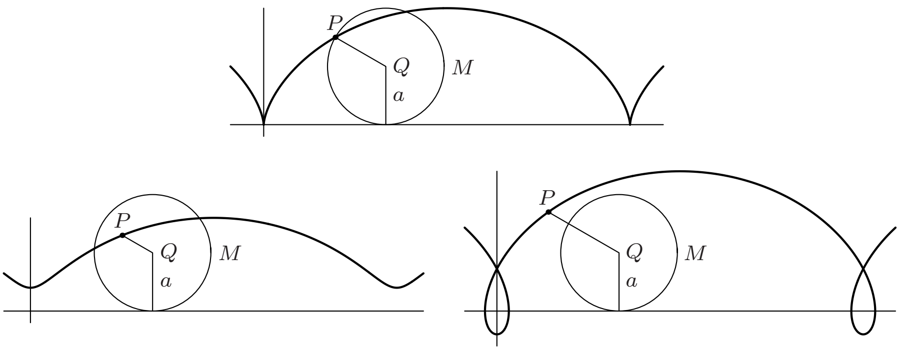


FIGURE 4.29
Cycloid (top) and trochoids with $k = 0.5a$ *and* $k = 1.6a$, *where* k *is the distance* PQ *from the center of the rolling circle to the pole.*



4.12.2 ROULETTES (SPIROGRAPH CURVES)

Suppose given a fixed curve C and a moving curve M, which rolls on C without slipping. The curve drawn by a point P kept fixed with respect to M is called a *roulette*, of which P is the *pole*.

The most important examples of roulettes arise when M is a circle and C is a straight line or a circle, but an interesting additional example is provided by the *catenary* $y = a\cosh(x/a)$, which arises by rolling the parabola $y = x^2/(4a)$ on the x-axis with pole the focus of the parabola (that is, $P = (0, a)$ in the initial position). The catenary is the shape taken under the action of gravity by a chain or string of uniform density whose ends are held in the air.

A circle rolling on a straight line gives a *trochoid*, with the *cycloid* as a special case when the pole P lies on the circle (Figure 4.29). If the moving circle M has radius a and the distance from the pole P to the center of M is k, the trochoid's parametric equation is

$$x = a\phi - k\sin\phi, \qquad y = a - k\cos\phi. \tag{4.12.2}$$

The cycloid, therefore, has the parametric equation

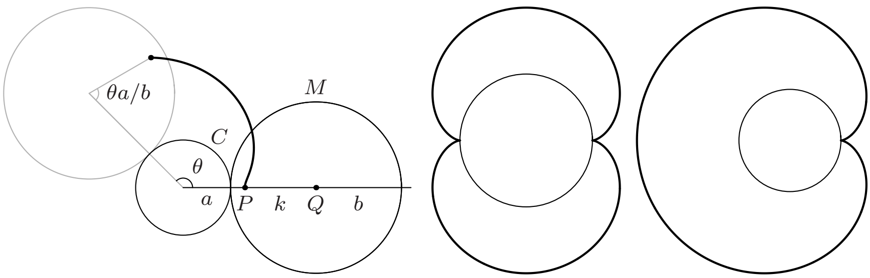
$$x = a(\phi - \sin\phi), \qquad y = a(1 - \cos\phi). \tag{4.12.3}$$

One can eliminate ϕ to get x as a (multivalued) function of y, which takes the following form for the cycloid:

$$x = \pm\left(a\cos^{-1}\left(\frac{a-y}{a}\right) - \sqrt{2ay - y^2}\right) \tag{4.12.4}$$

The length of one arch of the cycloid is $8a$, and the area under the arch is $3\pi a^2$.

FIGURE 4.30
Left: initial configuration for epitrochoid (black) and configuration at parameter value θ (gray). Middle: epicycloid with $b = \frac{1}{2}a$ (nephroid). Right: epicycloid with $b = a$ (cardioid).



A trochoid is also called a *curtate cycloid* when $k < a$ (that is, when P is inside the circle) and a *prolate cycloid* when $k > a$.

A circle rolling on another circle and exterior to it gives an *epitrochoid*. If a is the radius of the fixed circle, b that of the rolling circle, and k is the distance from P to the center of the rolling circle, the parametric equation of the epitrochoid is

$$x = (a+b)\cos\theta - k\cos((1+a/b)\theta), \qquad y = (a+b)\sin\theta - k\sin((1+a/b)\theta).$$

These equations assume that, at the start, everything is aligned along the positive x-axis, as in Figure 4.30, left. Usually one considers the case when a/b is a rational number, say $a/b = p/q$ where p and q are relatively prime. Then the rolling circle returns to its original position after rotating q times around the fixed circle, and the epitrochoid is a closed curve—in fact, an algebraic curve. One also usually takes $k = b$, so that P lies on the rolling circle; the curve in this case is called an *epicycloid*. The middle diagram in Figure 4.30 shows the case $b = k = \frac{1}{2}a$, called the *nephroid*; this curve is the cross-section of the caustic of a spherical mirror. The diagram on the right shows the case $b = k = a$, which gives the cardioid (compare to Figure 4.28, middle).

Hypotrochoids and *hypocycloids* are defined in the same way as epitrochoids and epicycloids, but the rolling circle is inside the fixed one. The parametric equation of the hypotrochoid is

$$x = (a-b)\cos\theta + k\cos((a/b-1)\theta), \qquad y = (a-b)\sin\theta - k\sin((a/b-1)\theta),$$

where the letters have the same meaning as for the epitrochoid. Usually one takes a/b rational and $k = b$. There are several interesting particular cases:

- $b = k = a$ gives a point.
- $b = k = \frac{1}{2}a$ gives a diameter of the circle C.
- $b = k = \frac{1}{3}a$ gives the *deltoid* (Figure 4.31, left), whose algebraic equation is

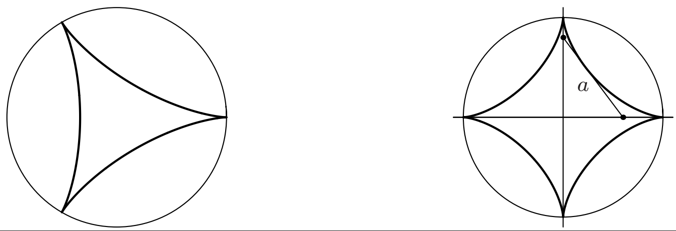
$$(x^2+y^2)^2 - 8ax^3 + 24axy^2 + 18a^2(x^2+y^2) - 27a^4 = 0. \qquad (4.12.5)$$

- $b = k = \frac{1}{4}a$ gives the *astroid* (Figure 4.31, right), an algebraic curve of degree six whose equation can be reduced to $x^{2/3} + y^{2/3} = a^{2/3}$. The figure illustrates another property of the astroid: its tangent intersects the coordinate axes at points that are always the same distance a apart. Otherwise said, the astroid is the envelope of a moving segment of fixed length whose endpoints are constrained to lie on the two coordinate axes.

4.12.3 CURVES IN POLAR COORDINATES

polar equation	type of curve
$r = a$	circle
$r = a\cos\theta$	circle
$r = a\sin\theta$	circle
$r^2 - 2br\cos(\theta-\beta) + (b^2-a^2) = 0$	circle at (b,β) of radius a
$r = \dfrac{k}{1-e\cos\theta}$	$\begin{cases} e=1 & \text{parabola} \\ 0<e<1 & \text{ellipse} \\ e>1 & \text{hyperbola} \end{cases}$

FIGURE 4.31
The hypocycloids with $a = 3b$ (deltoid) and $a = 4b$ (astroid).



4.12.4 SPIRALS

A number of interesting curves have polar equation $r = f(\theta)$, where f is a monotonic function (always increasing or decreasing). This property leads to a spiral shape. The *logarithmic spiral* or *Bernoulli spiral* (Figure 4.12.4, left) is self-similar: by rotation the curve can be made to match any scaled copy of itself. Its equation is $r = ke^{a\theta}$; the angle between the radius from the origin and the tangent to the curve is constant and equal to $\phi = \cot^{-1} a$. A curve parameterized by arc length and such that the radius of curvature is proportional to the parameter at each point is a Bernoulli spiral.

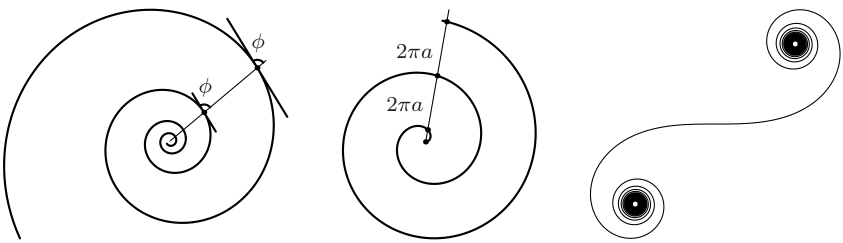


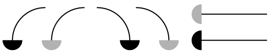
FIGURE 4.32
The Bernoulli or logarithmic spiral (left), the Archimedes or linear spiral (middle), and the Cornu spiral (right).

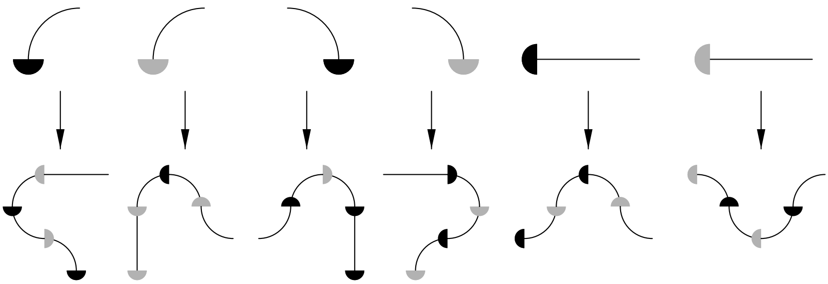
In the *Archimedean spiral* or *linear spiral* (Figure 4.12.4, middle), the spacing between intersections along a ray from the origin is constant. The equation of this spiral is $r = a\theta$; by scaling one can take $a = 1$. It has an inner endpoint, in contrast with the logarithmic spiral, which spirals down to the origin without reaching it. The *Cornu spiral* or *clothoid* (Figure 4.12.4, right), important in optics and engineering, has the following parametric representation in Cartesian coordinates:

$$X = aC(t) = a\int_0^t \cos(\tfrac{1}{2}\pi s^2)\,ds, \qquad y = aS(t) = a\int_0^t \sin(\tfrac{1}{2}\pi s^2)\,ds.$$

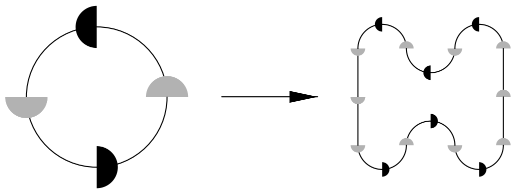
(C and S are the so-called Fresnel integrals; see page 476). A curve parameterized by arc length and such that the radius of curvature is inversely proportional to the parameter at each point is a Cornu spiral (compare to the Bernoulli spiral).

4.12.5 THE PEANO CURVE AND FRACTAL CURVES

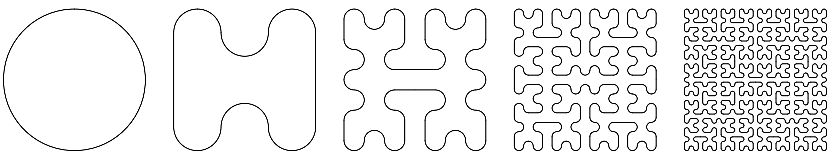
There are curves (in the sense of continuous maps from the real line to the plane) that completely cover a two-dimensional region of the plane. We give a construction of such a *Peano curve*, adapted from David Hilbert's example. The construction is inductive and is based on replacement rules. We consider building blocks of six shapes: , the length of the straight segments being twice the radius of the curved ones. A sequence of these patterns, end-to-end, represents a curve, if we disregard the gray and black half-disks. The replacement rules are the following:

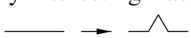


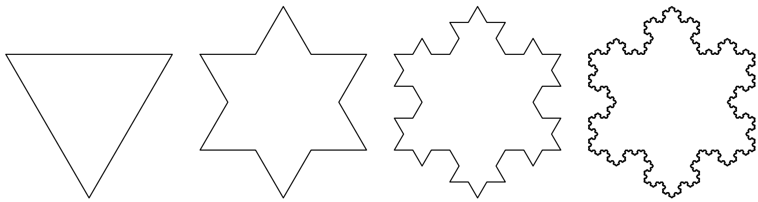
The rules are applied taking into account the way each piece is turned. Here we apply the replacement rules to a particular initial pattern:



(We scale the result so it has the same size as the original.) Applying the process repeatedly gives, in the limit, the Peano curve. Note that the sequence converges uniformly and thus the limit function is continuous. Here are the first five steps:



The same idea of replacement rules leads to many interesting fractal, and often self-similar, curves. For example, the substitution  leads to the *Koch snowflake* when applied to an initial equilateral triangle, like this (the first three stages and the sixth are shown):



4.12.6 FRACTAL OBJECTS

Given an object X, if $n(\epsilon)$ open sets of diameter of ϵ are required to cover X, then the capacity dimension of X is

$$d_{\text{capacity}} = \lim_{\epsilon \to 0} \frac{\ln n(\epsilon)}{\ln \epsilon} \tag{4.12.6}$$

indicating that $n(\epsilon)$ scales as $\epsilon^{d_{\text{capacity}}}$. The capacity dimension of various objects:

Object	Dimension
Logistic equation	0.538
Cantor set	$\frac{\ln 2}{\ln 3} \approx 0.6309$
Koch snowflake	$\frac{2\ln 2}{\ln 3} \approx 1.2619$
Cantor dust	$\frac{\ln 5}{\ln 3} \approx 1.4650$
Minkowski sausage	$\frac{3}{2} = 1.5$

Object	Dimension
Sierpiński sieve	$\frac{\ln 3}{\ln 2} \approx 1.5850$
Pentaflake	$\frac{\ln 2 + \ln 3}{\ln(1+\phi)3} \approx 1.8617$
Sierpiński carpet	$\frac{3\ln 2}{\ln 3} \approx 1.8928$
Tetrix	2
Menger sponge	$\frac{2\ln 2 + \ln 5}{\ln 3} \approx 2.7268$

The logistic equation is in section 3.5.7.3.

The Cantor set C_∞ is obtained by starting with the unit interval $C_0 = [0, 1]$ and sequentially removing the open middle third. Hence, $C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$ and $C_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]$.

4.12.7 CLASSICAL CONSTRUCTIONS

The ancient Greeks used straightedges and compasses to find the solutions to numerical problems. For example, they found square roots by constructing the geometric mean of two segments. Three famous problems that cannot be solved this way are:

1. The trisection of an arbitrary angle.
2. The squaring of the circle (the construction of a square whose area is equal to that of a given circle).
3. The doubling of the cube (the construction of a cube with double the volume of a given cube).

A regular n-gon inscribed in the unit circle can be constructed by straightedge and compass alone if and only if n has the form $n = 2^\ell p_1 p_2 \dots p_k$, where ℓ is a nonnegative integer and $\{p_i\}$ are distinct Fermat primes (primes of the form $2^{2^m}+1$). The only known Fermat primes are for 3, 5, 17, 257, and 65537, corresponding to $m = 1, 2, 3, 4$. Thus, regular n-gons can be constructed with this many sides $n = 3$, 4, 5, 6, 8, 10, 12, 15, 16, 17, 20, 24, ..., 257, ...

4.13 COORDINATE SYSTEMS IN SPACE

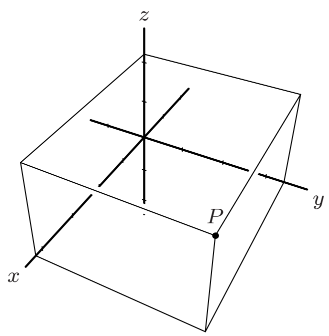
4.13.1 CONVENTIONS

When we talk about "the point with coordinates (x, y, z)" or "the surface with equation $f(x, y, z)$," we always mean Cartesian coordinates. If a formula involves another type of coordinates, this fact will be stated explicitly. Note that Section 4.3.2 has information on substitutions and transformations relevant to the three-dimensional case.

4.13.2 CARTESIAN COORDINATES IN SPACE

In *Cartesian coordinates* (or *rectangular coordinates*), a point P is referred to by three real numbers, indicating the positions of the perpendicular projections from the point to three fixed, perpendicular, graduated lines, called the *axes*. If the coordinates are denoted x, y, z, in that order, the axes are called the *x-axis*, etc., and we write $P = (x, y, z)$. Often the x-axis is imagined to be horizontal and pointing roughly toward the viewer (out of the page), the y-axis also horizontal and pointing more or less to the right, and the z-axis vertical, pointing up. The system is called *right-handed* if it can be rotated so the three axes are in this position. Figure 4.33 shows a right-handed system. The point $x = 0$, $y = 0$, $z = 0$ is the *origin*, where the three axes intersect.

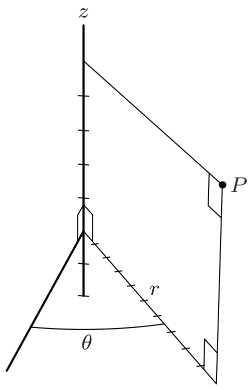
FIGURE 4.33
In Cartesian coordinates, $P = (4.2, 3.4, 2.2)$.



4.13.3 CYLINDRICAL COORDINATES IN SPACE

To define *cylindrical coordinates*, we take an axis (usually called the *z-axis*) and a perpendicular plane, on which we choose a ray (the *initial ray*) originating at the intersection of the plane and the axis (the *origin*). The coordinates of a point P are

FIGURE 4.34
Among the possible sets (r, θ, z) *of cylindrical coordinates for* P *are* $(10, 30°, 5)$ *and* $(10, 390°, 5)$.



the polar coordinates (r, θ) of the projection of P on the plane, and the coordinate z of the projection of P on the axis (Figure 4.34). See Section 4.3.4 for remarks on the values of r and θ.

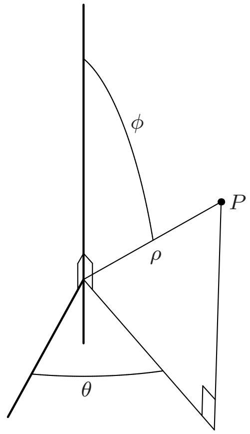
4.13.4 SPHERICAL COORDINATES IN SPACE

To define *spherical coordinates*, we take an axis (the *polar axis*) and a perpendicular plane (the *equatorial plane*), on which we choose a ray (the *initial ray*) originating at the intersection of the plane and the axis (the *origin* O). The coordinates of a point P are the distance ρ from P to the origin, the (*zenith*) angle ϕ between the line OP and the positive polar axis, and the (*azimuth*) angle θ between the initial ray and the projection of OP to the equatorial plane. See Figure 4.35. As in the case of polar and cylindrical coordinates, θ is only defined up to multiples of $360°$, and likewise ϕ. Usually ϕ is assigned a value between 0 and $180°$, but values of ϕ between $180°$ and $360°$ can also be used; the triples (ρ, ϕ, θ) and $(\rho,\ 360° - \phi,\ 180° + \theta)$ represent the same point. Similarly, one can extend ρ to negative values; the triples (ρ, ϕ, θ) and $(-\rho,\ 180° - \phi,\ 180° + \theta)$ represent the same point.

4.13.5 RELATIONS BETWEEN CARTESIAN, CYLINDRICAL, AND SPHERICAL COORDINATES

Consider a Cartesian, a cylindrical, and a spherical coordinate system, related as shown in Figure 4.36. The Cartesian coordinates (x, y, z), the cylindrical coordinates (r, θ, z), and the spherical coordinates (ρ, ϕ, θ) of a point are related as follows

FIGURE 4.35
A set of spherical coordinates for P is $(\rho, \theta, \phi) = (10, 60°, 30°)$.



(where the $\tan^{-1}$ function must be interpreted correctly in all quadrants):

cart ↔ cyl
$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \\ z = z \end{cases} \qquad \begin{cases} r = \sqrt{x^2+y^2} \\ \theta = \tan^{-1}\dfrac{y}{x} \\ z = z \end{cases} \qquad \begin{cases} \sin\theta = \dfrac{y}{\sqrt{x^2+y^2}} \\ \cos\theta = \dfrac{x}{\sqrt{x^2+y^2}} \\ z = z \end{cases}$$

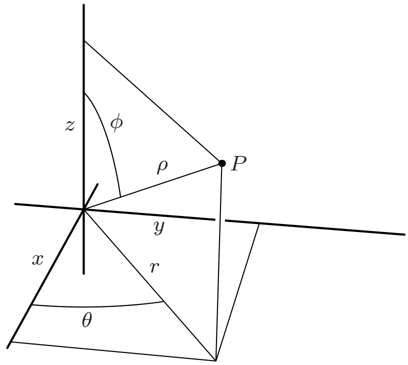
cyl ↔ sph
$$\begin{cases} r = \rho\sin\phi \\ z = \rho\cos\phi \\ \theta = \theta \end{cases} \qquad \begin{cases} \rho = \sqrt{r^2+z^2} \\ \phi = \tan^{-1}\dfrac{r}{z} \\ \theta = \theta \end{cases} \qquad \begin{cases} \sin\phi = \dfrac{r}{\sqrt{r^2+z^2}} \\ \cos\phi = \dfrac{z}{\sqrt{r^2+z^2}} \\ \theta = \theta \end{cases}$$

cart ↔ sph
$$\begin{cases} x = \rho\cos\theta\sin\phi \\ y = \rho\sin\theta\sin\phi \\ z = \rho\cos\phi \end{cases} \qquad \begin{cases} \rho = \sqrt{x^2+y^2+z^2} \\ \theta = \tan^{-1}\dfrac{y}{x} \\ \phi = \tan^{-1}\dfrac{\sqrt{x^2+y^2}}{z} \\ \quad = \cos^{-1}\dfrac{z}{\sqrt{x^2+y^2+z^2}} \end{cases}$$

4.13.6 HOMOGENEOUS COORDINATES IN SPACE

A quadruple of real numbers $(x : y : z : t)$, with $t \neq 0$, is a set of *homogeneous coordinates* for the point P with Cartesian coordinates $(x/t,\ y/t,\ z/t)$. Thus the same point has many sets of homogeneous coordinates: $(x : y : z : t)$ and $(x' : y' : z' : t')$ represent the same point if and only if there is some real number α such that $x' = \alpha x$, $y' = \alpha y$, $z' = \alpha z$, $t' = \alpha t$. If P has Cartesian coordinates (x_0, y_0, z_0),

FIGURE 4.36
Standard relations between Cartesian, cylindrical, and spherical coordinate systems. The origin is the same for all three. The positive z-axes of the Cartesian and cylindrical systems coincide with the positive polar axis of the spherical system. The initial rays of the cylindrical and spherical systems coincide with the positive x-axis of the Cartesian system, and the rays $\theta = 90^\circ$ coincide with the positive y-axis.



one set of homogeneous coordinates for P is $(x_0, y_0, z_0, 1)$.

Section 4.3.5 has more information on the relationship between Cartesian and homogeneous coordinates. Section 4.14.2 has formulas for space transformations in homogeneous coordinates.

4.14 SPACE SYMMETRIES OR ISOMETRIES

A transformation of space (invertible map of space to itself) that preserves distances is called an *isometry* of space. Every isometry of space is a composition of transformations of the following types:

1. The *identity* (which leaves every point fixed)
2. A *translation* by a vector **v**
3. A *rotation* through an angle α around a line L
4. A *screw motion* through an angle α around a line L, with displacement d
5. A *reflection* in a plane P
6. A *glide-reflection* in a plane P with displacement vector **v**
7. A *rotation-reflection* (rotation through an angle α around a line L composed with reflection in a plane perpendicular to L).

The identity is a particular case of a translation and of a rotation; rotations are particular cases of screw motions; reflections are particular cases of glide-reflections. However, as in the plane case, it is more intuitive to consider each case separately.

4.14.1 SYMMETRIES: CARTESIAN COORDINATES

In the formulas below, multiplication between a matrix and a triple of coordinates should be carried out regarding the triple as a column vector (or a matrix with three rows and one column).

1. *Translation* by (x_0, y_0, z_0):
$$(x, y, z) \mapsto (x + x_0,\ y + y_0,\ z + z_0). \tag{4.14.1}$$
2. *Rotation* through α (counterclockwise) around the line through the origin with direction cosines a, b, c (so that $a^2 + b^2 + c^2 = 1$, see page 257): $(x, y, z) \mapsto M(x, y, z)$, where M is the matrix
$$\begin{bmatrix} a^2(1-\cos\alpha)+\cos\alpha & ab(1-\cos\alpha)-c\sin\alpha & ac(1-\cos\alpha)+b\sin\alpha \\ ab(1-\cos\alpha)+c\sin\alpha & b^2(1-\cos\alpha)+\cos\alpha & bc(1-\cos\alpha)-a\sin\alpha \\ ac(1-\cos\alpha)-b\sin\alpha & bc(1-\cos\alpha)+a\sin\alpha & c^2(1-\cos\alpha)+\cos\alpha \end{bmatrix}. \tag{4.14.2}$$
3. *Rotation* through α (counterclockwise) around the line with direction cosines a, b, c through an arbitrary point (x_0, y_0, z_0):
$$(x, y, z) \mapsto (x_0, y_0, z_0) + M(x - x_0,\ y - y_0,\ z - z_0), \tag{4.14.3}$$
where M is given by Equation (4.14.2).
4. *Arbitrary rotations and Euler angles*: Any rotation of space fixing the origin can be decomposed as a rotation by ϕ about the z-axis, followed by a rotation by θ about the y-axis, followed by a rotation by ψ about the z-axis. The numbers ϕ, θ and ψ are called the *Euler angles* of the composite rotation, which acts as: $(x, y, z) \mapsto M(x, y, z)$, where M is the matrix

$$\begin{bmatrix} \cos\phi\cos\theta\cos\psi - \sin\phi\sin\psi & -\sin\phi\cos\theta\cos\psi - \cos\phi\sin\psi & \sin\theta\cos\psi \\ \sin\phi\cos\psi + \cos\phi\cos\theta\sin\psi & -\sin\phi\cos\theta\sin\psi + \cos\phi\cos\psi & \sin\theta\sin\psi \\ -\cos\phi\sin\theta & \sin\theta\sin\phi & \cos\theta \end{bmatrix} \tag{4.14.4}$$

(An alternative decomposition, more natural if we think of the coordinate system as a rigid trihedron that rotates in space, is the following: a rotation by ψ about the z-axis, followed by a rotation by θ about the *rotated* y-axis, followed by a rotation by ϕ about the *rotated* z-axis. Note that the order is reversed.)

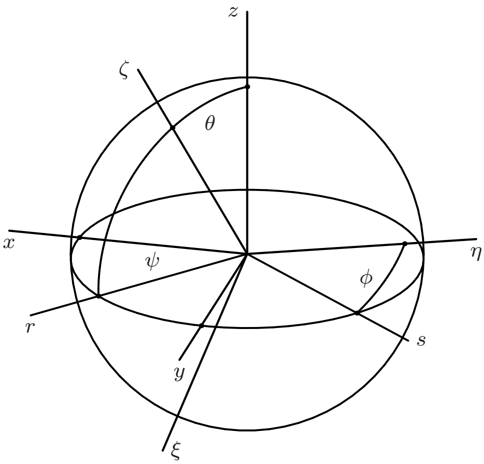
Provided that θ is not a multiple of $180°$, the decomposition of a rotation in this form is unique (apart from the ambiguity arising from the possibility of adding a multiple of $360°$ to any angle). Figure 4.37 shows how the Euler angles can be read off geometrically.

Warning: Some references define Euler angles differently; the most common variation is that the second rotation is taken about the x-axis instead of about the y-axis.

5. *Screw motion* with angle α and displacement d around the line with direction cosines a, b, c through an arbitrary point (x_0, y_0, z_0):
$$(x, y, z) \mapsto (x_0 + ad, y_0 + bd, z_0 + cd) + M(x - x_0,\ y - y_0,\ z - z_0), \tag{4.14.5}$$

FIGURE 4.37

The coordinate rays Ox, Oy, Oz, together with their images $O\xi$, $O\eta$, $O\zeta$ under a rotation, fix the Euler angles associated with that rotation, as follows: $\theta = zO\zeta$, $\psi = xOr = yOs$, and $\phi = sO\eta$. (Here the ray Or is the projection of $O\zeta$ to the xy-plane. The ray Os is determined by the intersection of the xy- and $\xi\eta$-planes.)



where M is given by (4.14.2).

6. *Reflection*

$$\begin{aligned} &\text{in the } xy\text{-plane:} && (x,y,z) \mapsto (x,\ y,\ -z). \\ &\text{in the } xz\text{-plane:} && (x,y,z) \mapsto (x,\ -y,\ z). \\ &\text{in the } yz\text{-plane:} && (x,y,z) \mapsto (-x,\ y,\ z). \end{aligned} \tag{4.14.6}$$

7. *Reflection* in a plane with equation $ax + by + cz + d = 0$:

$$(x,y,z) \mapsto \frac{1}{a^2+b^2+c^2}\big(M(x_0,y_0,z_0) - (2ad, 2bd, 2cd)\big), \tag{4.14.7}$$

where M is the matrix

$$M = \begin{bmatrix} -a^2+b^2+c^2 & -2ab & -2ac \\ -2ab & a^2-b^2+c^2 & -2bc \\ -2ac & -2bc & a^2+b^2-c^2 \end{bmatrix}. \tag{4.14.8}$$

8. *Reflection* in a plane going through (x_0, y_0, z_0) and whose normal has direction cosines a, b, c:

$$(x,y,z) \mapsto (x_0+y_0+z_0) + M(x-x_0,\ y-y_0,\ z-z_0), \tag{4.14.9}$$

where M is as in (4.14.8).

9. *Glide-reflection* in a plane P with displacement vector $\mathbf{v}$: Apply first a reflection in P, then a translation by the vector $\mathbf{v}$.

4.14.2 SYMMETRIES: HOMOGENEOUS COORDINATES

All isometries of space can be expressed in homogeneous coordinates in terms of multiplication by a matrix. As in the case of plane isometries (Section 4.4.2), this means that the successive application of transformations reduces to matrix multiplication. (In the formulas below, $\left[\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix}\right]$ is the 4×4 projective matrix obtained from the 3×3 matrix M by adding a row and a column as stated.)

1. *Translation* by (x_0, y_0, z_0):
$$\begin{bmatrix} 1 & 0 & 0 & x_0 \\ 0 & 1 & 0 & y_0 \\ 0 & 0 & 1 & z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$
2. *Rotation* through the origin:
$$\begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix},$$
where M is given in (4.14.2) or (4.14.4), as the case may be.
3. *Reflection* in a plane through the origin:
$$\begin{bmatrix} M & 0 \\ 0 & 1 \end{bmatrix},$$
where M is given in (4.14.8).

From this, one can deduce all other transformations, as in the case of plane transformations (see page 200).

4.15 OTHER TRANSFORMATIONS OF SPACE

4.15.1 SIMILARITIES

A transformation of space that preserves shapes is called a *similarity*. Every similarity of the plane is obtained by composing a *proportional scaling transformation* (also known as a *homothety*) with an isometry. A proportional scaling transformation centered at the origin has the form

$$(x, y, z) \mapsto (ax, ay, az), \tag{4.15.1}$$

where $a \neq 0$ is the scaling factor (a real number). The corresponding matrix in *homogeneous coordinates* is

$$H_a = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ 0 & 0 & a & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.15.2}$$

In *cylindrical coordinates*, the transformation is $(r, \theta, z) \mapsto (ar, \theta, az)$. In *spherical coordinates*, it is $(r, \phi, \theta) \mapsto (ar, \phi, \theta)$.

4.15.2 AFFINE TRANSFORMATIONS

A transformation that preserves lines and parallelism (maps parallel lines to parallel lines) is an *affine transformation*. There are two important particular cases of such transformations:

1. A *non-proportional scaling transformation* centered at the origin has the form $(x, y, z) \mapsto (ax, by, cz)$, where $a, b, c \neq 0$ are the scaling factors (real numbers). The corresponding matrix in *homogeneous coordinates* is
$$H_{a,b,c} = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.15.3}$$
2. A *shear* in the x-direction and preserving horizontal planes has the form $(x, y, z) \mapsto (x + rz, y, z)$, where r is the shearing factor. The corresponding matrix in *homogeneous coordinates* is
$$S_r = \begin{bmatrix} 1 & 0 & r & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4.15.4}$$

Every affine transformation is obtained by composing a non-proportional scaling transformation with an isometry, or one or two shears with a homothety and an isometry.

4.15.3 PROJECTIVE TRANSFORMATIONS

A transformation that maps lines to lines (but does not necessarily preserve parallelism) is a *projective transformation*. Any spatial projective transformation can be expressed by an invertible 4×4 matrix in homogeneous coordinates; conversely, any invertible 4×4 matrix defines a projective transformation of space. Projective transformations (if not affine) are not defined on all of space, but only on the complement of a plane (the missing plane is "mapped to infinity").

The following particular case is often useful, especially in computer graphics, in *projecting a scene* from space to the plane. Suppose an observer is at the point $E = (x_0, y_0, z_0)$ of space, looking toward the origin $O = (0, 0, 0)$. Let P, the *screen*, be the plane through O and perpendicular to the ray EO. Place a rectangular coordinate system $\xi\eta$ on P with origin at O so that the positive η-axis lies in the half-plane determined by E and the positive z-axis of space (that is, the z-axis is pointing "up" as seen from E). Then consider the transformation that associates with a point $X = (x, y, z)$ the triple (ξ, η, ζ), where (ξ, η) are the coordinates of the point, where the line EX intersects P (the *screen coordinates* of X as seen from E), and ζ is the inverse of the signed distance from X to E along the line EO (this distance is the *depth* of X as seen from E). This is a projective transformation, given by the matrix

$$\begin{bmatrix} -r^2y_0 & r^2x_0 & 0 & 0 \\ -rx_0z_0 & -ry_0z_0 & r\rho^2 & 0 \\ 0 & 0 & 0 & r\rho \\ -\rho x_0 & -\rho y_0 & -\rho z_0 & r^2\rho \end{bmatrix} \tag{4.15.5}$$

with $\rho = \sqrt{x_0^2 + y_0^2}$ and $r = \sqrt{x_0^2 + y_0^2 + z_0^2}$.

4.16 DIRECTION ANGLES AND DIRECTION COSINES

Given a vector (a, b, c) in three-dimensional space, the *direction cosines* of this vector are

$$\begin{aligned} \cos\alpha &= \frac{a}{\sqrt{a^2+b^2+c^2}}, \\ \cos\beta &= \frac{b}{\sqrt{a^2+b^2+c^2}}, \\ \cos\gamma &= \frac{c}{\sqrt{a^2+b^2+c^2}}. \end{aligned} \tag{4.16.1}$$

Here the *direction angles* α, β, γ are the angles that the vector makes with the positive x-, y- and z-axes, respectively. In formulas, usually the direction cosines appear, rather than the direction angles. We have

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1. \tag{4.16.2}$$

4.17 PLANES

The (Cartesian) equation of a *plane* is linear in the coordinates x, y, and z:

$$ax + by + cz + d = 0. \tag{4.17.1}$$

The *normal direction* to this plane is (a, b, c). The *intersection* of this plane with the x-axis, or *x-intercept*, is $x = -d/a$, the *y-intercept* is $y = -d/b$, and the *z-intercept* is $z = -d/c$. The plane is vertical (perpendicular to the xy-plane) if $c = 0$. It is perpendicular to the x-axis if $b = c = 0$, and likewise for the other coordinates.

When $a^2 + b^2 + c^2 = 1$ and $d \le 0$ in the equation $ax + by + cz + d = 0$, the equation is said to be in *normal form*. In this case $-d$ is the *distance of the plane to the origin*, and (a, b, c) are the *direction cosines* of the normal.

To reduce an arbitrary equation $ax + by + cz + d = 0$ to normal form, divide by $\pm\sqrt{a^2+b^2+c^2}$, where the sign of the radical is chosen opposite the sign of d when $d \ne 0$, the same as the sign of c when $d = 0$ and $c \ne 0$, and the same as the sign of b otherwise. (If b is also equal to 0, then there is no equation.)

4.17.1 PLANES WITH PRESCRIBED PROPERTIES

1. Plane through (x_0, y_0, z_0) and perpendicular to the direction (a, b, c):
$$a(x - x_0) + b(y - y_0) + c(z - z_0) = 0. \tag{4.17.2}$$
2. Plane through (x_0, y_0, z_0) and parallel to the directions (a_1, b_1, c_1) and (a_2, b_2, c_2):
$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0. \tag{4.17.3}$$
3. Plane through (x_0, y_0, z_0) and (x_1, y_1, z_1) and parallel to the direction (a, b, c):
$$\begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ a & b & c \end{vmatrix} = 0. \tag{4.17.4}$$
4. Plane going through (x_0, y_0, z_0), (x_1, y_1, z_1) and (x_2, y_2, z_2):
$$\begin{vmatrix} x & y & z & 1 \\ x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \end{vmatrix} = 0 \quad \text{or} \quad \begin{vmatrix} x - x_0 & y - y_0 & z - z_0 \\ x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \end{vmatrix} = 0. \tag{4.17.5}$$
(The last three formulas remain true in *oblique coordinates*; see page 198.)
5. The *distance* from the point (x_0, y_0, z_0) to the plane $ax + by + cz + d = 0$ is
$$\left| \frac{ax_0 + by_0 + cz_0 + d}{\sqrt{a^2 + b^2 + c^2}} \right|. \tag{4.17.6}$$
6. The *angle* between two planes $a_0x + b_0y + c_0z + d_0 = 0$ and $a_1x + b_1y + c_1z + d_1 = 0$ is
$$\cos^{-1} \frac{a_0a_1 + b_0b_1 + c_0c_1}{\sqrt{a_0^2 + b_0^2 + c_0^2}\sqrt{a_1^2 + b_1^2 + c_1^2}}. \tag{4.17.7}$$
In particular, the two planes are *parallel* when $a_0 : b_0 : c_0 = a_1 : b_1 : c_1$, and *perpendicular* when $a_0a_1 + b_0b_1 + c_0c_1 = 0$.

4.17.2 CONCURRENCE AND COPLANARITY

Four planes $a_0x+b_0y+c_0z+d_0 = 0$, $a_1x+b_1y+c_1z+d_1 = 0$, $a_2x+b_2y+c_2z+d_2 = 0$, and $a_3x + b_3y + c_3z + d_3 = 0$ are *concurrent* (share a point) if and only if
$$\begin{vmatrix} a_0 & b_0 & c_0 & d_0 \\ a_1 & b_1 & c_1 & d_1 \\ a_2 & b_2 & c_2 & d_2 \\ a_3 & b_3 & c_3 & d_3 \end{vmatrix} = 0. \tag{4.17.8}$$

Four points (x_0, y_0, z_0), (x_1, y_1, z_1), (x_2, y_2, z_2) and (x_3, y_3, z_3) are *coplanar* (lie on the same plane) if and only if

$$\begin{vmatrix} x_0 & y_0 & z_0 & 1 \\ x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ x_3 & y_3 & z_3 & 1 \end{vmatrix} = 0. \tag{4.17.9}$$

(Both of these assertions remain true in *oblique coordinates*.)

4.18 LINES IN SPACE

Two planes that are not parallel or coincident intersect in a *straight line*, such that one can express a line by a pair of linear equations

$$\begin{Bmatrix} ax + by + cz + d = 0 \\ a'x + b'y + c'z + d' = 0 \end{Bmatrix} \tag{4.18.1}$$

such that $bc' - cb'$, $ca' - ac'$, and $ab' - ba'$ are not all zero. The line thus defined is parallel to the vector $(bc' - cb', ca' - ac', ab' - ba')$. The *direction cosines* of the line are those of this vector. See Equation (4.16.1). (The direction cosines of a line are only defined up to a simultaneous change in sign, because the opposite vector still gives the same line.)

The following particular cases are important:

1. Line through (x_0, y_0, z_0) parallel to the vector (a, b, c):
$$\frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - z_0}{c}. \tag{4.18.2}$$
2. Line through (x_0, y_0, z_0) and (x_1, y_1, z_1):
$$\frac{x - x_0}{x_1 - x_0} = \frac{y - y_0}{y_1 - y_0} = \frac{z - z_0}{z_1 - z_0}. \tag{4.18.3}$$
This line is parallel to the vector $(x_1 - x_0,\ y_1 - y_0,\ z_1 - z_0)$.

4.18.1 DISTANCES

1. The *distance* between two points in space is the *length of the line segment* joining them. The distance between the points (x_0, y_0, z_0) and (x_1, y_1, z_1) is
$$\sqrt{(x_1 - x_0)^2 + (y_1 - y_0)^2 + (z_1 - z_0)^2}. \tag{4.18.4}$$
2. The point $k\%$ of the way from $P_0 = (x_0, y_0, z_0)$ to $P_1 = (x_1, y_1, z_1)$ is
$$\left(\frac{kx_1 + (100 - k)x_0}{100}, \frac{ky_1 + (100 - k)y_0}{100}, \frac{kz_1 + (100 - k)z_0}{100}\right). \tag{4.18.5}$$
(The same formula also applies in oblique coordinates.) This point divides the segment P_0P_1 in the ratio $k : (100 - k)$. As a particular case, the *midpoint* of

P_0P_1 is given by

$$\left(\frac{x_1+x_0}{2}, \frac{y_1+y_0}{2}, \frac{z_1+z_0}{2}\right). \tag{4.18.6}$$

3. The *distance* between the point (x_0, y_0, z_0) and the line through (x_1, y_1, z_1) in direction (a, b, c):

$$\sqrt{\frac{\begin{vmatrix} y_0-y_1 & z_0-z_1 \\ b & c \end{vmatrix}^2 + \begin{vmatrix} z_0-z_1 & x_0-x_1 \\ c & a \end{vmatrix}^2 + \begin{vmatrix} x_0-x_1 & y_0-y_1 \\ a & b \end{vmatrix}^2}{a^2+b^2+c^2}} \tag{4.18.7}$$

4. The *distance* between the line through (x_0, y_0, z_0) in direction (a_0, b_0, c_0) and the line through (x_1, y_1, z_1) in direction (a_1, b_1, c_1):

$$\left| \frac{\begin{vmatrix} x_1-x_0 & y_1-y_0 & z_1-z_0 \\ a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \end{vmatrix}}{\sqrt{\begin{vmatrix} b_0 & c_0 \\ b_1 & c_1 \end{vmatrix}^2 + \begin{vmatrix} c_0 & a_0 \\ c_1 & a_1 \end{vmatrix}^2 + \begin{vmatrix} a_0 & b_0 \\ a_1 & b_1 \end{vmatrix}^2}} \right|. \tag{4.18.8}$$

4.18.2 ANGLES

The *angle* between lines with directions (a_0, b_0, c_0) and (a_1, b_1, c_1):

$$\cos^{-1}\left(\frac{a_0a_1+b_0b_1+c_0c_1}{\sqrt{a_0^2+b_0^2+c_0^2}\sqrt{a_1^2+b_1^2+c_1^2}}\right). \tag{4.18.9}$$

In particular, the two lines are *parallel* when $a_0 : b_0 : c_0 = a_1 : b_1 : c_1$, and *perpendicular* when $a_0a_1 + b_0b_1 + c_0c_1 = 0$.

The *angle* between lines with direction angles $\alpha_0, \beta_0, \gamma_0$ and $\alpha_1, \beta_1, \gamma_1$:

$$\cos^{-1}(\cos\alpha_0\cos\alpha_1 + \cos\beta_0\cos\beta_1 + \cos\gamma_0\cos\gamma_1). \tag{4.18.10}$$

4.18.3 CONCURRENCE, COPLANARITY, PARALLELISM

Two lines, each specified by point and direction, are *coplanar* if and only if the determinant in the numerator of Equation (4.18.8) is zero. In this case they are *concurrent* (if the denominator is non-zero) or *parallel* (if the denominator is zero).

Three lines with directions (a_0, b_0, c_0), (a_1, b_1, c_1) and (a_2, b_2, c_2) are *parallel to a common plane* if and only if

$$\begin{vmatrix} a_0 & b_0 & c_0 \\ a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \end{vmatrix} = 0. \tag{4.18.11}$$

4.19 POLYHEDRA

For any polyhedron topologically equivalent to a sphere—in particular, for any *convex polyhedron*—the *Euler formula* holds:

$$v - e + f = 2, \tag{4.19.1}$$

where v is the number of vertices, e is the number of edges, and f is the number of faces.

Many common polyhedra are particular cases of cylinders (Section 4.20) or cones (Section 4.21). A cylinder with a polygonal base (the base is also called a directrix) is called a *prism*. A cone with a polygonal base is called a *pyramid*. A frustum of a cone with a polygonal base is called a *truncated pyramid*. Formulas (4.20.1), (4.21.1), and (4.21.2) give the volumes of a general cylinder, cone, and truncated cone.

A prism whose base is a parallelogram is a *parallelepiped*. The *volume* of a parallelepiped with one vertex at the origin and adjacent vertices at (x_1, y_1, z_1), (x_2, y_2, z_2), and (x_3, y_3, z_3) is given by

$$\text{volume} = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}. \tag{4.19.2}$$

The *rectangular parallelepiped* is a particular case: all of its faces are rectangles. If the side lengths are a, b, c, the *volume* is abc, the *total surface area* is $2(ab+ac+bc)$, and each *diagonal* has length $\sqrt{a^2+b^2+c^2}$. When $a = b = c$ we get the *cube*. See Section 4.19.1.

A pyramid whose base is a triangle is a *tetrahedron*. The *volume* of a tetrahedron with one vertex at the origin and the other vertices at (x_1, y_1, z_1), (x_2, y_2, z_2), and (x_3, y_3, z_3) is given by

$$\text{volume} = \frac{1}{6}\begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix}. \tag{4.19.3}$$

In a tetrahedron with vertices P_0, P_1, P_2, P_3, let d_{ij} be the distance (edge length) from P_i to P_j. Form the determinants

$$\Delta = \begin{vmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & d_{01}^2 & d_{02}^2 & d_{03}^2 \\ 1 & d_{01}^2 & 0 & d_{12}^2 & d_{13}^2 \\ 1 & d_{02}^2 & d_{12}^2 & 0 & d_{23}^2 \\ 1 & d_{03}^2 & d_{13}^2 & d_{23}^2 & 0 \end{vmatrix} \quad \text{and} \quad \Gamma = \begin{vmatrix} 0 & d_{01}^2 & d_{02}^2 & d_{03}^2 \\ d_{01}^2 & 0 & d_{12}^2 & d_{13}^2 \\ d_{02}^2 & d_{12}^2 & 0 & d_{23}^2 \\ d_{03}^2 & d_{13}^2 & d_{23}^2 & 0 \end{vmatrix}. \tag{4.19.4}$$

Then the *volume* of the tetrahedron is $\sqrt{|\Delta|/288}$, and the radius of the *circumscribed sphere* is $\frac{1}{2}\sqrt{|\Gamma/2\Delta|}$.

Expanding the determinant we find that the volume V satisfies the formula:

$$\begin{aligned}144V^2 = &- d_{01}^2 d_{12}^2 d_{02}^2 - d_{01}^2 d_{13}^2 d_{03}^2 - d_{12}^2 d_{13}^2 d_{23}^2 - d_{02}^2 d_{03}^2 d_{23}^2 \\ &+ d_{01}^2 d_{02}^2 d_{13}^2 + d_{12}^2 d_{02}^2 d_{13}^2 + d_{01}^2 d_{12}^2 d_{03}^2 + d_{12}^2 d_{02}^2 d_{03}^2 \\ &+ d_{12}^2 d_{13}^2 d_{03}^2 + d_{02}^2 d_{13}^2 d_{03}^2 + d_{01}^2 d_{12}^2 d_{23}^2 + d_{01}^2 d_{02}^2 d_{23}^2 \\ &+ d_{01}^2 d_{13}^2 d_{23}^2 + d_{02}^2 d_{13}^2 d_{23}^2 + d_{01}^2 d_{03}^2 d_{23}^2 + d_{12}^2 d_{03}^2 d_{23}^2 \\ &- d_{02}^2 d_{02}^2 d_{13}^2 - d_{02}^2 d_{13}^2 d_{13}^2 \\ &- d_{12}^2 d_{12}^2 d_{03}^2 - d_{12}^2 d_{03}^2 d_{03}^2 \\ &- d_{01}^2 d_{01}^2 d_{23}^2 - d_{01}^2 d_{23}^2 d_{23}^2 .\end{aligned} \tag{4.19.5}$$

(Mnemonic: Each of the first four negative terms corresponds to a closed path around a face; each positive term to an open path along three consecutive edges; each remaining negative term to a pair of opposite edges with weights 2 and 1. All such edge combinations are represented.)

For an arbitrary tetrahedron, let P be a vertex and let a, b, c be the lengths of the edges converging on P. If A, B, C are the angles between the same three edges, the volume of the tetrahedron is

$$V = \frac{1}{6}abc\sqrt{1 - \cos^2 A - \cos^2 B - \cos^2 C + 2\cos A \cos B \cos C}. \tag{4.19.6}$$

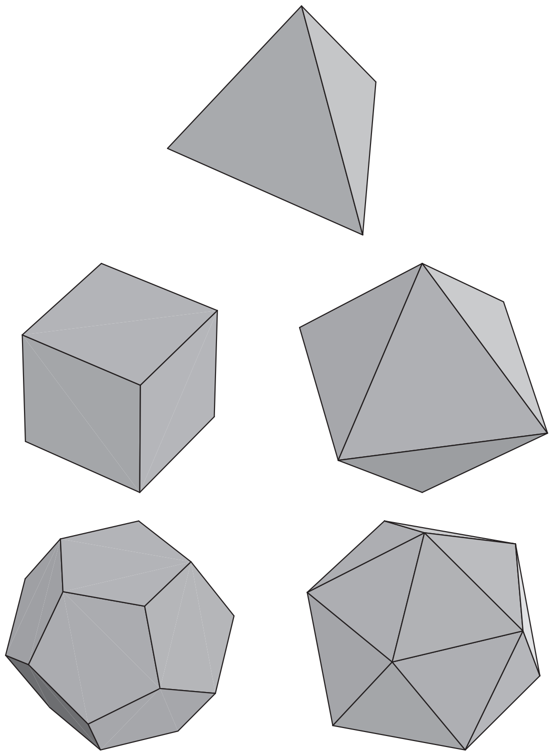
4.19.1 CONVEX REGULAR POLYHEDRA

Figure 4.38 shows the five regular polyhedra, or *Platonic solids*. In the following tables and formulas, a is the length of an edge, θ the dihedral angle at each edge, R the radius of the circumscribed sphere, r the radius of the inscribed sphere, V the volume, S the total surface area, v the total number of vertices, e the total number of edges, f the total number of faces, p the number of edges in a face (3 for equilateral triangles, 4 for squares, 5 for regular pentagons), and q the number of edges meeting at a vertex.

Note that $fp = vq = 2e$.

$$\begin{aligned}\theta &= 2\sin^{-1}\left(\frac{\cos(180^\circ/q)}{\sin(180^\circ/p)}\right), & \frac{R}{r} &= \tan\left(\frac{180^\circ}{p}\right)\tan\left(\frac{180^\circ}{q}\right), \\ \frac{R}{a} &= \frac{\frac{1}{2}\sin(180^\circ/q)}{\sin(180^\circ/p)\cos\frac{1}{2}\theta}, & \frac{S}{a^2} &= \frac{fp}{4}\cot\left(\frac{180^\circ}{p}\right), \\ \frac{r}{a} &= \frac{1}{2}\cot\left(\frac{180^\circ}{p}\right)\tan\left(\frac{\theta}{2}\right), & V &= \tfrac{1}{3}rS.\end{aligned} \tag{4.19.7}$$

FIGURE 4.38
The Platonic solids. Top: the tetrahedron (self-dual). Middle: the cube and the octahedron (dual to one another). Bottom: the dodecahedron and the icosahedron (dual to one another).



Name	v	e	f	p	q	$\sin\theta$	θ
Regular tetrahedron	4	6	4	3	3	$2\sqrt{2}/3$	$70°31'44''$
Cube	8	12	6	4	3	1	$90°$
Regular octahedron	6	12	8	3	4	$2\sqrt{2}/3$	$109°28'16''$
Regular dodecahedron	20	30	12	5	3	$2/\sqrt{5}$	$116°33'54''$
Regular icosahedron	12	30	20	3	5	$2/3$	$138°11'23''$

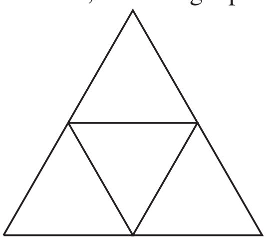
Name	R/a		r/a	
Tetrahedron	$\sqrt{6}/4$	0.612372	$\sqrt{6}/12$	0.204124
Cube	$\sqrt{3}/2$	0.866025	$\frac{1}{2}$	0.5
Octahedron	$\sqrt{2}/2$	0.707107	$\sqrt{6}/6$	0.408248
Dodecahedron	$\frac{1}{4}(\sqrt{15}+\sqrt{3})$	1.401259	$\frac{1}{20}\sqrt{250+110\sqrt{5}}$	1.113516
Icosahedron	$\frac{1}{4}\sqrt{10+2\sqrt{5}}$	0.951057	$\frac{1}{12}\sqrt{42+18\sqrt{5}}$	0.755761

Name	S/a^2		V/a^3	
Tetrahedron	$\sqrt{3}$	1.73205	$\sqrt{2}/12$	0.117851
Cube	6	$6.$	1	$1.$
Octahedron	$2\sqrt{3}$	3.46410	$\sqrt{2}/3$	0.471405
Dodecahedron	$3\sqrt{25+10\sqrt{5}}$	20.64573	$\frac{1}{4}(15+7\sqrt{5})$	7.663119
Icosahedron	$5\sqrt{3}$	8.66025	$\frac{5}{12}(3+\sqrt{5})$	2.181695

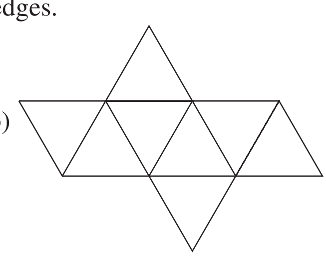
4.19.2 POLYHEDRA NETS

Nets for the five Platonic solids are shown: (a) tetrahedron, (b) octahedron, (c) icosahedron, (d) cube, and (e) dodecahedron. Paper models can be made by making an enlarged photocopy of each, cutting them out along the exterior lines, folding on the interior lines, and using tape to join the edges.

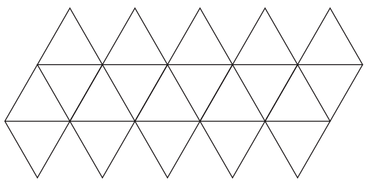
(a)



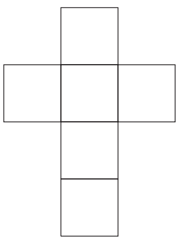
(b)



(c)



(d)



(e)

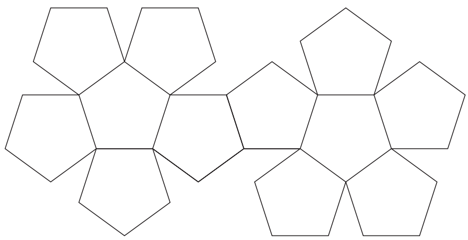
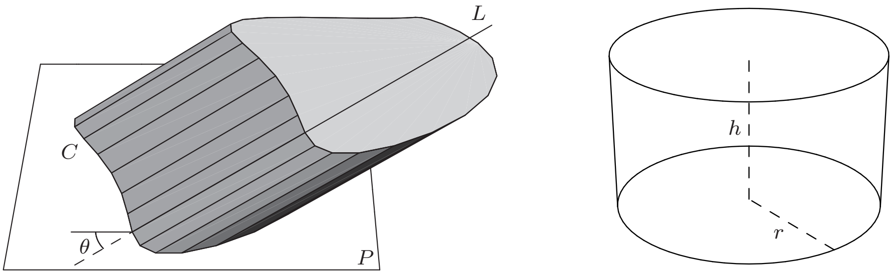


FIGURE 4.39
Left: an oblique cylinder with generator L and directrix C. Right: a right circular cylinder.



4.20 CYLINDERS

Given a line L and a curve C in a plane P, the *cylinder* with *generator* L and *directrix* C is the surface obtained by moving L parallel to itself, so that a point of L is always on C. If L is parallel to the z-axis, the surface's implicit equation does not involve the variable z. Conversely, any implicit equation that does not involve one of the variables (or that can be brought to that form by a change of coordinates) represents a cylinder.

If C is a simple closed curve, we also apply the word *cylinder* to the solid enclosed by the surface generated in this way (Figure 4.39, left). The *volume* contained between P and a plane P' parallel to P is

$$V = Ah = Al\sin\theta, \tag{4.20.1}$$

where A is the area in the plane P enclosed by C, h is the distance between P and P' (measured perpendicularly), l is the length of the segment of L contained between P and P', and θ is the angle that L makes with P. When $\theta = 90°$ we have a *right cylinder*, and $h = l$. For a right cylinder, the *lateral area* between P and P' is hs, where s is the length (circumference) of C.

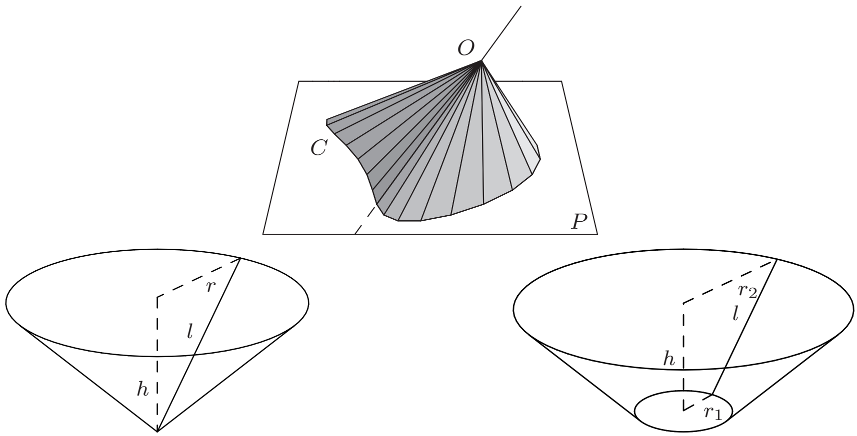
The most important particular case is the *right circular cylinder* (often simply called a *cylinder*). If r is the radius of the base and h is the altitude (Figure 4.39, right), the *lateral area* is $2\pi rh$, the *total area* is $2\pi r(r+h)$, and the *volume* is $\pi r^2 h$. The *implicit equation* of this surface can be written $x^2 + y^2 = r^2$; see also page 219.

4.21 CONES

Given a curve C in a plane P and a point O not in P, the *cone* with *vertex* O and *directrix* C is the surface obtained as the union of all rays that join O with points of C. If O is the origin and the surface is given implicitly by an algebraic equation, that equation is homogeneous (all terms have the same total degree in the variables). Conversely, any homogeneous implicit equation (or one that can be made homogeneous by a change of coordinates) represents a cone.

FIGURE 4.40

Top: a cone with vertex O and directrix C. Bottom left: a right circular cone. Bottom right: A frustum of the latter.



If C is a simple closed curve, we also apply the word *cone* to the solid enclosed by the surface generated in this way (Figure 4.40, top). The *volume* contained between P and the vertex O is

$$V = \tfrac{1}{3}Ah, \tag{4.21.1}$$

where A is the area in the plane P enclosed by C and h is the distance from O and P (measured perpendicularly).

The solid contained between P and a plane P' parallel to P (on the same side of the vertex) is called a *frustum*. Its volume is

$$V = \tfrac{1}{3}h(A + A' + \sqrt{AA'}), \tag{4.21.2}$$

where A and A' are the areas enclosed by the sections of the cone by P and P' (often called the *bases* of the frustum), and h is the distance between P and P'.

The most important particular case of a cone is the *right circular cone* (often simply called a *cone*). If r is the radius of the base, h is the altitude, and l is the length between the vertex and a point on the base circle (Figure 4.40, bottom left), the following relationships apply:

$$\begin{aligned} l &= \sqrt{r^2 + h^2}, \\ \text{Lateral area} &= \pi r l = \pi r \sqrt{r^2 + h^2}, \\ \text{Total area} &= \pi r l + \pi r^2 = \pi r (r + \sqrt{r^2 + h^2}), \text{ and} \\ \text{Volume} &= \tfrac{1}{3}\pi r^2 h. \end{aligned}$$

The *implicit equation* of this surface can be written $x^2 + y^2 = z^2$; see Section 4.9.

For a *frustum* of a right circular cone (Figure 4.40, bottom right),

$$l = \sqrt{(r_1 - r_2)^2 + h^2},$$
$$\text{Lateral area} = \pi(r_1 + r_2)l,$$
$$\text{Total area} = \pi(r_1^2 + r_2^2 + (r_1 + r_2)l), \quad \text{and}$$
$$\text{Volume} = \tfrac{1}{3}\pi h(r_1^2 + r_2^2 + r_1 r_2).$$

4.22 DIFFERENTIAL GEOMETRY

4.22.1 CURVES

4.22.1.1 Definitions

1. A *regular parametric representation of class C^k*, $k \geq 1$, is a vector valued function $\mathbf{f} : I \to \mathbb{R}^3$, where $I \subset \mathbb{R}$ is an interval that satisfies (i) $\mathbf{f}$ is of class C^k (i.e., has continuous k^{th} order derivatives), and (ii) $\mathbf{f}'(t) \neq 0$, for all $t \in I$. In terms of a standard basis of $\mathbb{R}^3$, we write $\mathbf{x} = \mathbf{f}(t) = (f_1(t), f_2(t), f_3(t))$, where the real valued functions f_i, $i = 1, 2, 3$ are the *component functions* of $\mathbf{f}$.
2. An *allowable change of parameter of class C^k* is any C^k function $\phi : J \to I$, where J is an interval and $\phi(J) \subset I$, that satisfies $\phi'(\tau) \neq 0$, for all $\tau \in J$.
3. A C^k regular parametric representation $\mathbf{f}$ is *equivalent* to a C^k regular parametric representation $\mathbf{g}$ if and only if an allowable change of parameter ϕ exists so that $\phi(I_g) = I_f$, and $\mathbf{g}(\tau) = \mathbf{f}(\phi(\tau))$, for all $\tau \in I_g$.
4. A *regular curve C of class C^k* is an equivalence class of C^k regular parametric representation under the equivalence relation on the set of regular parametric representations defined above.
5. The *arc length* of any regular curve C defined by the regular parametric representation $\mathbf{f}$, with $I_f = [a, b]$, is defined by $L = \displaystyle\int_a^b \left|\mathbf{f}'(u)\right| \, du$.
6. An *arc length parameter* along C is defined by $s = \alpha(t) = \pm \displaystyle\int_c^t \left|\mathbf{f}'(u)\right| \, du$. The choice of sign is arbitrary and c is any number in I_f.
7. A *natural representation of class C^k* of the regular curve defined by the regular parametric representation $\mathbf{f}$ is defined by $\mathbf{g}(s) = \mathbf{f}(\alpha^{-1}(s))$, for all $s \in [0, L]$.
8. A *property* of a regular curve C is any property of a regular parametric representation representing C which is invariant under any allowable change of parameter.
9. Let $\mathbf{g}$ be a natural representation of a regular curve C. The following quantities may be defined at each point $\mathbf{x} = \mathbf{g}(s)$ of C:

Binormal line	$\mathbf{y} = \lambda\mathbf{b}(s) + \mathbf{x}$
Curvature	$\kappa(s) = \mathbf{n}(s) \cdot \mathbf{k}(s)$
Curvature vector	$\mathbf{k}(s) = \dot{\mathbf{t}}(s)$
Moving trihedron	$\{\mathbf{t}(s), \mathbf{n}(s), \mathbf{b}(s)\}$
Normal plane	$(\mathbf{y} - \mathbf{x}) \cdot \mathbf{t}(s) = 0$

Osculating plane	$(\mathbf{y}-\mathbf{x})\cdot\mathbf{b}(s)=0$
Osculating sphere	$(\mathbf{y}-\mathbf{c})\cdot(\mathbf{y}-\mathbf{c})=r^2$ where $\mathbf{c}=\mathbf{x}+\rho(s)\mathbf{n}(s)-(\dot{\kappa}(s)/(\kappa^2(s)\tau(s)))\mathbf{b}(s)$ and $r^2=\rho^2(s)+\kappa^2(s)/(\kappa^4(s)\tau^2(s))$
Principal normal line	$\mathbf{y}=\lambda\mathbf{n}(s)+\mathbf{x}$
Principal normal unit vector	$\mathbf{n}(s)=\pm\mathbf{k}(s)/\lvert\mathbf{k}(s)\rvert$, for $\mathbf{k}(s)\neq 0$ defined to be continuous along C
Radius of curvature	$\rho(s)=1/\lvert\kappa(s)\rvert$, when $\kappa(s)\neq 0$
Rectifying plane	$(\mathbf{y}-\mathbf{x})\cdot\mathbf{n}(s)=0$
Tangent line	$\mathbf{y}=\lambda\mathbf{t}(s)+\mathbf{x}$
Torsion	$\tau(s)=-\mathbf{n}(s)\cdot\dot{\mathbf{b}}(s)$
Unit binormal vector	$\mathbf{b}(s)=\mathbf{t}(s)\times\mathbf{n}(s)$
Unit tangent vector	$\mathbf{t}(s)=\dot{\mathbf{g}}(s)$ with $\left(\dot{\mathbf{g}}(s)=\frac{d\mathbf{g}}{ds}\right)$

4.22.1.2 Results

The arc length L and the arc length parameter s of any regular parametric representation **f** are *invariant* under any allowable change of parameter. Thus, L is a property of the regular curve C defined by **f**.

The arc length parameter satisfies $\frac{ds}{dt}=\alpha'(t)=\pm\left|\mathbf{f}'(t)\right|$, which implies that $\left|\mathbf{f}'(s)\right|=1$, if and only if t is an arc length parameter. Thus, arc length parameters are uniquely determined up to the transformation $s\mapsto\tilde{s}=\pm s+s_0$, where s_0 is any constant.

The curvature, torsion, tangent line, normal plane, principal normal line, rectifying plane, binormal line, and osculating plane are properties of the regular curve C defined by any regular parametric representation **f**.

If $\mathbf{x}=(x(t),y(t),z(t))=\mathbf{f}(t)$ is any regular representation of a regular curve C, the following results hold at point $\mathbf{f}(t)$ of C:

$$\begin{aligned}|\kappa| &= \frac{|\mathbf{x}''\times\mathbf{x}'|}{|\mathbf{x}'|^3}=\frac{\sqrt{(z''y'-y''z')^2+(x''z'-z''x')^2+(y''x'-x''y')^2}}{(x'^2+y'^2+z'^2)^{3/2}}\\ \tau &= \frac{\det(\mathbf{x}',\mathbf{x}'',\mathbf{x}''')}{|\mathbf{x}'\times\mathbf{x}''|^2}=\frac{(\mathbf{x}'\times\mathbf{x}'')\cdot\mathbf{x}''')}{|\mathbf{x}'\times\mathbf{x}''|^2}\\ &= \frac{z'''(x'y''-y'x'')+z''(x'''y'-x'y''')+z'(x''y'''-x'''y'')}{(x'^2+y'^2+z'^2)(x''^2+y''^2+z''^2)}\end{aligned} \tag{4.22.1}$$

The vectors of the moving trihedron satisfy the *Serret–Frenet equations*

$$\dot{\mathbf{t}}=\kappa\mathbf{n},\quad \dot{\mathbf{n}}=-\kappa\mathbf{t}+\tau\mathbf{b},\quad \dot{\mathbf{b}}=-\tau\mathbf{n}. \tag{4.22.2}$$

For any plane curve represented parametrically by $\mathbf{x}=\mathbf{f}(t)=(t,f(t),0)$,

$$|\kappa|=\frac{\left|\frac{d^2x}{dt^2}\right|}{\left(1+\left(\frac{dx}{dt}\right)^2\right)^{3/2}}. \tag{4.22.3}$$

Expressions for the curvature vector **k** and the curvature $|\kappa| = \rho^{-1}$ of a *plane curve* corresponding to different representations are as follows:

- For $\{x = f(t), y = g(t)\}$ $\quad \mathbf{k} = \dfrac{(\dot{x}\ddot{y} - \dot{y}\ddot{x})}{(\dot{x}^2 + \dot{y}^2)^2}(-\dot{y}, \dot{x}) \qquad |\kappa| = \dfrac{|\dot{x}\ddot{y} - \dot{y}\ddot{x}|}{(\dot{x}^2 + \dot{y}^2)^{3/2}}$
- For $y = f(x)$ $\quad \mathbf{k} = \dfrac{y''}{(1 + y'^2)^2}(-y', 1) \qquad |\kappa| = \dfrac{|y''|}{(1 + y'^2)^{3/2}}$
- For $r = f(\theta)$ $\quad \mathbf{k} = \dfrac{(r^2 + 2r'^2 - rr'')}{(r^2 + r'^2)^2}(-\dot{r}\sin\theta - r\cos\theta, \dot{r}\cos\theta - r\sin\theta)$

$$|\kappa| = \frac{r^2 + 2r'^2 - rr''}{(r^2 + r'^2)^{3/2}}$$

The equation of the *osculating circle* of a plane curve is given by

$$(\mathbf{y} - \mathbf{c}) \cdot (\mathbf{y} - \mathbf{c}) = \rho^2, \tag{4.22.4}$$

where $\mathbf{c} = \mathbf{x} + \rho^2\mathbf{k}$ is the *center of curvature*.

THEOREM 4.22.1 *(Fundamental existence and uniqueness theorem)*
Let $\kappa(s)$ and $\tau(s)$ be any continuous functions defined for all $s \in [a, b]$. Then there exists, up to a congruence, a unique space curve C for which κ is the curvature function, τ is the torsion function, and s an arc length parameter along C.

4.22.1.3 Example

A regular parametric representation of the *circular helix* is given by $\mathbf{x} = \mathbf{f}(t) = (a\cos t, a\sin t, bt)$, for $t \in \mathbb{R}$, where $a > 0$ and $b \neq 0$ are constant. Differentiating with respect to t:

$$\begin{aligned} \mathbf{x}' &= (-a\sin t, \quad a\cos t, b), \\ \mathbf{x}'' &= (-a\cos t, -a\sin t, 0), \\ \mathbf{x}''' &= (\quad a\sin t, -a\cos t, 0), \end{aligned} \tag{4.22.5}$$

so that $\frac{ds}{dt} = |\mathbf{x}'| = \sqrt{a^2 + b^2}$. Hence,

1. *Arc length parameter*: $s = \alpha(t) = t(a^2 + b^2)^{\frac{1}{2}}$
2. *Curvature vector*: $\mathbf{k} = \frac{d\mathbf{t}}{ds} = \frac{dt}{ds}\frac{d\mathbf{t}}{dt} = (a^2 + b^2)^{-1}(-a\cos t, -a\sin t, 0)$
3. *Curvature*: $\kappa = |\mathbf{k}| = a(a^2 + b^2)^{-1}$
4. *Principal normal unit vector*: $\mathbf{n} = \mathbf{k}/\,|\mathbf{k}| = (-\cos t, -\sin t, 0)$
5. *Unit tangent vector*: $\mathbf{t} = \mathbf{x}'/|\mathbf{x}'| = (a^2 + b^2)^{-\frac{1}{2}}(-a\sin t, a\cos t, b)$
6. *Unit binormal vector*:
$$\mathbf{b} = \mathbf{t} \times \mathbf{n} = (a^2 + b^2)^{-\frac{1}{2}}(b\sin t, b\cos t, a)$$
$$\dot{\mathbf{b}} = \frac{dt}{ds}\frac{d\mathbf{b}}{dt} = b(a^2 + b^2)^{-1}(\cos t, \sin t, 0)$$
7. *Torsion*: $\tau = -\mathbf{n} \cdot \dot{\mathbf{b}} = b(a^2 + b^2)^{-1}$

The values of $|\kappa|$ and τ can be verified using (4.22.1). The sign of (the invariant) τ determines whether the helix is right-handed, $\tau > 0$, or left-handed, $\tau < 0$.

4.22.2 SURFACES

4.22.2.1 Definitions

1. A *coordinate patch of class C^k*, $k \geq 1$ on a surface $S \subset \mathbb{R}^3$ is a vector valued function $f : U \to S$, where $U \subset \mathbb{R}^2$ is an open set, that satisfies (i) **f** is class C^k on U, (ii) $\frac{\partial \mathbf{f}}{\partial u}(u, v) \times \frac{\partial \mathbf{f}}{\partial v}(u, v) \neq 0$, for all $(u, v) \in U$, and (iii) **f** is one-to-one and bi-continuous on U.

2. In terms of a standard basis of $\mathbb{R}^3$ we write $\mathbf{x} = \mathbf{f}(u, v) = (f_1(u, v), f_2(u, v), f_3(u, v))$, where the real valued functions $\{f_1, f_2, f_3\}$ are the *component functions* of **f**. The notation $\mathbf{x}_1 = \mathbf{x}_u = \frac{\partial \mathbf{f}}{\partial u}$, $\mathbf{x}_2 = \mathbf{x}_v = \frac{\partial \mathbf{f}}{\partial v}$, $u^1 = u, u^2 = v$, is frequently used.

3. A *Monge patch* is a coordinate patch where **f** has the form $\mathbf{f}(u, v) = (u, v, f(u, v))$, where f is a real valued function of class C^k.

4. The *u-parameter curves* $v = v_0$ on S are the images of the lines $v = v_0$ in U. They are parametrically represented by $\mathbf{x} = \mathbf{f}(u, v_0)$. The *$v$-parameter curves* $u = u_0$ are defined similarly.

5. An *allowable parameter transformation* of class C^k is a one-to-one function $\phi : U \to V$, where $U, V \subset \mathbb{R}^2$ are open, that satisfies
$$\det \begin{bmatrix} \frac{\partial \phi^1}{\partial u}(u, v) & \frac{\partial \phi^1}{\partial v}(u, v) \\ \frac{\partial \phi^2}{\partial u}(u, v) & \frac{\partial \phi^2}{\partial v}(u, v) \end{bmatrix} \neq 0, \tag{4.22.6}$$
for all $(u, v) \in U$, where the real valued functions, ϕ^1 and ϕ^2, defined by $\phi(u, v) = (\phi^1(u, v), \phi^2(u, v))$ are the component functions of ϕ. One may also write the parameter transformation as $\tilde{u}^1 = \phi^1(u^1, u^2)$, $\tilde{u}^2 = \phi^2(u^1, u^2)$.

6. A *local property* of surface S is any property of a coordinate patch that is invariant under any allowable parameter transformation.

7. Let **f** define a coordinate patch on a surface S. The following quantities may be defined at each point $\mathbf{x} = \mathbf{f}(u, v)$ on the patch:

Asymptotic direction	A direction $du : dv$ for which $\kappa_n = 0$
Asymptotic line	A curve on S whose tangent line at each point coincides with an asymptotic direction
Dupin's indicatrix	$ex_1^2 + 2fx_1x_2 + gx_2^2 = \pm 1$
Elliptic point	$eg - f^2 > 0$
First fundamental form	$I = d\mathbf{x} \cdot d\mathbf{x} = g_{\alpha\beta}(u, v)\, du^\alpha\, du^\beta$ $= E(u, v)\, du^2 + 2F(u, v)\, du\, dv + G(u, v)\, dv^2$
First fundamental metric coefficients	$\begin{cases} E(u, v) = g_{11}(u, v) = \mathbf{x}_1 \cdot \mathbf{x}_1 \\ F(u, v) = g_{12}(u, v) = \mathbf{x}_1 \cdot \mathbf{x}_2 \\ G(u, v) = g_{22}(u, v) = \mathbf{x}_2 \cdot \mathbf{x}_2 \end{cases}$

Fundamental differential	$d\mathbf{x} = \mathbf{x}_\alpha\, du^\alpha = \mathbf{x}_u\, du + \mathbf{x}_v\, dv$ (a repeated upper and lower index signifies a summation for $\alpha = 1, 2$)
Gaussian curvature	$K = \kappa_1\kappa_2 = \dfrac{eg - f^2}{EG - F^2}$
Geodesic curvature vector of curve C on S through x	$\mathbf{k}_g = \mathbf{k} - (\mathbf{k}\cdot\mathbf{N})\mathbf{N} = [\ddot{u}^\alpha + \Gamma^\alpha_{\beta\gamma}\dot{u}^\beta\dot{u}^\gamma]\mathbf{x}_\alpha$ where $\Gamma^\alpha_{\beta\gamma}$ denote the Christoffel symbols of the second kind for the metric $g_{\alpha\beta}$, defined in Section 5.9.3
Geodesic on S	A curve on S which satisfies $\mathbf{k}_g = 0$ at each point
Hyperbolic point	$eg - f^2 < 0$
Line of curvature	A curve on S whose tangent line at each point coincides with a principal direction
Mean curvature	$H = \dfrac{\kappa_1 + \kappa_2}{2} = \dfrac{gE + eG - 2fF}{2(EG - F^2)}$
Normal curvature vector of curve C on S through x	$\mathbf{k}_n = (\mathbf{k}\cdot\mathbf{N})\mathbf{N}$
Normal curvature in the $du : dv$ direction	$\kappa_n = \mathbf{k}\cdot\mathbf{N} = \dfrac{II}{I}$
Normal line	$\mathbf{y} = \lambda\mathcal{N} + \mathbf{x}$
Normal vector	$\mathcal{N} = \mathbf{x}_u \times \mathbf{x}_v$
Parabolic point	$eg - f^2 = 0$ not all of $e, f, g = 0$
Planar point	$e = f = g = 0$
Principal curvatures	The extreme values κ_1 and κ_2 of κ_n
Principal directions	The perpendicular directions $du : dv$ in which κ_n attains its extreme values
Second fundamental form	$II = -d\mathbf{x}\cdot d\mathbf{N} = b_{\alpha\beta}(u,v)\, du^\alpha\, du^\beta$ $= e(u,v)\, du^2 + 2f(u,v)\, du\, dv + g(u,v)\, dv^2$
Second fundamental metric coefficients	$\begin{cases} e(u,v) = b_{11}(u,v) = \mathbf{x}_{11}\cdot\mathbf{N} \\ f(u,v) = b_{12}(u,v) = \mathbf{x}_{12}\cdot\mathbf{N} \\ g(u,v) = b_{22}(u,v) = \mathbf{x}_{22}\cdot\mathbf{N} \end{cases}$
Tangent plane	$(\mathbf{y} - \mathbf{x})\cdot\mathcal{N} = 0$, or $\mathbf{y} = \mathbf{x} + \lambda\mathbf{x}_u + \mu\mathbf{x}_v$
Umbilical point	$\kappa_n =$ constant for all directions $du : dv$
Unit normal vector	$\mathbf{N} = \dfrac{\mathbf{x}_u \times \mathbf{x}_v}{\lvert\mathbf{x}_u \times \mathbf{x}_v\rvert}$

4.22.2.2 Results

1. The tangent plane, normal line, first fundamental form, second fundamental form and all derived quantities thereof are local properties of any surface S.

2. The transformation laws for the first and second fundamental metric coefficients under any allowable parameter transformation are given, respectively, by
$$\tilde{g}_{\alpha\beta} = g_{\gamma\delta}\frac{\partial u^\gamma}{\partial \tilde{u}^\alpha}\frac{\partial u^\delta}{\partial \tilde{u}^\beta}, \quad \text{and} \quad \tilde{b}_{\alpha\beta} = b_{\gamma\delta}\frac{\partial u^\gamma}{\partial \tilde{u}^\alpha}\frac{\partial u^\delta}{\partial \tilde{u}^\beta}. \tag{4.22.7}$$
Thus $g_{\alpha\beta}$ and $b_{\alpha\beta}$ are the components of type $(0, 2)$ tensors.

3. $I \geq 0$ for all directions $du : dv$; and $I = 0$ if and only if $du = dv = 0$.

4. The angle θ between two tangent lines to S at $\mathbf{x} = \mathbf{f}(u, v)$ defined by the directions $du : dv$ and $\delta u : \delta v$, is given by
$$\cos\theta = \frac{g_{\alpha\beta}\, du^\alpha \delta u^\beta}{(g_{\alpha\beta}\, du^\alpha\, du^\beta)^{\frac{1}{2}}(g_{\alpha\beta}\delta u^\alpha \delta u^\beta)^{\frac{1}{2}}}. \tag{4.22.8}$$
The angle between the u-parameter curves and the v-parameter curves is given by $\cos\theta = F(u,v)/(E(u,v)G(u,v))^{\frac{1}{2}}$. The u-parameter curves and v-parameter curves are orthogonal if and only if $F(u, v) = 0$.

5. Suppose two curves, $y = f_1(x)$ and $y = f_2(x)$, intersect at the point $P(X, Y)$. If the derivatives exist, then the angle of intersection (α) is given by:
$$\tan\alpha = \frac{f_2'(X) - f_1'(X)}{1 + f_1'(X)f_2'(X)}. \tag{4.22.9}$$
If $(\tan\alpha) > 0$ then α is an acute angle; otherwise, α is an obtuse angle.

6. The arc length of a curve C on S, defined by $\mathbf{x} = \mathbf{f}(u^1(t), u^2(t))$, with $a \leq t \leq b$, is given by
$$\begin{aligned} L &= \int_a^b \sqrt{g_{\alpha\beta}(u^1(t), u^2(t))\dot{u}^\alpha \dot{u}^\beta}\, dt \qquad (4.22.10) \\ &= \int_a^b \sqrt{E(u(t), v(t))\dot{u}^2 + 2F(u(t), v(t))\dot{u}\dot{v} + G(u(t), v(t))\dot{v}^2}\, dt. \end{aligned}$$

7. The area of $S = \mathbf{f}(U)$ is given by
$$\begin{aligned} A &= \iint\limits_U \sqrt{\det(g_{\alpha\beta}(u^1, u^2))}\, du^1\, du^2 \\ &= \iint\limits_U \sqrt{E(u,v)G(u,v) - F^2(u,v)}\, du\, dv. \end{aligned} \tag{4.22.11}$$

8. The principal curvatures are the roots of the *characteristic equation,* $\det(b_{\alpha\beta} - \lambda g_{\alpha\beta}) = 0$, which may be written as $\lambda^2 - b_{\alpha\beta}g^{\alpha\beta}\lambda + b/g = 0$, where $g^{\alpha\beta}$ is the inverse of $g_{\alpha\beta}$, $b = \det(b_{\alpha\beta})$, and $g = \det(g_{\alpha\beta})$. This expands into
$$(EG - F^2)\lambda^2 - (eG - 2fF + gE)\lambda + eg - f^2 = 0. \tag{4.22.12}$$

9. The principal directions $du : dv$ are obtained by solving the homogeneous equation

$$b_{1\alpha}g_{2\beta}\,du^\alpha\,du^\beta - b_{2\alpha}g_{1\beta}\,du^\alpha\,du^\beta = 0, \tag{4.22.13}$$

or

$$(eF - fE)\,du^2 + (eG - gE)\,du\,dv + (fG - gF)\,dv^2 = 0. \tag{4.22.14}$$

10. *Rodrigues formula*: $du : dv$ is a principal direction with principal curvature κ if and only if $d\mathbf{N} + \kappa\,d\mathbf{x} = 0$.

11. A point $\mathbf{x} = \mathbf{f}(u, v)$ on S is an umbilical point if and only if there exists a constant k such that $b_{\alpha\beta}(u, v) = kg_{\alpha\beta}(u, v)$.

12. The principal directions at $\mathbf{x}$ are orthogonal if $\mathbf{x}$ is not an umbilical point.

13. The u- and v-parameter curves at any non-umbilical point $\mathbf{x}$ are tangent to the principal directions if and only if $f(u, v) = F(u, v) = 0$. If $\mathbf{f}$ defines a coordinate patch without umbilical points, the u- and v-parameter curves are lines of curvature if and only if $f = F = 0$.

14. If $f = F = 0$ on a coordinate patch, the principal curvatures are $\kappa_1 = e/E$ and $\kappa_2 = g/G$. It follows that the Gaussian and mean curvatures are

$$K = \frac{eg}{EG}, \quad \text{and} \quad H = \frac{1}{2}\left(\frac{e}{E} + \frac{g}{G}\right). \tag{4.22.15}$$

15. *Gauss equation*: $\mathbf{x}_{\alpha\beta} = \Gamma^\gamma_{\alpha\beta}\mathbf{x}_\gamma + b_{\alpha\beta}\mathbf{N}$.

16. *Weingarten equation*: $\mathbf{N}_\alpha = -b_{\alpha\beta}g^{\beta\gamma}\mathbf{x}_\gamma$.

17. *Gauss–Mainardi–Codazzi equations*: $b_{\alpha\beta}b_{\gamma\delta} - b_{\alpha\gamma}b_{\beta\delta} = R_{\delta\alpha\beta\gamma}$, $b_{\alpha\beta,\gamma} - b_{\alpha\gamma,\beta} + \Gamma^\delta_{\alpha\beta}b_{\delta\gamma} - \Gamma^\delta_{\alpha\gamma}b_{\delta\beta} = 0$, where $R_{\delta\alpha\beta\gamma}$ is the Riemann curvature tensor.

THEOREM 4.22.2 *(Gauss' theorema egregium)*

The Gaussian curvature K depends only on the components of the first fundamental metric $g_{\alpha\beta}$ and their derivatives.

THEOREM 4.22.3 *(Fundamental theorem of surface theory)*

If $g_{\alpha\beta}$ and $b_{\alpha\beta}$ are sufficiently differentiable functions of u and v which satisfy the Gauss–Mainardi–Codazzi equations, $\det(g_{\alpha\beta}) > 0$, $g_{11} > 0$, and $g_{22} > 0$, then a surface exists with $I = g_{\alpha\beta}\,du^\alpha\,du^\beta$ and $II = b_{\alpha\beta}\,du^\alpha\,du^\beta$ as its first and second fundamental forms. This surface is unique up to a congruence.

4.22.2.3 Example: paraboloid of revolution

A Monge patch for a paraboloid of revolution is given by $\mathbf{x} = \mathbf{f}(u, v) = (u, v, u^2 + v^2)$, for $(u, v) \in U = \mathbb{R}^2$. By successive differentiation: $\mathbf{x}_u = (1, 0, 2u)$, $\mathbf{x}_v = (0, 1, 2v)$, $\mathbf{x}_{uu} = (0, 0, 2)$, $\mathbf{x}_{uv} = (0, 0, 0)$, and $\mathbf{x}_{vv} = (0, 0, 2)$.

1. *Unit normal vector*: $\mathbf{N} = (1 + 4u^2 + 4v^2)^{-\frac{1}{2}}(-2u, -2v, 1)$.
2. *First fundamental coefficients*: $E(u, v) = g_{11}(u, v) = 1 + 4u^2$, $F(u, v) = g_{12}(u, v) = 4uv$, $G(u, v) = g_{22}(u, v) = 1 + 4v^2$.
3. *First fundamental form*: $I = (1+4u^2)\,du^2 + 8uv\,du\,dv + (1+4v^2)\,dv^2$. Since $F(u, v) = 0 \Rightarrow u = 0$ or $v = 0$, it follows that the u-parameter curve $v = 0$, is orthogonal to any v-parameter curve, and the v-parameter curve $u = 0$ is orthogonal to any u-parameter curve. Otherwise the u- and v-parameter curves are *not* orthogonal.
4. *Second fundamental coefficients*: $e(u, v) = b_{11}(u, v) = 2(1 + 4u^2 + 4v^2)^{-\frac{1}{2}}$, $f(u, v) = b_{12}(u, v) = 0$, $g(u, v) = b_{22}(u, v) = 2(1 + 4u^2 + 4v^2)^{-\frac{1}{2}}$.
5. *Second fundamental form*: $II = 2(1 + 4u^2 + 4v^2)^{-\frac{1}{2}}(du^2 + dv^2)$.
6. *Classification of points*: $e(u, v)g(u, v) = 4(1+4u^2+4v^2) > 0$ implies that all points on S are elliptic points. The point $(0, 0, 0)$ is the only umbilical point.
7. *Equation for the principal directions*: $uv\,du^2 + (v^2 - u^2)\,du\,dv + uv\,dv^2 = 0$ factors to read $(u\,du + v\,dv)(v\,du - u\,dv) = 0$.
8. *Lines of curvature*: Integrate the differential equations, $u\,dv + v\,dv = 0$, and $v\,du - v\,du = 0$, to obtain, respectively, the equations of the lines of curvature, $u^2 + v^2 = r^2$, and $u/v = \cot\theta$, where r and θ are constant.
9. *Characteristic equation*: $(1 + 4u^2 + 4v^2)\lambda^2 - 4(1 + 2u^2 + 2v^2)\,(1 + 4u^2 + 4v^2)^{-\frac{1}{2}}\lambda + 4(1 + 4u^2 + 4v^2)^{-1} = 0$.
10. *Principal curvatures*: $\kappa_1 = 2(1+4u^2+4v^2)^{-\frac{1}{2}}$ and $\kappa_2 = 2(1+4u^2+4v^2)^{-\frac{3}{2}}$. The paraboloid of revolution may also be represented by $\mathbf{x} = \hat{\mathbf{f}}(r, \theta) = (r\cos\theta.\ r\sin\theta, r^2)$. In this representation the r- and θ-parameter curves are lines of curvature.
11. *Gaussian curvature*: $K = 4(1 + 4u^2 + 4v^2)^{-2}$.
12. *Mean curvature*: $H = 2(1 + 2u^2 + 2v^2)(1 + 4u^2 + 4v^2)^{-\frac{3}{2}}$.

Chapter 5

Analysis

5.1 DIFFERENTIAL CALCULUS

5.1.1 LIMITS

If $\lim_{x\to a} f(x) = A < \infty$ and $\lim_{x\to a} g(x) = B < \infty$ then

1. $\lim_{x\to a} (f(x) \pm g(x)) = A \pm B$
2. $\lim_{x\to a} f(x)g(x) = AB$
3. $\lim_{x\to a} \dfrac{f(x)}{g(x)} = \dfrac{A}{B}$ (if $B \neq 0$)
4. $\lim_{x\to a} [f(x)]^{g(x)} = A^B$ (if $A > 0$)
5. $\lim_{x\to a} h(f(x)) = h(A)$ (if h continuous)
6. If $f(x) \leq g(x)$, then $A \leq B$
7. If $A = B$ and $f(x) \leq h(x) \leq g(x)$, then $\lim_{x\to a} h(x) = A$

EXAMPLES

1. $\lim_{x\to\infty} \left(1 + \dfrac{t}{x}\right)^x = e^t$
2. $\lim_{x\to\infty} x^{1/x} = 1$
3. $\lim_{x\to\infty} \dfrac{(\log x)^p}{x^q} = 0$ (if $q > 0$)
4. $\lim_{x\to 0+} x^p \left|\log x\right|^q = 0$ (if $p > 0$)
5. $\lim_{x\to 0} \dfrac{\sin ax}{x} = a$
6. $\lim_{x\to 0} \dfrac{a^x - 1}{x} = \log a$
7. $\lim_{x\to 0} \dfrac{\log(1+x)}{x} = 1$

5.1.2 DERIVATIVES

The *derivative* of the function $f(x)$, written $f'(x)$, is defined as

$$f'(x) = \lim_{\Delta x\to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x} \tag{5.1.1}$$

if the limit exists. If $y = f(x)$, then $\frac{dy}{dx} = f'(x)$. The n^{th} derivative is

$$y^{(n)} = \frac{dy^{(n-1)}}{dx} = \frac{d}{dx}\left(\frac{d^{n-1}y}{dx^{n-1}}\right) = \frac{d^n y}{dx^n} = f^{(n)}(x).$$

The second and third derivatives are usually written as y'' and y'''. Sometimes the fourth and fifth derivatives are written as $y^{(iv)}$ and $y^{(v)}$.

The partial derivative of $f(x, y)$ with respect to x, written $f_x(x, y)$ or $\frac{\partial f}{\partial x}$, is defined as

$$f_x(x, y) = \lim_{\Delta x\to 0} \frac{f(x+\Delta x, y) - f(x, y)}{\Delta x}. \tag{5.1.2}$$

5.1.3 DERIVATIVE FORMULAS

Let u, v, w be functions of x, and let a, c, and n be constants. Appropriate non-zero values, differentiability, and invertibility are assumed.

(a) $\frac{d}{dx}(a) = 0$

(b) $\frac{d}{dx}(x) = 1$

(c) $\frac{d}{dx}(au) = a\frac{du}{dx}$

(d) $\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}$

(e) $\frac{d}{dx}(uv) = v\frac{du}{dx} + u\frac{dv}{dx}$

(f) $\frac{d}{dx}(uvw) = uv\frac{dw}{dx} + uw\frac{dv}{dx} + vw\frac{du}{dx}$

(g) $\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{1}{v}\frac{du}{dx} - \frac{u}{v^2}\frac{dv}{dx} = \frac{v(du/dx) - u(dv/dx)}{v^2}$

(h) $\frac{d}{dx}(u^n) = nu^{n-1}\frac{du}{dx}$

(i) $\frac{d}{dx}(u^v) = vu^{v-1}\frac{du}{dx} + (\log_e u)u^v\frac{dv}{dx}$

(j) $\frac{d}{dx}\left(\sqrt{u}\right) = \frac{1}{2\sqrt{u}}\frac{du}{dx}$

(k) $\frac{d}{dx}(\log_e u) = \frac{1}{u}\frac{du}{dx}$

(l) $\frac{d}{dx}(\log_a u) = (\log_a e)\frac{1}{u}\frac{du}{dx}$

(m) $\frac{d}{dx}\left(\frac{1}{u}\right) = -\frac{1}{u^2}\frac{du}{dx}$

(n) $\frac{d}{dx}\left(\frac{1}{u^n}\right) = -\frac{n}{u^{n+1}}\frac{du}{dx}$

(o) $\frac{d}{dx}\left(\frac{u^n}{v^m}\right) = \frac{u^{n-1}}{v^{m+1}}\left(nv\frac{du}{dx} - mu\frac{dv}{dx}\right)$

(p) $\frac{d}{dx}(u^n v^m) = u^{n-1}v^{m-1}\left(nv\frac{du}{dx} + mu\frac{dv}{dx}\right)$

(q) $\frac{d}{dx}(f(u)) = \frac{df}{du}\cdot\frac{du}{dx}$

(r) $\frac{d^2}{dx^2}(f(u)) = \frac{df}{du}\cdot\frac{d^2u}{dx^2} + \frac{d^2f}{du^2}\cdot\left(\frac{du}{dx}\right)^2$

(s) $\frac{d^3}{dx^3}(f(u)) = \frac{df}{du}\cdot\frac{d^3u}{dx^3} + 3\frac{d^2f}{du^2}\cdot\frac{du}{dx}\cdot\frac{d^2u}{dx^2} + \frac{d^3f}{du^3}\cdot\left(\frac{du}{dx}\right)^3$

(t) $\frac{d^n}{dx^n}(uv) = \binom{n}{0}v\frac{d^nu}{dx^n} + \binom{n}{1}\frac{dv}{dx}\frac{d^{n-1}u}{dx^{n-1}} + \cdots + \binom{n}{n}\frac{d^nv}{dx^n}u$

(u) $\frac{d}{dx}\int_c^x f(t)\,dt = f(x)$

(v) $\frac{d}{dx}\int_x^c f(t)\,dt = -f(x)$

(w) $\frac{dx}{dy} = \left(\frac{dy}{dx}\right)^{-1}$ and $\frac{d^2x}{dy^2} = -\frac{d^2y}{dx^2}\Big/\left(\frac{dy}{dx}\right)^3$

(x) If $F(x,y) = 0$, then $\frac{dy}{dx} = -\frac{F_x}{F_y}$ and

$$\frac{d^2y}{dx^2} = -\frac{\left(F_{xx}F_y^2 - 2F_{xy}F_xF_y + F_{yy}F_x^2\right)}{F_y^3}$$

(y) Leibniz's rule gives the derivative of an integral:

$$\frac{d}{dx}\left(\int_{f(x)}^{g(x)} h(x,t)\,dt\right) = g'(x)h(x,g(x)) - f'(x)h(x,f(x)) + \int_{f(x)}^{g(x)} \frac{\partial h}{\partial x}(x,t)\,dt.$$

(z) If $x = x(t)$ and $y = y(t)$ then (the dots denote differentiation with respect to t):

$$\frac{dy}{dx} = \frac{\dot{y}(t)}{\dot{x}(t)}, \qquad \frac{d^2y}{dx^2} = \frac{\dot{x}\ddot{y} - \ddot{x}\dot{y}}{(\dot{x})^3}.$$

5.1.4 DERIVATIVES OF COMMON FUNCTIONS

Let a be a constant.

$f(x)$	$f'(x)$	$f(x)$	$f'(x)$	$f(x)$	$f'(x)$
$\sin x$	$\cos x$	$\sinh x$	$\cosh x$	x^a	ax^{a-1}
$\cos x$	$-\sin x$	$\cosh x$	$\sinh x$	$\frac{1}{x^a}$	$-\frac{a}{x^{a+1}}$
$\tan x$	$\sec^2 x$	$\tanh x$	$\operatorname{sech}^2 x$	$\sqrt{x}$	$\frac{1}{2\sqrt{x}}$
$\csc x$	$-\csc x \cot x$	$\operatorname{csch} x$	$-\operatorname{csch} x \coth x$	$\ln\|x\|$	$1/x$
$\sec x$	$\sec x \tan x$	$\operatorname{sech} x$	$-\operatorname{sech} x \tanh x$	e^x	e^x
$\cot x$	$-\csc^2 x$	$\coth x$	$-\operatorname{csch}^2 x$	a^x $(a > 0)$	$a^x \ln a$
$\sin^{-1} x$	$\frac{1}{\sqrt{1-x^2}}$	$\sinh^{-1} x$	$\frac{1}{\sqrt{x^2+1}}$	$\|x\|$	$x/\|x\|$
$\cos^{-1} x$	$-\frac{1}{\sqrt{1-x^2}}$	$\cosh^{-1} x$	$\frac{1}{\sqrt{x^2-1}}$		
$\tan^{-1} x$	$\frac{1}{1+x^2}$	$\tanh^{-1} x$	$\frac{1}{1-x^2}$		
$\csc^{-1} x$	$-\frac{1}{x\sqrt{x^2-1}}$	$\operatorname{csch}^{-1} x$	$\frac{-1}{\|x\|\sqrt{1+x^2}}$		
$\sec^{-1} x$	$\frac{1}{x\sqrt{x^2-1}}$	$\operatorname{sech}^{-1} x$	$\frac{-1}{\|x\|\sqrt{1-x^2}}$		
$\cot^{-1} x$	$-\frac{1}{1+x^2}$	$\coth^{-1} x$	$\frac{1}{1-x^2}$		

5.1.5 DERIVATIVE THEOREMS

1. *Fundamental theorem of calculus*: Suppose f is continuous on $[a,b]$.
   (a) If G is defined as $G(x) = \int_a^x f(t)\,dt$ for all x in $[a,b]$, then G is an anti-derivative of f on $[a,b]$.
   (b) If F is any anti-derivative of f, then $\int_a^b f(t)\,dt = F(b) - F(a)$.
2. *Intermediate value theorem*: If $f(x)$ is continuous on $[a,b]$ and if $f(a) \neq f(b)$, then f takes on every value between $f(a)$ and $f(b)$ in the interval (a,b).
3. *Rolle's theorem*: If $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b), and if $f(a) = f(b)$, then $f'(c) = 0$ for at least one number c in (a,b).
4. *Mean value theorem*: If $f(x)$ is continuous on $[a,b]$ and differentiable on (a,b), then a number c exists in (a,b) such that $f(b) - f(a) = (b-a)f'(c)$.

5.1.6 THE TWO-DIMENSIONAL CHAIN RULE

If $x = x(t)$, $y = y(t)$, and $z = z(x, y)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}, \qquad \text{and}$$

$$\begin{aligned}\frac{d^2z}{dt^2} &= \frac{\partial z}{\partial x}\frac{d^2x}{dt^2} + \frac{dx}{dt}\left(\frac{\partial^2 z}{\partial x^2}\frac{dx}{dt} + \frac{\partial^2 z}{\partial x \partial y}\frac{dy}{dt}\right) \\ &\quad + \frac{\partial z}{\partial y}\frac{d^2y}{dt^2} + \frac{dy}{dt}\left(\frac{\partial^2 z}{\partial y^2}\frac{dy}{dt} + \frac{\partial^2 z}{\partial x \partial y}\frac{dx}{dt}\right).\end{aligned} \tag{5.1.3}$$

If $x = x(u, v)$, $y = y(u, v)$, and $z = z(x, y)$, then

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u} \qquad \text{and} \qquad \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v}. \tag{5.1.4}$$

If $u = u(x, y)$, $v = v(x, y)$, and $f = f(x, y)$, then the partial derivative of f with respect to u, holding v constant, written $\left(\frac{\partial f}{\partial u}\right)_v$, can be expressed as

$$\left(\frac{\partial f}{\partial u}\right)_v = \left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\partial x}{\partial u}\right)_v + \left(\frac{\partial f}{\partial y}\right)_x \left(\frac{\partial y}{\partial u}\right)_v. \tag{5.1.5}$$

5.1.7 THE CYCLIC RULE

If x, y, and z all depend on one another (say, through $f(x, y, z) = 0$), then

$$\left(\frac{\partial x}{\partial y}\right)_z = \left[\left(\frac{\partial y}{\partial x}\right)_z\right]^{-1} = -\frac{(\partial f/\partial y)_{x,z}}{(\partial f/\partial x)_{y,z}}, \tag{5.1.6}$$

and we find (this is the *cyclic rule*)

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1. \tag{5.1.7}$$

EXAMPLES

1. If $x + y + z = 3$ then

$$\begin{aligned} \left(\frac{\partial}{\partial y}\right)_z (x + y + z = 3) &\Longrightarrow \left(\frac{\partial x}{\partial y}\right)_z + 1 = 0 \quad \text{or} \quad \left(\frac{\partial x}{\partial y}\right)_z = -1, \\ \left(\frac{\partial}{\partial z}\right)_x (x + y + z = 3) &\Longrightarrow \left(\frac{\partial y}{\partial z}\right)_x + 1 = 0 \quad \text{or} \quad \left(\frac{\partial y}{\partial z}\right)_x = -1, \\ \left(\frac{\partial}{\partial x}\right)_y (x + y + z = 3) &\Longrightarrow \left(\frac{\partial z}{\partial x}\right)_y + 1 = 0 \quad \text{or} \quad \left(\frac{\partial z}{\partial x}\right)_y = -1 \end{aligned}$$

2. If $x^2 + y^3 z = 5$ then

$$\begin{aligned} \left(\frac{\partial}{\partial y}\right)_z (x^2 + y^3 z = 5) &\Longrightarrow 2x\left(\frac{\partial x}{\partial y}\right)_z + 3y^2 z = 0 \quad \text{or} \quad \left(\frac{\partial x}{\partial y}\right)_z = -\frac{3y^2 z}{2x}, \\ \left(\frac{\partial}{\partial z}\right)_x (x^2 + y^3 z = 5) &\Longrightarrow 3y^2\left(\frac{\partial y}{\partial z}\right)_x z + y^3 = 0 \quad \text{or} \quad \left(\frac{\partial y}{\partial z}\right)_x = -\frac{y}{3z}, \\ \left(\frac{\partial}{\partial x}\right)_y (x^2 + y^3 z = 5) &\Longrightarrow 2x + y^3\left(\frac{\partial z}{\partial x}\right)_y = 0 \quad \text{or} \quad \left(\frac{\partial z}{\partial x}\right)_y = -\frac{2x}{y^3} \end{aligned}$$

5.1.8 MAXIMA AND MINIMA OF FUNCTIONS

1. If a function $f(x)$ has a local extremum at a number c, then either $f'(c) = 0$ or $f'(c)$ does not exist.
2. If $f'(c) = 0$, $f(x)$ and $f'(x)$ are differentiable on an open interval containing c, and
   (a) if $f''(c) < 0$, then f has a local maximum at c;
   (b) if $f''(c) > 0$, then f has a local minimum at c.

5.1.8.1 Two variables

Suppose that extreme points of $f(x, y)$ are desired. The stationary points are the solutions to

$$f_x(x, y) = 0, \qquad f_y(x, y) = 0. \tag{5.1.8}$$

For each stationary point we compute

$$\Delta = f_{xx} f_{yy} - f_{xy}^2 \tag{5.1.9}$$

to determine the type of stationary point.

1. If $\Delta > 0$ and $f_{xx} > 0$, then the stationary point is a local minimum.
2. If $\Delta > 0$ and $f_{xx} < 0$, then the stationary point is a local maximum.
3. If $\Delta < 0$, then the stationary point cannot be an extermum.

5.1.8.2 Three variables

Suppose that extreme points of $\Phi(x, y, z)$ are desired. The stationary points are the solutions to

$$\Phi_x(x, y, z) = 0, \qquad \Phi_y(x, y, z) = 0, \qquad \Phi_z(x, y, z) = 0. \tag{5.1.10}$$

For each stationary point we compute

$$\Delta_1 = \Phi_{xx}, \qquad \Delta_2 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} \\ \Phi_{xy} & \Phi_{yy} \end{vmatrix} \qquad \Delta_3 = \begin{vmatrix} \Phi_{xx} & \Phi_{xy} & \Phi_{xz} \\ \Phi_{xy} & \Phi_{yy} & \Phi_{yz} \\ \Phi_{xz} & \Phi_{yz} & \Phi_{zz} \end{vmatrix}. \tag{5.1.11}$$

to determine the type of stationary point.

1. If $\Delta_1 > 0$, $\Delta_2 > 0$, and $\Delta_3 > 0$, then the stationary point is a local minimum.
2. If $\Delta_1 < 0$, $\Delta_2 > 0$, and $\Delta_3 < 0$, then the stationary point is a local maximum.

5.1.8.3 Lagrange multipliers

To extremize the function $f(x_1, x_2, \ldots, x_n) = f(\mathbf{x})$ subject to the m side constraints $\mathbf{g}(\mathbf{x}) = 0$, introduce an m-dimensional vector of Lagrange multipliers λ and define $F(\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda^{\mathrm{T}} \mathbf{g}(\mathbf{x})$. Then extremize F with respect to all of its arguments:

$$\frac{\partial F}{\partial x_i} = \frac{\partial f}{\partial x_i} + \lambda^{\mathrm{T}} \frac{\partial \mathbf{g}}{\partial x_i} = 0, \qquad \text{and} \qquad \frac{\partial F}{\partial \lambda_j} = g_j = 0. \tag{5.1.12}$$

For the case of extremizing $f(x, y)$ subject to $g(x, y) = 0$ (i.e., $n = 2$ and $m = 1$):

$$f_x + \lambda g_x = 0, \qquad f_y + \lambda g_y = 0, \quad \text{and} \quad g = 0. \tag{5.1.13}$$

EXAMPLE Find points on the unit circle (given by $g(x,y) = (x-1)^2+(y-2)^2-1 = 0$) that are the closest to and furthest from the origin (the distance squared from the origin is $f(x,y) = x^2 + y^2$). This is solved by creating, and solving, the three (non-linear) algebraic equations:

$$2x + 2\lambda(x-1) = 0, \qquad 2y + 2\lambda(y-2) = 0, \qquad (x-1)^2 + (y-2)^2 = 1.$$

The solutions are $\left\{x = 1 + 1/\sqrt{5}, y = 2 + 2/\sqrt{5}, \lambda = -1 - \sqrt{5}\right\}$ (furthest), and $\left\{x = 1 - 1/\sqrt{5}, y = 2 - 2/\sqrt{5}, \lambda = -1 + \sqrt{5}\right\}$ (closest).

5.1.9 L'HÔPITAL'S RULE

If $f(x)$ and $g(x)$ are differentiable in a punctured neighborhood of point a, and if $f(x)$ and $g(x)$ both tend to 0 or ∞ as $x \to a$, then

$$\lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)} \tag{5.1.14}$$

if the right-hand side exists. Sometimes, by manipulating functions, this rule can determine the value of the indeterminate forms $\{1^\infty, 0^0, \infty^0, 0\cdot\infty, \infty-\infty\}$. The rule can also be applied repeatedly.

EXAMPLES

- $\displaystyle \lim_{x\to 0} \frac{x - \sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2} = \lim_{x\to 0} \frac{\sin x}{6x} = \lim_{x\to 0} \frac{\cos x}{6} = \frac{1}{6}$
- $\displaystyle \lim_{x\to\infty} \frac{x^n}{e^x} = \lim_{x\to\infty} \frac{nx^{n-1}}{e^x} = \lim_{x\to\infty} \frac{n(n-1)x^{n-2}}{e^x} = \cdots = \lim_{x\to\infty} \frac{n!}{e^x} = 0$
- $\displaystyle \lim_{x\to 1} \left(\frac{x}{x-1} - \frac{1}{\ln x}\right) = \lim_{x\to 1} \frac{x\ln x - x + 1}{(x-1)\ln x} = \lim_{x\to 1} \frac{\ln x}{\frac{x-1}{x} + \ln x}$
  $\displaystyle = \lim_{x\to 1} \frac{x\ln x}{x - 1 + x\ln x} = \lim_{x\to 1} \frac{1+\ln x}{2+\ln x} = \frac{1}{2}$

5.1.10 COMMON LIMITS

Assume $a > 0$ in the following.

1. $\displaystyle \lim_{x\to 0} \frac{e^x - 1}{x} = 1$
2. $\displaystyle \lim_{x\to 0} \frac{a^x - 1}{x} = \ln a$
3. $\displaystyle \lim_{x\to 0} \frac{\sinh x}{x} = 1$
4. $\displaystyle \lim_{x\to 0} \frac{\sin x}{x} = 1$
5. $\displaystyle \lim_{x\to 0} \frac{\tanh x}{x} = 1$
6. $\displaystyle \lim_{x\to 0} \frac{\tan x}{x} = 1$
7. $\displaystyle \lim_{x\to 0} \frac{1-\cos x}{x} = \frac{1}{2}$
8. $\displaystyle \lim_{x\to 0} \frac{\ln(1+x)}{x} = 1$
9. $\displaystyle \lim_{x\to 0} \frac{\log_a(1+x)}{x} = \log_a e$
10. $\displaystyle \lim_{x\to 0+} x^x = 1$
11. $\displaystyle \lim_{x\to 0+} x^a \ln x = 0$
12. $\displaystyle \lim_{x\to +\infty} x^{-a} \ln x = 0$

5.1.11 VECTOR CALCULUS

1. Definitions of "div," "grad," and "curl" are on page 388.

2. A vector field $\mathbf{F}$ is *irrotational* if $\nabla \times \mathbf{F} = 0$.

3. A vector field $\mathbf{F}$ is *solenoidal* if $\nabla \cdot \mathbf{F} = 0$.

4. In Cartesian coordinates, $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right) = \mathbf{i}\frac{\partial}{\partial x} + \mathbf{j}\frac{\partial}{\partial y} + \mathbf{k}\frac{\partial}{\partial z}$.

5. $\nabla \cdot u = \sum_{i=1}^{3} \frac{\partial u}{\partial x_i}$ is a scalar while $\mathbf{F} \cdot \nabla = \sum_{i=1}^{3} F_i \frac{\partial}{\partial x_i}$ is an operator.

6. If u and v are scalars and $\mathbf{F}$ and $\mathbf{G}$ are vectors in $\mathbb{R}^3$, then

   (a) $\nabla(u + v) = \nabla u + \nabla v$

   (b) $\nabla(uv) = u\nabla v + v\nabla u$

   (c) $\nabla(\mathbf{F} + \mathbf{G}) = \nabla\mathbf{F} + \nabla\mathbf{G}$

   (d) $\nabla(\mathbf{F} \cdot \mathbf{G}) = (\mathbf{F} \cdot \nabla)\mathbf{G} + (\mathbf{G} \cdot \nabla)\mathbf{F} + \mathbf{F} \times (\nabla \times \mathbf{G}) + \mathbf{G} \times (\nabla \times \mathbf{F})$

   (e) $\nabla \cdot (u\mathbf{F}) = u(\nabla \cdot \mathbf{F}) + \mathbf{F} \cdot \nabla u$

   (f) $\nabla \cdot (\mathbf{F} \times \mathbf{G}) = \mathbf{G} \cdot (\nabla \times \mathbf{F}) - \mathbf{F} \cdot (\nabla \times \mathbf{G})$

   (g) $\nabla \times (u\mathbf{F}) = u(\nabla \times \mathbf{F}) + (\nabla u) \times \mathbf{F}$

   (h) $\nabla \times (\mathbf{F} + \mathbf{G}) = \nabla \times \mathbf{F} + \nabla \times \mathbf{G}$

   (i) $\nabla \times (\mathbf{F} \times \mathbf{G}) = \mathbf{F}(\nabla \cdot \mathbf{G}) - \mathbf{G}(\nabla \cdot \mathbf{F}) + (\mathbf{G} \cdot \nabla)\mathbf{F} - (\mathbf{F} \cdot \nabla)\mathbf{G}$

   (j) $\mathbf{F} \cdot \frac{d\mathbf{F}}{dt} = |\mathbf{F}| \frac{d|\mathbf{F}|}{dt}$

   (k) $\nabla \times (\nabla \times \mathbf{F}) = \nabla(\nabla \cdot \mathbf{F}) - \nabla^2\mathbf{F}$

   (l) $\nabla \times (\nabla u) = \mathbf{0}$

   (m) $\nabla \cdot (\nabla \times \mathbf{F}) = 0$

   (n) $\nabla^2(uv) = u\nabla^2 v + 2(\nabla u) \cdot (\nabla v) + v\nabla^2 u$

7. If $r = |\mathbf{r}|$, $\mathbf{a}$ is a constant vector, and n is an integer, then

Φ	$\nabla\Phi$	$\nabla^2\Phi$
$\mathbf{a} \cdot \mathbf{r}$	$\mathbf{a}$	0
r^n	$nr^{n-2}\mathbf{r}$	$n(n+1)r^{n-2}$
$\log r$	$\mathbf{r}/r^2$	$1/r^2$

$\mathbf{F}$	$\nabla \cdot \mathbf{F}$	$\nabla \times \mathbf{F}$	$(\mathbf{G} \cdot \nabla)\mathbf{F}$
$\mathbf{r}$	3	0	$\mathbf{G}$
$\mathbf{a} \times \mathbf{r}$	0	$2\mathbf{a}$	$\mathbf{a} \times \mathbf{G}$
$\mathbf{a}r^n$	$nr^{n-2}(\mathbf{r} \cdot \mathbf{a})$	$nr^{n-2}(\mathbf{r} \times \mathbf{a})$	$nr^{n-2}(\mathbf{r} \cdot \mathbf{G})\mathbf{a}$
$\mathbf{r}r^n$	$(n+3)r^n$	0	$r^n\mathbf{G} + nr^{n-2}(\mathbf{r} \cdot \mathbf{G})\mathbf{r}$
$\mathbf{a}\log r$	$\mathbf{r} \cdot \mathbf{a}/r^2$	$\mathbf{r} \times \mathbf{a}/r^2$	$(\mathbf{G} \cdot \mathbf{r})\mathbf{a}/r^2$

$\mathbf{F}$	$\nabla^2\mathbf{F}$	$\nabla\nabla\cdot\mathbf{F}$
$\mathbf{r}$	0	0
$\mathbf{a}\times\mathbf{r}$	0	0
$\mathbf{a}r^n$	$n(n+1)r^{n-2}\mathbf{a}$	$nr^{n-2}\mathbf{a}+n(n-2)r^{n-4}(\mathbf{r}\cdot\mathbf{a})\mathbf{r}$
$\mathbf{r}r^n$	$n(n+3)r^{n-2}\mathbf{r}$	$n(n+3)r^{n-2}\mathbf{r}$
$\mathbf{a}\log r$	$\mathbf{a}/r^2$	$[r^2\mathbf{a}-2(\mathbf{r}\cdot\mathbf{a})\mathbf{r}]/r^4$

8. $\dfrac{d}{dt}(\mathbf{F}+\mathbf{G}) = \dfrac{d\mathbf{F}}{dt}+\dfrac{d\mathbf{G}}{dt}$

9. $\dfrac{d}{dt}(\mathbf{F}\cdot\mathbf{G}) = \mathbf{F}\cdot\dfrac{d\mathbf{G}}{dt}+\dfrac{d\mathbf{F}}{dt}\cdot\mathbf{G}$

10. $\dfrac{d}{dt}(\mathbf{F}\times\mathbf{G}) = \mathbf{F}\times\dfrac{d\mathbf{G}}{dt}+\dfrac{d\mathbf{F}}{dt}\times\mathbf{G}$

11. $\dfrac{d}{dt}(\mathbf{V}_1\times\mathbf{V}_2\times\mathbf{V}_3) = \left(\dfrac{d\mathbf{V}_1}{dt}\right)\times(\mathbf{V}_2\times\mathbf{V}_3)+\mathbf{V}_1\times\left(\left(\dfrac{d\mathbf{V}_2}{dt}\right)\times\mathbf{V}_3\right)+\mathbf{V}_1\times\left(\mathbf{V}_2\times\left(\dfrac{d\mathbf{V}_3}{dt}\right)\right)$

12. $\dfrac{d}{dt}[\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3] = \left[\left(\dfrac{d\mathbf{V}_1}{dt}\right)\mathbf{V}_2\mathbf{V}_3\right]+\left[\mathbf{V}_1\left(\dfrac{d\mathbf{V}_2}{dt}\right)\mathbf{V}_3\right]+\left[\mathbf{V}_1\mathbf{V}_2\left(\dfrac{d\mathbf{V}_3}{dt}\right)\right]$,
where $[\mathbf{V}_1\mathbf{V}_2\mathbf{V}_3]=\mathbf{V}_1\cdot(\mathbf{V}_2\times\mathbf{V}_3)$ is the scalar triple product (see page 82).

13. If $A=A(t)$ and $B=B(t)$ are matrices then

    (a) $\dfrac{d(AB)}{dt} = \dfrac{dA}{dt}B + A\dfrac{dB}{dt}$

    (b) $\dfrac{d(A\otimes B)}{dt} = \dfrac{dA}{dt}\otimes B + A\otimes\dfrac{dB}{dt}$

    (c) $\dfrac{d(A^{-1})}{dt} = -A^{-1}\dfrac{dA}{dt}A^{-1}$

5.1.12 MATRIX AND VECTOR DERIVATIVES

5.1.12.1 Definitions

1. The derivative of the row vector $\mathbf{y}=\begin{bmatrix}y_1 & y_2 & \dots & y_m\end{bmatrix}$ with respect to the scalar x is

$$\frac{\partial\mathbf{y}}{\partial x} = \begin{bmatrix}\frac{\partial y_1}{\partial x} & \frac{\partial y_2}{\partial x} & \dots & \frac{\partial y_m}{\partial x}\end{bmatrix}. \tag{5.1.15}$$

2. The derivative of a scalar y with respect to the vector $\mathbf{x}$ is the column vector

$$\frac{\partial y}{\partial\mathbf{x}} = \begin{bmatrix}\frac{\partial y}{\partial x_1} & \frac{\partial y}{\partial x_2} & \dots & \frac{\partial y}{\partial x_n}\end{bmatrix}^{\mathrm{T}}. \tag{5.1.16}$$

3. Let $\mathbf{x}$ be a $n \times 1$ vector and let $\mathbf{y}$ be a $m \times 1$ vector. The derivative of $\mathbf{y}$ with respect to $\mathbf{x}$ is the matrix

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_2}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_1} \\ \frac{\partial y_1}{\partial x_2} & \frac{\partial y_2}{\partial x_2} & \cdots & \frac{\partial y_m}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y_1}{\partial x_n} & \frac{\partial y_2}{\partial x_n} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix} \tag{5.1.17}$$

In multivariate analysis, if $\mathbf{x}$ and $\mathbf{y}$ have the same length, then the absolute value of the determinant of $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$ is called the *Jacobian* of the transformation determined by $\mathbf{y} = \mathbf{y}(\mathbf{x})$; written $\frac{\partial(y_1,y_2,\ldots,y_n)}{\partial(x_1,x_2,\ldots,x_n)}$.

4. The Jacobian of the derivatives $\frac{\partial \phi}{\partial x_1}, \frac{\partial \phi}{\partial x_2}, \ldots \frac{\partial \phi}{\partial x_n}$ of the function $\phi(x_1, \ldots . x_n)$ with respect to $x_1, \ldots, x_n$ is called the *Hessian* H of ϕ:

$$H = \begin{vmatrix} \frac{\partial^2 \phi}{\partial x_1^2} & \frac{\partial^2 \phi}{\partial x_1\, \partial x_2} & \frac{\partial^2 \phi}{\partial x_1\, \partial x_3} & \cdots & \frac{\partial^2 \phi}{\partial x_1\, \partial x_n} \\ \frac{\partial^2 \phi}{\partial x_2\, \partial x_1} & \frac{\partial^2 \phi}{\partial x_2^2} & \frac{\partial^2 \phi}{\partial x_2\, \partial x_3} & \cdots & \frac{\partial^2 \phi}{\partial x_2\, \partial x_n} \\ \vdots & \vdots & \vdots & & \vdots \\ \frac{\partial^2 \phi}{\partial x_n\, \partial x_1} & \frac{\partial^2 \phi}{\partial x_n\, \partial x_2} & \frac{\partial^2 \phi}{\partial x_n\, \partial x_3} & \cdots & \frac{\partial^2 \phi}{\partial x_n^2} \end{vmatrix}$$

5. The derivative of the matrix $A(t) = (a_{ij}(t))$, with respect to the scalar t, is the matrix $\frac{dA(t)}{dt} = \left(\frac{da_{ij}(t)}{dt}\right)$.

6. If $X = (x_{ij})$ is a $m \times n$ matrix and if y is a scalar function of X, then the derivative of y with respect to X is (here, $E_{ij} = \mathbf{e}_i \mathbf{e}_j^{\mathrm{T}}$):

$$\frac{\partial y}{\partial X} = \begin{bmatrix} \frac{\partial y}{\partial x_{11}} & \frac{\partial y}{\partial x_{12}} & \cdots & \frac{\partial y}{\partial x_{1n}} \\ \frac{\partial y}{\partial x_{21}} & \frac{\partial y}{\partial x_{22}} & \cdots & \frac{\partial y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial y}{\partial x_{m1}} & \frac{\partial y}{\partial x_{m2}} & \cdots & \frac{\partial y}{\partial x_{mn}} \end{bmatrix} = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} E_{ij} \frac{\partial y}{\partial x_{ij}}. \tag{5.1.18}$$

7. If $Y = (y_{ij})$ is a $p \times q$ matrix and X is a $m \times n$ matrix, then the derivative of Y with respect to X is

$$\frac{\partial Y}{\partial X} = \begin{bmatrix} \frac{\partial Y}{\partial x_{11}} & \frac{\partial Y}{\partial x_{12}} & \cdots & \frac{\partial Y}{\partial x_{1n}} \\ \frac{\partial Y}{\partial x_{21}} & \frac{\partial Y}{\partial x_{22}} & \cdots & \frac{\partial Y}{\partial x_{2n}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial Y}{\partial x_{m1}} & \frac{\partial Y}{\partial x_{m2}} & \cdots & \frac{\partial Y}{\partial x_{mn}} \end{bmatrix} = \sum_{\substack{1 \le i \le m \\ 1 \le j \le n}} E_{ij} \otimes \frac{\partial Y}{\partial x_{ij}}. \tag{5.1.19}$$

5.1.12.2 Properties

1. If $Y = AX^{-1}B$, then
   (a) $\frac{\partial Y}{\partial x_{rs}} = -AX^{-1}E_{rs}X^{-1}B$.
   (b) $\frac{\partial y_{ij}}{\partial X} = -(X^{-1})^{\mathrm{T}} A^{\mathrm{T}} E_{ij} B^{\mathrm{T}} (X^{-1})^{\mathrm{T}}$.

2. If $Y = AXB$ then $\frac{\partial Y}{\partial x_{ij}} = AE_{ij}B$ where $E_{ij} = \mathbf{e}_i\mathbf{e}_j^{\mathrm{T}}$ has the same size as X.
3. If $Y = AX^{\mathrm{T}}B$, then $\frac{\partial y_{ij}}{\partial X} = BE_{ij}^{\mathrm{T}}A$.
4. If $Y = X^{\mathrm{T}}AX$, then
   (a) $\frac{\partial Y}{\partial x_{rs}} = E_{rs}^{\mathrm{T}}AX + X^{\mathrm{T}}AE_{rs}$.
   (b) $\frac{\partial y_{ij}}{\partial X} = AXE_{ij}^{\mathrm{T}} + A^{\mathrm{T}}XE_{ij}$.
5. If $\mathbf{y} = \operatorname{Vec} Y$ and $\mathbf{x} = \operatorname{Vec} X$ (see page 103), then
   (a) If $Y = AX$, then $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = I \otimes A^{\mathrm{T}}$.
   (b) If $Y = XA$, then $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = A \otimes I$.
   (c) If $Y = AX^{-1}B$, then $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = -(X^{-1}B) \otimes (X^{-1})^{\mathrm{T}}A^{\mathrm{T}}$.
6. The derivative of the determinant of a matrix can be written:
   (a) If Y_{ij} is the cofactor of element y_{ij} in $|Y|$, then $\frac{\partial |Y|}{\partial x_{rs}} = \sum_i \sum_j Y_{ij} \frac{\partial y_{ij}}{\partial x_{rs}}$.
   (b) If all of the components (x_{ij}) of X are independent, then $\frac{\partial |X|}{\partial X} = |X|\,(X^{-1})^{\mathrm{T}}$.
7. Derivatives of powers of matrices are obtained as follows:
   (a) If $Y = X^r$, then $\frac{\partial Y}{\partial x_{rs}} = \sum_{k=0}^{r-1} X^k E_{rs} X^{n-k-1}$.
   (b) If $Y = X^{-r}$, then $\frac{\partial Y}{\partial x_{rs}} = -X^{-r}\left(\sum_{k=0}^{r-1} X^k E_{rs} X^{n-k-1}\right) X^{-r}$.
   (c) The n^{th} derivative of the r^{th} power of the matrix A^{-1}, in terms of derivatives of the matrix A, is

$$\frac{d^n A^{-r}}{dx^n} = n! \left(\sum_{k=1}^{n} (-1)^k \frac{\mathcal{P}_{i_1}}{i_1!} \frac{\mathcal{P}_{i_2}}{i_2!} \cdots \frac{\mathcal{P}_{i_k}}{i_k!} \right) A^{-r} \tag{5.1.20}$$

   where $\mathcal{P}_i = A^{-r} \frac{d^i A^r}{dx^i}$ and the summation is taken over all positive integers $(i_1, i_2, \ldots, i_k)$, distinct or otherwise, such that $\sum_{m=1}^{k} i_m = n$. Setting $n = r = 1$ results in

$$\frac{dA^{-1}}{dx} = -A^{-1} \frac{dA}{dx} A^{-1}. \tag{5.1.21}$$

8. Derivative formulas:
   (a) If $\mathbf{z} = \mathbf{z}(\mathbf{y}(\mathbf{x}))$, then $\dfrac{\partial \mathbf{z}}{\partial \mathbf{y}} = \dfrac{\partial \mathbf{z}}{\partial \mathbf{x}} \dfrac{\partial \mathbf{x}}{\partial \mathbf{y}}$.
   (b) If X and Y are matrices, then $\left(\frac{\partial Y}{\partial X}\right)^{\mathrm{T}} = \frac{\partial Y^{\mathrm{T}}}{\partial X^{\mathrm{T}}}$.
   (c) If X, Y, and Z are matrices of size $m \times n$, $n \times v$, and $p \times q$, then $\frac{\partial (XY)}{\partial Z} = \frac{\partial X}{\partial Z}(I_q \otimes Y) + (I_p \otimes X)\frac{\partial Y}{\partial Z}$.

9. Differentiation with respect to a vector $\mathbf{x}$. (Here A is constant.) $\dfrac{\partial A\mathbf{x}}{\partial \mathbf{x}^{\mathrm{T}}} = A$ and

y (a scalar or a vector)	$\frac{\partial y}{\partial \mathbf{x}}$ (where $\mathbf{x}$ is a vector)
$\mathbf{x}^{\mathrm{T}}$	I
$A\mathbf{x}$	A^{T}
$\mathbf{x}^{\mathrm{T}}A$	A
$\mathbf{x}^{\mathrm{T}}\mathbf{x}$	$2\mathbf{x}$
$\mathbf{x}^{\mathrm{T}}A\mathbf{x}$	$A\mathbf{x} + A^{\mathrm{T}}\mathbf{x}$

10. Differentiation with respect to a matrix X. (Here $\{\mathbf{a}, \mathbf{b}, A, B\}$ are constants, $Y = Y(X)$, and $Z = Z(X)$.)

y (a scalar, vector, or matrix)	$\frac{\partial y}{\partial X}$ (where X is a matrix)
YZ	$Y\frac{dZ}{dX} + \frac{dY}{dX}Z$
AXB	$A^{\mathrm{T}}B^{\mathrm{T}}$
$\mathbf{a}^{\mathrm{T}}X^{\mathrm{T}}X\mathbf{b}$	$X(\mathbf{a}\mathbf{b}^{\mathrm{T}} + \mathbf{b}\mathbf{a}^{\mathrm{T}})$
$\mathbf{a}^{\mathrm{T}}X^{\mathrm{T}}X\mathbf{a}$	$2X\mathbf{a}\mathbf{a}^{\mathrm{T}}$
$\mathbf{a}^{\mathrm{T}}X^{\mathrm{T}}CX\mathbf{b}$	$C^{\mathrm{T}}X\mathbf{a}\mathbf{b}^{\mathrm{T}} + CX\mathbf{b}\mathbf{a}^{\mathrm{T}}$
$\mathbf{a}^{\mathrm{T}}X^{\mathrm{T}}CX\mathbf{a}$	$(C + C^{\mathrm{T}})X\mathbf{a}\mathbf{a}^{\mathrm{T}}$
$(X\mathbf{a} + \mathbf{b})^{\mathrm{T}}C(X\mathbf{a} + \mathbf{b})$	$(C + C^{\mathrm{T}})(X\mathbf{a} + \mathbf{b})\mathbf{a}^{\mathrm{T}}$

11. Differentiation of specific scalar functions with respect to a matrix X. (Here $\{A, B\}$ are constants.)

y (a scalar)	$\frac{\partial y}{\partial X}$ (where X is a matrix)
$\mathrm{tr}(X)$	I
$\mathrm{tr}(A^{\mathrm{T}}X)$, $\mathrm{tr}(XA^{\mathrm{T}})$, $\mathrm{tr}(AX^{\mathrm{T}})$, or $\mathrm{tr}(X^{\mathrm{T}}A)$	A
$\mathrm{tr}(AXB)$	$A^{\mathrm{T}}B^{\mathrm{T}}$
$\mathrm{tr}(XAX^{\mathrm{T}})$	$X^{\mathrm{T}}(A + A^{\mathrm{T}})$
$\mathrm{tr}(X^{\mathrm{T}}AX)$	$(A + A^{\mathrm{T}})X$
$\mathrm{tr}(X^{\mathrm{T}}AXB)$	$AXB + A^{\mathrm{T}}XB^{\mathrm{T}}$
$\mathrm{tr}(e^X)$	$e^{X^{\mathrm{T}}}$
$\det(X)$ or $\det(X^{\mathrm{T}})$	$\det(X)\,(X^{-1})^{\mathrm{T}}$
$\det(AXB)$	$\det(AXB)\,(X^{-1})^{\mathrm{T}}$
$\log\lvert X\rvert$	$(X^{-1})^{\mathrm{T}}$

5.2 DIFFERENTIAL FORMS

Let $dx_k(\cdot)$ be the function that assigns a vector its k^{th} coordinate; for the vector $\mathbf{a} = (a_1, \dots, a_k, \dots, a_n)$ we have $dx_k(\mathbf{a}) = a_k$. Geometrically, $dx_k(\mathbf{a})$ is the length, with appropriate sign, of the projection of **a** on the k^{th} coordinate axis. If the $\{F_i\}$ are functions then the following linear combination of the functions $\{dx_k\}$

$$\omega_{\mathbf{x}} = F_1(\mathbf{x})\, dx_1 + F_2(\mathbf{x})\, dx_2 + \cdots + F_n(\mathbf{x})\, dx_n \tag{5.2.1}$$

is a new function $\omega_{\mathbf{x}}$ that acts on vectors **a** as

$$\omega_{\mathbf{x}}(\mathbf{a}) = F_1(\mathbf{x})\, dx_1(\mathbf{a}) + F_2(\mathbf{x})\, dx_2(\mathbf{a}) + \cdots + F_n(\mathbf{x})\, dx_n(\mathbf{a}). \tag{5.2.2}$$

Such a function is a *differential 1-form* or a *1-form*. For example:

1. If $\mathbf{a} = (-2, 0, 4)$ then $dx_1(\mathbf{a}) = -2$, $dx_2(\mathbf{a}) = 0$, and $dx_3(\mathbf{a}) = 4$.
2. If in $\mathbb{R}^2$, $\omega_{\mathbf{x}} = \omega_{(x,y)} = x^2\, dx + y^2\, dy$, then $\omega_{(x,y)}(a, b) = ax^2 + by^2$ and $\omega_{(1,-3)}(a, b) = a + 9b$.
3. Note that $\nabla_{\mathbf{x}} f$, the differential of f at **x**, is a 1-form with

$$\begin{aligned}\nabla_{\mathbf{x}} f(\mathbf{a}) &= \frac{\partial f}{\partial x_1}(\mathbf{x})\, dx_1(\mathbf{a}) + \frac{\partial f}{\partial x_2}(\mathbf{x})\, dx_2(\mathbf{a}) + \frac{\partial f}{\partial x_3}(\mathbf{x})\, dx_3(\mathbf{a}) \\ &= \frac{\partial f}{\partial x_1}(\mathbf{x}) a_1 + \frac{\partial f}{\partial x_2}(\mathbf{x}) a_2 + \frac{\partial f}{\partial x_3}(\mathbf{x}) a_3.\end{aligned}$$

5.2.1 PRODUCTS OF 1-FORMS

The basic 1-forms in $\mathbb{R}^3$ are dx_1, dx_2, and dx_3. The *wedge product* (or *exterior product*) $dx_1 \wedge dx_2$ is defined so that it is a function of ordered pairs of vectors in $\mathbb{R}^2$. Geometrically, $dx_1 \wedge dx_2(\mathbf{a}, \mathbf{b})$ is the area of the parallelogram spanned by the projections of **a** and **b** into the (x_1, x_2)-plane. The sign of the area is determined so that if the projections of **a** and **b** have the same orientation as the positive x_1 and x_2 axes, then the area is positive; it is negative when these orientations are opposite. Thus, if $\mathbf{a} = (a_1, a_2, a_3)$ and $\mathbf{b} = (b_1, b_2, b_3)$, then

$$dx_1 \wedge dx_2(\mathbf{a}, \mathbf{b}) = \det \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} = a_1 b_2 - a_2 b_1, \tag{5.2.3}$$

and the determinant automatically gives the correct sign. This generalizes to

$$dx_i \wedge dx_j(\mathbf{a}, \mathbf{b}) = \det \begin{bmatrix} dx_i(\mathbf{a}) & dx_i(\mathbf{b}) \\ dx_j(\mathbf{a}) & dx_j(\mathbf{b}) \end{bmatrix} = \det \begin{bmatrix} a_i & b_i \\ a_j & b_j \end{bmatrix}. \tag{5.2.4}$$

1. If f and g are real-valued functions and
   (a) If ω and μ are 1-forms, then $f\omega + g\mu$ is a 1-form.
   (b) If ω, ν, and μ are 1-forms, then $(f\omega + g\nu) \wedge \mu = f\, \omega \wedge \mu + g\, \nu \wedge \mu$.
2. $dx_i \wedge dx_j = -dx_j \wedge dx_i$
3. $dx_i \wedge dx_i = 0$
4. $dx_i \wedge dx_j(\mathbf{b}, \mathbf{a}) = -dx_i \wedge dx_j(\mathbf{a}, \mathbf{b})$

5.2.2 DIFFERENTIAL 2-FORMS

In $\mathbb{R}^3$, the most general linear combination of the functions $dx_i \wedge dx_j$ has the form $c_1\, dx_2 \wedge dx_3 + c_2\, dx_3 \wedge dx_1 + c_3\, dx_1 \wedge dx_2$. If $\mathbf{F} = (F_1, F_2, F_3)$ is a vector field, then the function of ordered pairs,

$$\tau_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) = F_1(\mathbf{x})\, dx_2 \wedge dx_3 + F_2(\mathbf{x})\, dx_3 \wedge dx_1 + F_3(\mathbf{x})\, dx_1 \wedge dx_2 \tag{5.2.5}$$

is a *differential 2-form* or *2-form.*

EXAMPLES

1. For the specific 2-form $\tau_{\mathbf{x}} = 2\, dx_2 \wedge dx_3 + dx_3 \wedge dx_1 + 5\, dx_1 \wedge dx_2$, if $\mathbf{a} = (1, 2, 3)$ and $\mathbf{b} = (0, 1, 1)$, then

$$\begin{aligned} \tau_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) &= 2 \det \begin{bmatrix} 2 & 1 \\ 3 & 1 \end{bmatrix} + \det \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} + 5 \det \begin{bmatrix} 1 & 0 \\ 2 & 1 \end{bmatrix} \\ &= 2 \cdot (-1) + 1 \cdot (-1) + 5 \cdot (1) = 2 \end{aligned}$$

independent of $\mathbf{x}$. Note that $\mathbf{a} \times \mathbf{b} = \det \begin{bmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ 1 & 2 & 3 \\ 0 & 1 & 1 \end{bmatrix} = (-1, -1, 1)$, and so $\tau_{\mathbf{x}}(\mathbf{a}, \mathbf{b}) = (2, 1, 5) \cdot (\mathbf{a} \times \mathbf{b})$. (Of course, this only works in 2 dimensions since cross-products are not defined in higher dimensions.)

2. When changing from Cartesian coordinates to polar coordinates, the element of area dA can be written

$$\begin{aligned} dA &= dx \wedge dy \\ &= (-r \sin\theta\, d\theta + \cos\theta\, dr) \wedge (r \cos\theta\, d\theta + \sin\theta\, dr) \\ &= -r^2 \sin\theta \cos\theta\, (d\theta \wedge d\theta) + \sin\theta \cos\theta\, (dr \wedge dr) \\ &\quad - r \sin^2\theta\, (d\theta \wedge dr) + r \cos^2\theta\, (dr \wedge d\theta) \\ &= r\, dr \wedge d\theta. \end{aligned} \tag{5.2.6}$$

5.2.3 THE 2-FORMS IN $\mathbb{R}^N$

Every 2-form can be written as a linear combination of "basic 2-forms." For example, in $\mathbb{R}^2$ there is only one basic 2-form (which may be taken to be $dx_1 \wedge dx_2$) and in $\mathbb{R}^3$ there are 3 basic 2-forms (possibly the set $\{dx_1 \wedge dx_2, dx_2 \wedge dx_3, dx_3 \wedge dx_1\}$). The exterior product of any two 1-forms (in, say, $\mathbb{R}^n$) is found by multiplying the 1-forms as if they were ordinary polynomials in the variables $dx_1, \ldots, dx_n$, and then simplifying using the rules for $dx_i \wedge dx_j$.

EXAMPLE Denoting the basic 1-forms in $\mathbb{R}^3$ as dx, dy, and dz then

$$\begin{aligned} (x\, dx + y^2\, dy) \wedge (dx + x\, dy) &= x\, (dx \wedge dx) + y^2\, (dy \wedge dx), \\ &\quad + x^2\, (dx \wedge dy) + xy^2\, (dy \wedge dy), \\ &= 0 - y^2\, (dx \wedge dy) + x^2\, (dx \wedge dy) + 0, \\ &= (x^2 - y^2)\, dx \wedge dy. \end{aligned} \tag{5.2.7}$$

5.2.4 HIGHER DIMENSIONAL FORMS

The basic 3-form $dx_1 \wedge dx_2 \wedge dx_3$ represents a signed volume:

$$dx_1 \wedge dx_2 \wedge dx_3(\mathbf{a}, \mathbf{b}, \mathbf{c}) = \det \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \tag{5.2.8}$$

which is a 3-dimensional oriented volume of the parallelepiped defined by the vectors **a**, **b**, and **c**.

The basic p-forms in $\mathbb{R}^n$, with $p \geq 1$, are

$$dx_{k_1} \wedge dx_{k_2} \wedge \cdots \wedge dx_{k_p}(\mathbf{a}_1, \dots, \mathbf{a}_p) = \det(dx_{k_i}(\mathbf{a}_j))_{\substack{i=1,\dots,p \\ j=1,\dots,p}}. \tag{5.2.9}$$

The general p-forms are linear combinations of the basic p-forms.

1. The general p-form can be written $\omega^p = \sum_{i_1 < \cdots < i_p} f_{i_1,\dots,i_p} dx_{i_1} \wedge \cdots \wedge dx_{i_k}$, where $1 \leq i_k \leq n$ for $k = 1, \dots, p$. This sum has $\binom{n}{p}$ distinct non-zero terms.
2. If ω^p is a p-form in $\mathbb{R}^n$ and ω^q is a q-form in $\mathbb{R}^n$, then $\omega^p \wedge \omega^q = (-1)^{pq}\, \omega^q \wedge \omega^p$.
3. A basic p-form with a repeated factor is zero. Thus if $p > n$, then any p-form is identically zero.
4. *Stoke's Theorem*: $\int_M d\omega = \int_{\partial M} \omega$ where ω is a differential form.

5.2.5 THE EXTERIOR DERIVATIVE

The exterior differentiation operator is denoted by d. When d is applied to a scalar function $f(\mathbf{x})$, the result is a 1-form $df = \nabla_{\mathbf{x}} f = \frac{\partial f}{\partial x_1}\, dx_1 + \cdots + \frac{\partial f}{\partial x_n}\, dx_n$. For the 1-form $\omega^1 = f_1\, dx + \cdots + f_n\, dx_n$ the *exterior derivative* is $d\omega^1 = (df_1) \wedge dx_1 + \cdots + (df_n) \wedge dx_n$. This generalizes to higher dimensional forms.

EXAMPLE

If $\omega^1_{(x_1,x_2)} = x_1 x_2\, dx_1 + (x_1^2 + x_2^2)\, dx_2$, then $d\omega^1$ is given by

$$\begin{aligned} d\omega^1 &= d(x_1 x_2\, dx_1 + (x_1^2 + x_2^2)\, dx_2) \\ &= (x_2\, dx_1 + x_1\, dx_2) \wedge dx_1 + (2x_1\, dx_1 + 2x_2\, dx_2) \wedge dx_2 \\ &= x_1\, dx_1 \wedge dx_2. \end{aligned}$$

5.2.6 PROPERTIES OF THE EXTERIOR DERIVATIVE

1. If $f_1(x_1, x_2)$ and $f_2(x_1, x_2)$ are differentiable functions, then
$$df_1 \wedge df_2 = \det\left(\frac{\partial(f_1, f_2)}{\partial(x_1, x_2)}\right) dx_1 \wedge dx_2. \tag{5.2.10}$$
2. If ω^p and ω^q represent a p-form and a q-form, then
$$d(\omega^p \wedge \omega^q) = (d\omega^p) \wedge \omega^q + (-1)^{pq} \omega^p \wedge (d\omega^q). \tag{5.2.11}$$
3. If ω^p is a p-form with at least two derivatives, then $d(d\omega^p) = 0$.
   - $d(d\omega^0) = 0$ is equivalent to $\operatorname{curl}(\operatorname{grad} f) = \nabla \times (\nabla f) = 0$.
   - $d(d\omega^1) = 0$ is equivalent to $\operatorname{div}(\operatorname{curl} \mathbf{F}) = \nabla \cdot (\nabla \mathbf{F}) = 0$.

5.3 INTEGRATION

5.3.1 DEFINITIONS

The following definitions apply to the expression $I = \int_a^b f(x)\,dx$:

1. The *integrand*	is $f(x)$.
2. The *upper limit*	is b.
3. The *lower limit*	is a.
4. I	is *the integral of $f(x)$ from a to b.*

It is conventional to indicate the indefinite integral of a function represented by a lowercase letter by the corresponding uppercase letter. For example, $F(x) = \int_a^x f(t)\,dt$ and $G(x) = \int_a^x g(t)\,dt$. Note that all functions that differ from $F(x)$ by a constant are also indefinite integrals of $f(x)$. We have the following notation:

1. $\int f(x)\,dx$	indefinite integral of $f(x)$ (also written $\int^x f(t)\,dt$)
2. $\int_a^b f(x)\,dx$	definite integral of $f(x)$; for a continuous function defined as $\lim_{n\to\infty}\left(\frac{b-a}{n}\sum_{k=1}^{n} f\left[a+\frac{k}{n}(b-a)\right]\right)$
3. $\oint_{\mathcal{C}} f(x)\,dx$	definite integral of $f(x)$, taken along the contour $\mathcal{C}$
4. $\int_a^\infty f(x)\,dx$	defined as $\lim_{R\to\infty}\int_a^R f(x)\,dx$
5. $\int_{-\infty}^\infty f(x)\,dx$	defined as the limit of $\int_{-S}^R f(x)\,dx$ as R and S independently go to ∞
6. *Improper integral*	integral for which the region of integration is not bounded, or the integrand is not bounded

7. *Cauchy principal value*
   (a) The Cauchy principal value of the integral $\int_a^b f(x)\,dx$, denoted $⨍_a^b f(x)\,dx$, is defined as $\lim_{\epsilon\to 0+}\left(\int_a^{c-\epsilon} f(x)\,dx + \int_{c+\epsilon}^b f(x)\,dx\right)$, assuming that f is singular only at c.
   (b) The Cauchy principal value of the integral $\int_{-\infty}^\infty f(x)\,dx$ is defined as the limit of $\int_{-R}^R f(x)\,dx$ as $R\to\infty$.
8. If, at the complex point $z = a$, $f(z)$ is either analytic or has an isolated singularity, then the residue of $f(z)$ at $z = a$ is given by the contour integral $\text{Res}_f(a) = \frac{1}{2\pi i}\oint_{\mathcal{C}} f(\xi)\,d\xi$; where $\mathcal{C}$ is a closed contour around a in a positive direction.

5.3.2 PROPERTIES OF INTEGRALS

Indefinite integrals have the properties (here F is the anti-derivative of f):

1. $\int [af(x) + bg(x)]\, dx = a \int f(x)\, dx + b \int g(x)\, dx$ — linearity
2. $\int f(x)g(x)\, dx = F(x)g(x) - \int F(x)g'(x)\, dx$ — integration by parts
3. $\int f(g(x))g'(x)\, dx = F(g(x))$ — substitution
4. $\int f(ax + b)\, dx = \frac{1}{a}F(ax + b)$
5. If $f(x)$ is an odd function and $F(0) = 0$, then $F(x)$ is an even function.
6. If $f(x)$ is an even function and $F(0) = 0$, then $F(x)$ is an odd function.
7. If $f(x)$ has a finite number of discontinuities, then the integral $\int f(x)\, dx$ is the sum of the integrals over those subintervals where $f(x)$ is continuous (provided they exist).
8. *Fundamental theorem of calculus*: If $f(x)$ is bounded, and integrable on $[a, b]$, and there exists a function $F(x)$ such that $F'(x) = f(x)$ for $a \le x \le b$ then for $a \le x \le b$

$$\int_a^x f(x)\, dx = F(x) \Big|_a^x = F(x) - F(a) \tag{5.3.1}$$

Definite integrals have the properties:

(a) $\int_a^a f(x)\, dx = 0$

(b) $\int_a^b f(x)\, dx = -\int_b^a f(x)\, dx$

(c) $\int_a^b f(x)\, dx + \int_b^c f(x)\, dx = \int_a^c f(x)\, dx$ — additivity

(d) $\int_a^b [cf(x) + dg(x)]\, dx = c\int_a^b f(x)\, dx + d\int_a^b g(x)\, dx$ — linearity

5.3.2.1 Convergence tests

1. If $\int_a^b |f(x)|\, dx$ is convergent, and f is integrable, then $\int_a^b f(x)\, dx$ is convergent.
2. If $0 \le f(x) \le g(x)$, $\int_a^b g(x)\, dx$ is convergent, and f is integrable, then $\int_a^b f(x)\, dx$ is convergent.
3. If $0 \le g(x) \le f(x)$ and $\int_a^b g(x)\, dx$ is divergent, then $\int_a^b f(x)\, dx$ is divergent.

The following integrals are often used with the above tests:

(a) $\displaystyle\int_2^\infty \frac{dx}{x(\log x)^p}$ and $\displaystyle\int_1^\infty \frac{dx}{x^p}$ converge when $p > 1$, and diverge when $p \le 1$.

(b) $\displaystyle\int_0^1 \frac{dx}{x^p}$ converges when $p < 1$, and diverges when $p \ge 1$.

5.3.2.2 Double Integrals

1. *Bounds on Integration*: If $m \leq f(x,y) \leq M$ in a domain D of area A, then
$$mA \leq \iint_D f(x,y)\,dx\,dy \leq MA.$$
2. *Integration of Inequalities*: If $a(x,y) \leq f(x,y) \leq b(x,y)$ in a domain D, then
$$\iint_D a(x,y)\,dx\,dy \leq \iint_D f(x,y)\,dx\,dy \leq \iint_D b(x,y)\,dx\,dy.$$
3. *Absolute value theorem*: $$\left|\iint_D f(x,y)\,dx\,dy\right| \leq \iint_D |f(x,y)|\,dx\,dy.$$
4. *Mean value theorem*: If $f(x,y)$ is continuous in a domain D of area A, then there exists at least one point $(\overline{x},\overline{y})$ in D such that
$$\iint_D f(x,y)\,dx\,dy = f(\overline{x},\overline{y})\,A.$$
5. *Area of a surface*: If a surface is defined parametrically by the equations $\{x = x(u,v), y = y(u,v), z = z(u,v)\}$ for values of (u,v) defined in D_{uv}, then the area of the surface is
$$\text{Area} = \iint_{D_{uv}} \sqrt{EG - F^2}\,du\,dv \qquad \text{where}$$
$$E = \left(\frac{\partial x}{\partial u}\right)^2 + \left(\frac{\partial y}{\partial u}\right)^2 + \left(\frac{\partial z}{\partial u}\right)^2$$
$$G = \left(\frac{\partial x}{\partial v}\right)^2 + \left(\frac{\partial y}{\partial v}\right)^2 + \left(\frac{\partial z}{\partial v}\right)^2$$
$$F = \frac{\partial x}{\partial u}\frac{\partial x}{\partial v} + \frac{\partial y}{\partial u}\frac{\partial y}{\partial v} + \frac{\partial z}{\partial u}\frac{\partial z}{\partial v}$$
6. *Change of Variable*: Let $x(u,v)$ and $y(u,v)$ be continuously differentiable functions that map one-to-one the domain D_{uv} in the uv-plane to the domain D_{xy} in the xy-plane. Let $f(x,y)$ be a continuous function in D_{xy}. Then
$$\iint_{D_{xy}} f(x,y)\,dx\,dy = \iint_{D_{uv}} f\left(x(u,v), y(u,v)\right)\,|J(u,v)|\,du\,dv \quad (5.3.2)$$
where $J(u,v)$ is the Jacobian of the mapping of D_{uv} onto D_{xy}:
$$J(u,v) = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u} \quad (5.3.3)$$
   - Not that the Jacobian of the mapping from Cartesian Coordinates (x,y) to polar coordinates (r,θ) via $\{x = r\cos\theta, y = r\sin\theta\}$ is $J(r,\theta) = r$.

EXAMPLE A circle of radius R in Cartesian coordinates has area $A = \iint_{x^2+y^2\leq R^2} 1\,dx\,dy$. Changing to polar coordinates this becomes $A = \int_0^{2\pi}\int_0^R (1 \times J(r,\theta))\,dr\,d\theta = \left(\int_0^{2\pi} d\theta\right)\left(\int_0^R r\,dr\right) = (2\pi)\left(\frac{R^2}{2}\right) = \pi R^2$

5.3.3 APPLICATIONS OF INTEGRATION

1. Using Green's theorems, the area bounded by the simple, closed, positively oriented contour $\mathcal{C}$ is

$$\text{area} = \oint_{\mathcal{C}} x\,dy = -\oint_{\mathcal{C}} y\,dx. \tag{5.3.4}$$

2. Arc length:

   (a) $s = \displaystyle\int_{x_1}^{x_2} \sqrt{1 + y'^2}\,dx$ for $y = f(x)$.

   (b) $s = \displaystyle\int_{t_1}^{t_2} \sqrt{\dot{\phi}^2 + \dot{\psi}^2}\,dt$ for $x = \phi(t)$, $y = \psi(t)$.

   (c) $s = \displaystyle\int_{\theta_1}^{\theta_2} \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\,d\theta = \int_{r_1}^{r_2} \sqrt{1 + r^2\left(\frac{dr}{d\theta}\right)^2}\,dr$ for $r = f(\theta)$.

3. Surface area for surfaces of revolution:

   (a) $A = 2\pi \displaystyle\int_{x_1}^{x_2} f(x)\sqrt{1 + (f'(x))^2}\,dx$
   when $y = f(x)$ is rotated about the x-axis.

   (b) $A = 2\pi \displaystyle\int_{y_1}^{y_2} x\sqrt{1 + (f'(x))^2}\,dy$
   when $y = f(x)$ is rotated about the y-axis and f is one-to-one.

   (c) $A = 2\pi \displaystyle\int_{t_1}^{t_2} \psi\sqrt{\dot{\phi}^2 + \dot{\psi}^2}\,dt$
   for $x = \phi(t)$, $y = \psi(t)$ rotated about the x-axis.

   (d) $A = 2\pi \displaystyle\int_{t_1}^{t_2} \phi\sqrt{\dot{\phi}^2 + \dot{\psi}^2}\,dt$
   for $x = \phi(t)$, $y = \psi(t)$ rotated about the y-axis.

   (e) $A = 2\pi \displaystyle\int_{\phi_1}^{\phi_2} r\sin\phi\sqrt{r^2 + \left(\frac{dr}{d\phi}\right)^2}\,d\phi$
   for $r = r(\phi)$ rotated about the x-axis.

   (f) $A = 2\pi \displaystyle\int_{\phi_1}^{\phi_2} r\cos\phi\sqrt{r^2 + \left(\frac{dr}{d\phi}\right)^2}\,d\phi$
   for $r = r(\phi)$ rotated about the y-axis.

4. Volumes of revolution:

   (a) $V = \pi \displaystyle\int_{x_1}^{x_2} f^2(x)\,dx$ for $y = f(x)$ rotated about the x-axis

   (b) $V = \pi \displaystyle\int_{x_1}^{x_2} x^2 f'(x)\,dx$ for $y = f(x)$ rotated about the y-axis

(c) $V = \pi \int_{y_1}^{y_2} g^2(y)\,dy$ for $x = g(y)$ rotated about the y-axis

(d) $V = \pi \int_{t_1}^{t_2} \psi^2 \dot{\phi}\,dt$ for $x = \phi(t)$, $y = \psi(t)$ rotated about the x-axis

(e) $V = \pi \int_{t_1}^{t_2} \phi^2 \dot{\psi}\,dt$ for $x = \phi(t)$, $y = \psi(t)$ rotated about the y-axis

(f) $V = \pi \int_{\phi_1}^{\phi_2} \sin^2 \phi \left(\frac{dr}{d\phi} \cos\phi - r \sin\phi \right) d\phi$

for $r = f(\phi)$ rotated about the x-axis

(g) $V = \pi \int_{\phi_1}^{\phi_2} \cos^2 \phi \left(\frac{dr}{d\phi} \sin\phi - r \cos\phi \right) d\phi$

for $r = f(\phi)$ rotated about the y-axis

5. The area enclosed by the curve $x^{b/c} + y^{b/c} = a^{b/c}$, where $a > 0$, c is an odd integer and b is an even integer, is $A = \frac{2ca^2}{b} \frac{\left[\Gamma\left(\frac{c}{b}\right)\right]^2}{\Gamma\left(\frac{2c}{b}\right)}$.

5.3.4 INTEGRAL INEQUALITIES

1. **Schwartz inequality:** $\int_a^b |fg| \leq \sqrt{\left(\int_a^b |f|^2\right)\left(\int_a^b |g|^2\right)}$.

2. **Minkowski's inequality:**

$$\left(\int_a^b |f+g|^p\right)^{1/p} \leq \left(\int_a^b |f|^p\right)^{1/p} + \left(\int_a^b |g|^p\right)^{1/p} \text{ when } p \geq 1.$$

3. **Hölder's inequality:**

$$\int_a^b |fg| \leq \left[\int_a^b |f|^p\right]^{1/p} \left[\int_a^b |g|^q\right]^{1/q} \text{ when } \tfrac{1}{p} + \tfrac{1}{q} = 1,\ p > 1, \text{ and } q > 1.$$

4. $\left|\int_a^b f(x)\,dx\right| \leq \int_a^b |f(x)|\,dx \leq \left(\max_{x\in[a,b]} |f(x)|\right)(b-a)$ assuming $a \leq b$.

5. If $f(x) \leq g(x)$ on the interval $[a,b]$, then $\int_a^b f(x)\,dx \leq \int_a^b g(x)\,dx$.

5.3.5 METHODS OF EVALUATING INTEGRALS

5.3.5.1 Substitution

Substitution can be used to change integrals to simpler forms. When the transform $t = g(x)$ is chosen, the integral $I = \int f(t)\,dt$ becomes $I = \int f(g(x))\,dt = \int f(g(x))g'(x)\,dx$. Several precautions must be taken when using substitutions:

1. Make the substitution in the dx term, as well as everywhere else in the integral.
2. Ensure that the function substituted is one-to-one and differentiable. If this is not true, then the integral must be restricted in such a way as to make it true.
3. With definite integrals, the limits should also be expressed in terms of the new dependent variables. With indefinite integrals, it is necessary to perform the reverse substitution to obtain the answer in terms of the original independent variable. This may also be done for definite integrals, but it is usually easier to change the limits.

EXAMPLE Consider the integral

$$I = \int \frac{x^4}{\sqrt{a^2 - x^2}}\,dx \tag{5.3.5}$$

for $a \neq 0$. Choosing the substitution $x = |a| \sin\theta$ we find $dx = |a| \cos\theta\,d\theta$ and

$$\sqrt{a^2 - x^2} = \sqrt{a^2 - a^2 \sin^2\theta} = |a| \sqrt{1 - \sin^2\theta} = |a|\,|\cos\theta|\,. \tag{5.3.6}$$

Note the absolute value signs. It is important to interpret the square root radical consistently as the positive square root. Thus $\sqrt{x^2} = |x|$.

Note that the substitution chosen is not a one-to-one function, that is, it does not have a unique inverse. Thus the range of θ must be restricted in such a way as to make the function one-to-one. In this case we can solve for θ to obtain

$$\theta = \sin^{-1} \frac{x}{|a|}. \tag{5.3.7}$$

This will be unique if we restrict the inverse sine to the principal values $-\frac{\pi}{2} \leq \theta \leq \frac{\pi}{2}$.

Thus, the integral becomes (with $dx = |a| \cos\theta\,d\theta$)

$$I = \int \frac{a^4 \sin^4\theta}{|a|\,|\cos\theta|}\,|a| \cos\theta\,d\theta. \tag{5.3.8}$$

For the range of values chosen for θ, we find that $\cos\theta$ is always non-negative. Thus, removing the absolute value signs from $\cos\theta$ results in

$$I = a^4 \int \sin^4\theta\,d\theta. \tag{5.3.9}$$

Using integration formula #283 on page 324, and simplifying, this becomes

$$I = -\frac{a^4}{4} \sin^3\theta \cos\theta - \frac{3a^4}{8} \sin\theta \cos\theta + \frac{3a^4}{8}\theta + C. \tag{5.3.10}$$

To obtain an evaluation of I as a function of x, we must transform variables from θ to x. We have

$$\cos\theta = \pm\sqrt{1-\sin^2\theta} = \pm\sqrt{1-\frac{x^2}{a^2}} = \pm\frac{\sqrt{a^2-x^2}}{|a|}. \tag{5.3.11}$$

Because of the previously recorded fact that $\cos\theta$ is non-negative for our range of θ, we may omit the $\pm$ sign. Using $\sin\theta = x/|a|$ and $\cos\theta = \sqrt{a^2-x^2}/|a|$ we can evaluate Equation (5.3.10) to obtain the final result,

$$I = \int \frac{x^4}{\sqrt{a^2-x^2}} = -\frac{x^3}{4}\sqrt{a^2-x^2} - \frac{3a^2x}{8}\sqrt{a^2-x^2} + \frac{3a^4}{8}\sin^{-1}\frac{x}{|a|} + C. \tag{5.3.12}$$

5.3.5.2 Partial fraction decomposition

An integral of the form $\int R(x)\,dx$, where R is a rational function, can be evaluated in terms of elementary functions. The technique is to factor the denominator of R and create a partial fraction decomposition; each resulting sub-integral is elementary.

EXAMPLE Consider $I = \int \frac{2x^3-10x^2+13x-4}{x^2-5x+6}\,dx$. This can be written as

$$I = \int\left(2x + \frac{x-4}{x^2-5x+6}\right)dx = \int\left(2x + \frac{2}{x-2} - \frac{1}{x-3}\right)dx$$

which can be readily integrated $I = x^2 + 2\ln(x-2) - \ln(x-3)$.

5.3.5.3 Useful transformations

The following transformations may make evaluation of an integral easier:

1. $\int f\left(x, \sqrt{x^2+a^2}\right)\,dx = a\int f(a\tan u, a\sec u)\sec^2 u\,du$ when $u = \tan^{-1}\frac{x}{a}$ and $a > 0$.
2. $\int f\left(x, \sqrt{x^2-a^2}\right)\,dx = a\int f(a\sec u, a\tan u)\sec u\tan u\,du$ when $u = \sec^{-1}\frac{x}{a}$ and $a > 0$.
3. $\int f\left(x, \sqrt{a^2-x^2}\right)\,dx = a\int f(a\sin u, a\cos u)\cos u\,du$ when $u = \sin^{-1}\frac{x}{a}$ and $a > 0$.
4. $\int f(\sin x)\,dx = 2\int f\left(\frac{2z}{1+z^2}\right)\frac{dz}{1+z^2}$ when $z = \tan\frac{x}{2}$.
5. $\int f(\cos x)\,dx = 2\int f\left(\frac{1-z^2}{1+z^2}\right)\frac{dz}{1+z^2}$ when $z = \tan\frac{x}{2}$.
6. $\int f(\cos x)\,dx = -\int f(v)\frac{dv}{\sqrt{1-v^2}}$ when $v = \cos x$.
7. $\int f(\sin x)\,dx = \int f(u)\frac{du}{\sqrt{1-u^2}}$ when $u = \sin x$.
8. $\int f(\sin x, \cos x)\,dx = \int f\left(u, \sqrt{1-u^2}\right)\frac{du}{\sqrt{1-u^2}}$ when $u = \sin x$.
9. $\int f(\sin x, \cos x)\,dx = 2\int f\left(\frac{2z}{1+z^2}, \frac{1-z^2}{1+z^2}\right)\frac{dz}{1+z^2}$ when $z = \tan\frac{x}{2}$.
10. $\int_{-\infty}^{\infty} f(u)\,dx = \int_{-\infty}^{\infty} f(x)\,dx$ when $u = x - \sum_{j=1}^{n}\frac{a_j}{x-c_j}$ where $\{a_i\}$ is any sequence of positive constants and the $\{c_j\}$ are any real constants whatsoever.

Several transformations of the integral $\int_0^\infty f(x)\,dx$, with an infinite integration range, to an integral with a finite integration range, are shown:

$t(x)$	$x(t)$	$\frac{dx}{dt}$	Finite interval integral
e^{-x}	$-\log t$	$-\frac{1}{t}$	$\int_0^1 \frac{f(-\log t)}{t}\,dt$
$\frac{x}{1+x}$	$\frac{t}{1-t}$	$\frac{1}{(1-t)^2}$	$\int_0^1 f\left(\frac{t}{1-t}\right)\frac{dt}{(1-t)^2}$
$\tanh x$	$\frac{1}{2}\log\frac{1+t}{1-t}$	$\frac{1}{1-t^2}$	$\int_0^1 f\left(\frac{1}{2}\log\frac{1+t}{1-t}\right)\frac{dt}{1-t^2}$

5.3.5.4 Integration by parts

In one dimension, the integration by parts formula is

$$\int u\,dv = uv - \int v\,du \qquad \text{for indefinite integrals,} \tag{5.3.13}$$

$$\int_a^b u\,dv = uv\Big|_a^b - \int_a^b v\,du \qquad \text{for definite integrals.} \tag{5.3.14}$$

When evaluating a given integral by this method, u and v must be chosen so that the form $\int u\,dv$ becomes identical to the given integral. This is usually accomplished by specifying u and dv and deriving du and v. Then the integration by parts formula will produce a boundary term and another integral to be evaluated. If u and v were well chosen, then this second integral may be easier to evaluate.

EXAMPLE Consider the integral

$$I = \int x \sin x\,dx.$$

Two obvious choices for the integration by parts formula are $\{u = x, dv = \sin x\,dx\}$ and $\{u = \sin x, dv = x\,dx\}$. We try each of them in turn.

1. Using $\{u = x, dv = \sin x\,dx\}$, we compute $du = dx$ and $v = \int dv = \int \sin x\,dx = -\cos x$. Hence, we can represent I in the alternative form as

$$I = \int x \sin x\,dx = \int u\,dv = uv - \int v\,du = -x\cos x + \int \cos x\,dx.$$

In this representation of I, we must evaluate the last integral. Because we know $\int \cos x\,dx = \sin x$ the final result is $I = \sin x - x\cos x$.

2. Using $\{u = \sin x, dv = x\,dx\}$ we compute $du = \cos x\,dx$ and $v = \int dv = \int x\,dx = x^2/2$. Hence, we can represent I in the alternative form as

$$I = \int x \sin x\,dx = \int u\,dv = uv - \int v\,du = \frac{x^2}{2}\sin x - \int \frac{x^2}{2}\cos x\,dx.$$

In this case, we have actually made the problem "worse" since the remaining integral appearing in I is "harder" than the one we started with.

EXAMPLE Consider the integral

$$I = \int e^x \sin x \, dx.$$

We choose to use the integration by parts formula with $u = e^x$ and $dv = \sin x \, dx$. From these we compute $du = e^x \, dx$ and $v = \int dv = \int \sin x \, dx = -\cos x$. Hence, we can represent I in the alternative form as

$$I = \int e^x \sin x \, dx = \int u \, dv = uv - \int v \, du = -e^x \cos x + \int e^x \cos x \, dx$$

If we write this as

$$I = -e^x \cos x + J \qquad \text{with} \qquad J = \int e^x \cos x \, dx, \tag{5.3.15}$$

then we can apply integration by parts to J using $\{u = e^x,\, dv = \cos x \, dx\}$. From these we compute $du = e^x \, dx$ and $v = \int dv = \int \cos x \, dx = \sin x$. Hence, we can represent J in the alternative form as

$$J = \int e^x \cos x \, dx = \int u \, dv = uv - \int v \, du = e^x \sin x - \int e^x \sin x \, dx$$

If we write this as

$$J = e^x \sin x - I, \tag{5.3.16}$$

then we can solve the linear Equations (5.3.15) and (5.3.16) simultaneously to determine both I and J. We find

$$\begin{aligned} I &= \int e^x \sin x \, dx = \frac{1}{2}\left(e^x \sin x - e^x \cos x\right), \quad \text{and} \\ J &= \int e^x \cos x \, dx = \frac{1}{2}\left(e^x \sin x + e^x \cos x\right). \end{aligned} \tag{5.3.17}$$

5.3.5.5 Extended integration by parts rule

The following rule is the result of $n+1$ applications of integration by parts. Let

$$g_1(x) = \int g(x) \, dx, \qquad g_2(x) = \int g_1(x) \, dx, \quad \dots, \quad g_m(x) = \int g_{m-1}(x) \, dx. \tag{5.3.18}$$

Then

$$\begin{aligned} \int f(x)g(x) \, dx &= f(x)g_1(x) - f'(x)g_2(x) + f''(x)g_3(x) - \dots \\ &\quad + (-1)^n f^{(n)}(x) g_{n+1}(x) + (-1)^{n+1} \int f^{(n+1)}(x) g_{n+1}(x) \, dx. \end{aligned} \tag{5.3.19}$$

5.3.6 TYPES OF INTEGRALS

5.3.6.1 Line and surface integrals

A **line integral** is a definite integral whose path of integration is along a specified curve; it can be evaluated by reducing it to ordinary integrals. If $f(x, y)$ is continuous on $\mathcal{C}$, and the integration contour $\mathcal{C}$ is parameterized by $(\phi(t), \psi(t))$ as t varies from a to b, then

$$\begin{aligned}\int_{\mathcal{C}} f(x, y)\, dx &= \int_a^b f\left(\phi(t), \psi(t)\right) \phi'(t)\, dt, \\ \int_{\mathcal{C}} f(x, y)\, dy &= \int_a^b f\left(\phi(t), \psi(t)\right) \psi'(t)\, dt.\end{aligned} \tag{5.3.20}$$

In a simply connected domain, the line integral $I = \int_{\mathcal{C}} X\, dx + Y\, dy + Z\, dz$ is independent of the closed curve $\mathcal{C}$ if and only if $\mathbf{u} = (X, Y, Z)$ is a gradient vector, $\mathbf{u} = \operatorname{grad} F$ (that is, $F_x = X$, $F_y = Y$, and $F_z = Z$).

Green's theorem: Let D be a domain of the xy plane, and let $\mathcal{C}$ be a piecewise smooth, simple closed curve in D whose interior R is also in D. Let $P(x, y)$ and $Q(x, y)$ be functions defined in D with continuous first partial derivatives in D. Then

$$\oint_{\mathcal{C}} (P\, dx + Q\, dy) = \iint\limits_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx\, dy. \tag{5.3.21}$$

This theorem may be written in the two alternative forms (using $\mathbf{u} = P(x, y)\mathbf{i} + Q(x, y)\mathbf{j}$ and $\mathbf{v} = Q(x, y)\mathbf{i} - P(x, y)\mathbf{j}$),

$$\oint_{\mathcal{C}} \mathbf{u}_T\, ds = \iint\limits_R \operatorname{curl} \mathbf{u}\, dx\, dy \quad \text{and} \quad \oint_{\mathcal{C}} \mathbf{v}_n\, ds = \iint\limits_R \operatorname{div} \mathbf{v}\, dx\, dy. \tag{5.3.22}$$

where $_T$ represents the tangent vector and $_n$ represents the normal vector. The first equation above is a simplification of Stokes's theorem; the second equation is the divergence theorem.

Stokes's theorem: Let S be a piecewise smooth oriented surface in space, whose boundary $\mathcal{C}$ is a piecewise smooth simple closed curve, directed in accordance with the given orientation of S. Let $\mathbf{u} = L\mathbf{i} + M\mathbf{j} + N\mathbf{k}$ be a vector field with continuous and differentiable components in a domain D of space including S. Then, $\int_{\mathcal{C}} u_T\, ds = \iint\limits_S (\operatorname{curl} \mathbf{u}) \cdot \mathbf{n}\, d\sigma$, where $\mathbf{n}$ a unit normal vector on S, that is

$$\begin{aligned}\int_{\mathcal{C}} L\, dx + M\, dy + N\, dz = \iint\limits_S &\left(\frac{\partial N}{\partial y} - \frac{\partial M}{\partial z}\right) dy\, dz \\ &+ \left(\frac{\partial L}{\partial z} - \frac{\partial N}{\partial x}\right) dz\, dx + \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\right) dx\, dy.\end{aligned} \tag{5.3.23}$$

Divergence theorem: Let $\mathbf{v} = L\mathbf{i} + M\mathbf{j} + N\mathbf{k}$ be a vector field in a domain D of space. Let L, M, and N be continuous with continuous derivatives in D. Let S be a piecewise smooth surface in D that forms the complete boundary of a bounded closed region R in D. Let $\mathbf{n}$ be the outer normal of S with respect to R. Then $\iint\limits_S v_n \, d\sigma = \iiint\limits_R \operatorname{div} \mathbf{v} \, dx \, dy \, dz$, that is

$$\iint\limits_S L \, dy \, dz + M \, dz \, dx + N \, dx \, dy$$

$$= \iiint\limits_R \left(\frac{\partial L}{\partial x} + \frac{\partial M}{\partial y} + \frac{\partial N}{\partial z} \right) dx \, dy \, dz. \quad (5.3.24)$$

If D is a three-dimensional domain with boundary B, let dV represent the volume element of D, let dS represent the surface element of B, and let $d\mathbf{S} = \mathbf{n} \, dS$, where $\mathbf{n}$ is the outer normal vector of the surface B. Then **Gauss' formulas** are

$$\begin{aligned} \iiint\limits_D \nabla \cdot \mathbf{A} \, dV &= \iint\limits_B d\mathbf{S} \cdot \mathbf{A} = \iint\limits_B (\mathbf{n} \cdot \mathbf{A}) \, dS, \\ \iiint\limits_D \nabla \times \mathbf{A} \, dV &= \iint\limits_B d\mathbf{S} \times \mathbf{A} = \iint\limits_B (\mathbf{n} \times \mathbf{A}) \, dS, \quad \text{and} \\ \iiint\limits_D \nabla \phi \, dV &= \iint\limits_B \phi \, d\mathbf{S}, \end{aligned} \quad (5.3.25)$$

where ϕ is an arbitrary scalar and $\mathbf{A}$ is an arbitrary vector.

Green's theorems also relate a volume integral to a surface integral: Let V be a volume with surface S, which we assume is simple and closed. Define n as the outward normal to S. Let ϕ and ψ be scalar functions which, together with $\nabla^2\phi$ and $\nabla^2\psi$, are defined in V and on S. Then

1. **Green's first theorem** states that
$$\int_S \phi \frac{\partial \psi}{\partial n} \, dS = \int_V \left(\phi \nabla^2 \psi + \nabla\phi \cdot \nabla\psi \right) \, dV. \quad (5.3.26)$$

2. **Green's second theorem** states that
$$\int_S \left(\phi \frac{\partial \psi}{\partial n} - \psi \frac{\partial \phi}{\partial n} \right) \, dS = \int_V \left(\phi \nabla^2 \psi - \psi \nabla^2 \phi \right) \, dV. \quad (5.3.27)$$

5.3.6.2 Contour integrals

If $f(z)$ is analytic inside a simple closed curve C (with proper orientation), then

1. The **Cauchy–Goursat integral theorem** is $\oint_{\mathcal{C}} f(\xi)\,d\xi = 0$.
2. **Cauchy's integral formula** is

$$f(z) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f(\xi)}{\xi - z}\,d\xi \qquad \text{and} \qquad f'(z) = \frac{1}{2\pi i}\oint_{\mathcal{C}} \frac{f(\xi)}{(\xi - z)^2}\,d\xi. \tag{5.3.28}$$

In general, $f^{(n)}(z) = \dfrac{n!}{2\pi i}\oint_{\mathcal{C}} \dfrac{f(\xi)}{(\xi - z)^{n+1}}\,d\xi$.

3. Use of residues
   (a) If $f(z)$ is written as a Laurent series about the point z_0

$$f(z) = \sum_{n=-\infty}^{\infty} a_n (z - z_0)^n = \cdots + \frac{a_{-2}}{(z - z_0)^2} + \frac{a_{-1}}{z - z_0} + a_0 + a_1(z - z_0) + \ldots, \tag{5.3.29}$$

   then the residue of $f(z)$ at $z = z_0$, denoted $\text{Res}_f(z_0)$, is the value a_{-1}.

   (b) **Residue theorem**: For every simple closed contour $\mathcal{C}$ enclosing a finite number of (necessarily isolated) singularities $\{z_1, z_2, \ldots, z_n\}$ of a single-valued analytic function $f(z)$ continuous on $\mathcal{C}$,

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} f(\xi)\,d\xi = \sum_{k=1}^{n} \text{Res}_f(z_k) \tag{5.3.30}$$

EXAMPLE Consider the integral

$$I = \int_0^{2\pi} \frac{d\theta}{2 + \cos\theta}. \tag{5.3.31}$$

Change the variable from θ to z via $z = e^{i\theta}$, so that $d\theta = -iz^{-1}\,dz$. Since $\cos\theta = \frac{1}{2}\left(z + z^{-1}\right)$, Equation (5.3.31) becomes

$$I = \oint_{\mathcal{C}} \frac{-iz^{-1}\,dz}{2 + \frac{1}{2}\left(z + z^{-1}\right)} = \oint_{\mathcal{C}} \frac{-2i}{z^2 + 4z + 1}\,dz. \tag{5.3.32}$$

As θ varies from 0 to 2π, z traces out a circle of radius one in the clockwise sense. Hence, the contour $\mathcal{C}$ is the unit circle.

The quadradic in the denominator factors as $z^2 + 4z + 1 = (z - \alpha_+)(z - \alpha_-)$, where $\alpha_\pm = -2 \pm \sqrt{3}$. Hence, $I = \oint_{\mathcal{C}} \dfrac{-2i}{(z - \alpha_+)(z - \alpha_-)}\,dz$. Of the two poles in the integrand ($z = \alpha_\pm$) only $z = \alpha_+$ is inside the unit circle. The Laurent series of the integrand about the pole $z = \alpha_+$ is

$$\frac{-2i}{(z - \alpha_+)(z - \alpha_-)} = \left(\frac{-2i}{\alpha_+ - \alpha_-}\right)\frac{1}{z - \alpha_+} + h(z), \tag{5.3.33}$$

where $h(z) = \left(\frac{-2i}{\alpha_- - \alpha_+}\right)\frac{1}{z - \alpha_-}$ is analytic at $z = \alpha_+$. Hence, the residue of the integrand at α_+ is $\text{Res}_f(\alpha_+) = \left(\frac{-2i}{\alpha_+ - \alpha_-}\right) = \frac{-i}{\sqrt{3}}$. Using this in the residue theorem (Equation (5.3.30)) results in

$$I = \oint_{\mathcal{C}} f(z)\,dz = 2\pi i \sum_{k=1}^{n} \text{Res}_f(z_k) = 2\pi i\left(\frac{-i}{\sqrt{3}}\right) = \frac{2\pi}{\sqrt{3}}.$$

5.3.7 VARIATIONAL PRINCIPLES

If J depends on a function $g(x)$ and its derivatives through an integral of the form $J[g] = \int F(g, g', \dots)\, dx$, then J will be stationary to small perturbations (i.e., first derivative zero) if F satisfies the corresponding Euler–Lagrange equation.

Function	Euler–Lagrange equation
$\int_R F\left(x, y, y'\right)\, dx$	$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) = 0$
$\int_R F\left(x, y, y', \dots, y^{(n)}\right)\, dx$	$\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial y''}\right) - \cdots + (-1)^n \frac{d^n}{dx^n}\left(\frac{\partial F}{\partial y^{(n)}}\right) = 0$
$\iint\limits_R \left[a\left(\frac{\partial u}{\partial x}\right)^2 + b\left(\frac{\partial u}{\partial x}\right)^2 + cu^2 + 2fu\right]\, dx\, dy$	$\frac{\partial}{\partial x}\left(a\frac{\partial u}{\partial x}\right) + \frac{\partial}{\partial y}\left(b\frac{\partial u}{\partial y}\right) - cu = f$
$\iint\limits_R F(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy})\, dx\, dy$	$\frac{\partial F}{\partial u} - \frac{\partial}{\partial x}\left(\frac{\partial F}{\partial u_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial F}{\partial u_y}\right) + \frac{\partial^2}{\partial x^2}\left(\frac{\partial F}{\partial u_{xx}}\right) + \frac{\partial^2}{\partial x \partial y}\left(\frac{\partial F}{\partial u_{xy}}\right) + \frac{\partial^2}{\partial y^2}\left(\frac{\partial F}{\partial u_{yy}}\right) = 0$

5.3.8 CONTINUITY OF INTEGRAL ANTI-DERIVATIVES

Consider the following different anti-derivatives of an integral

$$F(x) = \int f(x)\, dx = \int \frac{3}{5 - 4\cos x} dx = \begin{cases} 2\tan^{-1}\left(3\tan\frac{x}{2}\right) \\ 2\tan^{-1}(3\sin x/(\cos x + 1)) \\ -\tan^{-1}(-3\sin x/(5\cos x - 4)) \\ 2\tan^{-1}\left(3\tan\frac{x}{2}\right) + 2\pi\left\lfloor\frac{x}{2\pi} + \frac{1}{2}\right\rfloor \end{cases}$$

where $\lfloor\cdot\rfloor$ denotes the floor function. These anti-derivatives are all "correct" because differentiating any of them results in the original integrand (except at isolated points). However, if we desire $\int_0^{4\pi} f(x)\, dx = F(4\pi) - F(0)$ to hold, then only the last anti-derivative is correct. This is true because the other anti-derivatives of $F(x)$ are discontinuous when x is a multiple of π.

In general, if $\widehat{F}(x) = \int^x f(t)\, dt$ is a discontinuous evaluation (with $\hat{F}(x)$ discontinuous at the single point $x = b$), then a continuous evaluation on a finite interval is given by $\int_a^c f(x)\, dx = F(c) - F(a)$, where

$$F(x) = \hat{F}(x) - \hat{F}(a) + H(x - b)\left[\lim_{x \to b-} \hat{F}(x) - \lim_{x \to b+} \hat{F}(x)\right] \tag{5.3.34}$$

and where $H(\cdot)$ is the Heaviside function. For functions with an infinite number of discontinuities, note that $\sum_{n=1}^{\infty} H(x - pn - q) = \left\lfloor \frac{x - q}{p} \right\rfloor$.

5.3.9 ASYMPTOTIC INTEGRAL EVALUATION

1. *Laplace's method*: If $g(x_0) \neq 0$, $f'(x_0) = 0$, $f''(x_0) < 0$, and $\lambda \to \infty$, then
$$I_{x_0}(\lambda) \equiv \int_{x_0-\epsilon}^{x_0+\epsilon} g(x)e^{\lambda f(x)}\,dx \sim g(x_0)\sqrt{\frac{2\pi}{\lambda\,|f''(x_0)|}}\exp\left[\lambda f(x_0)\right] + \dots. \tag{5.3.35}$$
Hence, if points of local maximum $\{x_i\}$ satisfy $f'(x_i) = 0$ and $f''(x_i) < 0$, then $\int_{-\infty}^{\infty} g(x)e^{\lambda f(x)}\,dx \sim \sum_i I_{x_i}(\lambda)$.
2. *Method of stationary phase*: If $g(x_0) \neq 0$, $f(x_0) \neq 0$, $f'(x_0) = 0$, $f''(x_0) \neq 0$, and $\lambda \to \infty$, then
$$\int_{x_0-\epsilon}^{x_0+\epsilon} g(x)e^{i\lambda f(x)}dx \sim g(x_0)\sqrt{\frac{2\pi}{\lambda\,|f''(x_0)|}}\exp\left[i\lambda f(x_0) + \frac{i\pi}{4}\operatorname{sgn} f''(x_0)\right] + \dots.$$
3. *Power Laplace Integral*: If $g(x_0) \neq 0$, $f(x) > 0$, $f'(x_0) = 0$, $f''(x_0) < 0$, and $\lambda \to \infty$, then
$$\int_{x_0-\epsilon}^{x_0+\epsilon} g(x)\left[f(x)\right]^{\lambda}\,dx \sim g(x_0)\sqrt{\frac{2\pi}{\lambda\,|f''(x_0)|}}\left[f(x_0)\right]^{\lambda+\frac{1}{2}} + \dots.$$

5.3.10 TABLES OF INTEGRALS

Even with an extensive integral table it is uncommon to find a specific desired integral. Often a transformation is required. Simple transformations, such as substitutions (e.g., $y = ax$) are employed, almost unconsciously, by experienced users.

We adopt the following conventions in the integral tables:

1. A constant of integration must be included with all indefinite integrals.
2. Angles are measured in radians.
3. Inverse trigonometric and hyperbolic functions represent principal values.
4. Logarithmic expressions are to base e, unless otherwise specified, and are to be evaluated for the absolute value of the arguments involved therein.
5. The natural logarithm function is denoted as $\log x$.
6. The variables n and m usually denote integers. The denominator of an expression is not allowed to be zero; this may require that $m \neq 0$ or $m \neq n$ or some other similar statement.
7. When inverse trigonometric functions occur in the integrals, be careful with substitutions. There is often no problem for positive arguments of inverse trigonometric functions. However, if the argument is negative, special care must be used. Thus, if $u > 0$ then
$$\sin^{-1} u = \cos^{-1}\sqrt{1-u^2} = \csc^{-1}\frac{1}{u} = \dots.$$
However, if $u < 0$, then
$$\sin^{-1} u = -\cos^{-1}\sqrt{1-u^2} = -\pi - \csc^{-1}\frac{1}{u} = \dots.$$

5.4 TABLE OF INDEFINITE INTEGRALS

5.4.1 ELEMENTARY FORMS

1. $\int a\,dx = ax.$

2. $\int a\,f(x)\,dx = a \int f(x)\,dx.$

3. $\int \phi(y(x))\,dx = \int \frac{\phi(y)}{y'}\,dy,$ where $y' = \frac{dy}{dx}.$

4. $\int (u+v)\,dx = \int u\,dx + \int v\,dx,$ where u and v are any functions of x.

5. $\int u\,dv = u \int dv - \int v\,du = uv - \int v\,du.$

6. $\int u \frac{dv}{dx}\,dx = uv - \int v \frac{du}{dx}\,dx.$

7. $\int x^n\,dx = \frac{x^{n+1}}{n+1},$ except when $n = -1$.

8. $\int \frac{dx}{x} = \log x.$

9. $\int \frac{f'(x)}{f(x)}\,dx = \log f(x),\ (df(x) = f'(x)\,dx).$

10. $\int \frac{f'(x)}{2\sqrt{f(x)}}\,dx = \sqrt{f(x)},\ (df(x) = f'(x)\,dx).$

11. $\int e^x\,dx = e^x.$

12. $\int e^{ax}\,dx = \frac{e^{ax}}{a}.$

13. $\int b^{ax}\,dx = \frac{b^{ax}}{a \log b},\quad b > 0.$

14. $\int \log x\,dx = x \log x - x.$

15. $\int a^x\,dx = \frac{a^x}{\log a},\quad a > 0.$

16. $\int \frac{dx}{a^2+x^2} = \frac{1}{a} \tan^{-1} \frac{x}{a}.$

17. $\int \frac{dx}{a^2-x^2} = \begin{cases} \frac{1}{a} \tanh^{-1} \frac{x}{a}, \\ \text{or} \\ \frac{1}{2a} \log \frac{a+x}{a-x}, \quad a^2 > x^2. \end{cases}$

18. $$\int \frac{dx}{x^2-a^2} = \begin{cases} -\frac{1}{a}\coth^{-1}\frac{x}{a}, \\ \text{or} \\ \frac{1}{2a}\log\frac{x-a}{x+a}, \qquad x^2>a^2. \end{cases}$$

19. $$\int \frac{dx}{\sqrt{a^2-x^2}} = \begin{cases} \sin^{-1}\frac{x}{|a|}, \\ \text{or} \\ -\cos^{-1}\frac{x}{|a|}, \qquad a^2>x^2. \end{cases}$$

20. $$\int \frac{dx}{\sqrt{x^2\pm a^2}} = \log\left(x+\sqrt{x^2\pm a^2}\right).$$

21. $$\int \frac{dx}{x\sqrt{x^2-a^2}} = \frac{1}{|a|}\sec^{-1}\frac{x}{a}.$$

22. $$\int \frac{dx}{x\sqrt{a^2\pm x^2}} = -\frac{1}{a}\log\left(\frac{a+\sqrt{a^2\pm x^2}}{x}\right).$$

5.4.2 FORMS CONTAINING $a + bx$

23. $$\int (a+bx)^n\,dx = \frac{(a+bx)^{n+1}}{(n+1)b}, \quad n\neq -1.$$

24. $$\int x(a+bx)^n\,dx = \frac{1}{b^2(n+2)}(a+bx)^{n+2} - \frac{a}{b^2(n+1)}(a+bx)^{n+1}, \quad n\neq -1,\ n\neq -2.$$

25. $$\int x^2(a+bx)^n\,dx = \frac{1}{b^3}\left[\frac{(a+bx)^{n+3}}{n+3} - 2a\frac{(a+bx)^{n+2}}{n+2} + a^2\frac{(a+bx)^{n+1}}{n+1}\right], \quad n\neq -1, \quad n\neq -2,\ n\neq -3.$$

26. $$\int x^m(a+bx)^n\,dx = \begin{cases} \frac{x^{m+1}(a+bx)^n}{m+n+1} + \frac{an}{m+n+1}\int x^m(a+bx)^{n-1}\,dx, \\ \text{or} \\ \frac{1}{a(n+1)}\left[-x^{m+1}(a+bx)^{n+1} + (m+n+2)\int x^m(a+bx)^{n+1}\,dx\right], \\ \text{or} \\ \frac{1}{b(m+n+1)}\left[x^m(a+bx)^{n+1} - ma\int x^{m-1}(a+bx)^n\,dx\right]. \end{cases}$$

27. $$\int \frac{dx}{a+bx} = \frac{1}{b}\log|a+bx|.$$

28. $$\int \frac{dx}{(a+bx)^2} = -\frac{1}{b(a+bx)}.$$

29. $$\int \frac{dx}{(a+bx)^3} = -\frac{1}{2b(a+bx)^2}.$$

30. $$\int \frac{x}{a+bx}\,dx = \begin{cases} \frac{1}{b^2}\left[a+bx-a\log(a+bx)\right], \\ \text{or} \\ \frac{x}{b} - \frac{a}{b^2}\log(a+bx). \end{cases}$$

31. $\displaystyle\int \frac{x}{(a+bx)^2}\,dx = \frac{1}{b^2}\left[\log(a+bx) + \frac{a}{a+bx}\right].$

32. $\displaystyle\int \frac{x}{(a+bx)^n}\,dx = \frac{1}{b^2}\left[\frac{-1}{(n-2)(a+bx)^{n-2}} + \frac{a}{(n-1)(a+bx)^{n-1}}\right],$
$n \neq 1,\ n \neq 2.$

33. $\displaystyle\int \frac{x^2}{a+bx}\,dx = \frac{1}{b^3}\left(\frac{1}{2}(a+bx)^2 - 2a(a+bx) + a^2\log(a+bx)\right).$

34. $\displaystyle\int \frac{x^2}{(a+bx)^2}\,dx = \frac{1}{b^3}\left(a+bx - 2a\log(a+bx) - \frac{a^2}{a+bx}\right).$

35. $\displaystyle\int \frac{x^2}{(a+bx)^3}\,dx = \frac{1}{b^3}\left(\log(a+bx) + \frac{2a}{a+bx} - \frac{a^2}{2(a+bx)^2}\right).$

36. $\displaystyle\int \frac{x^2}{(a+bx)^n}\,dx = \frac{1}{b^3}\left[\frac{-1}{(n-3)(a+bx)^{n-3}} + \frac{2a}{(n-2)(a+bx)^{n-2}} - \frac{a^2}{(n-1)(a+bx)^{n-1}}\right],\quad n\neq 1,\ n\neq 2,\ n\neq 3.$

37. $\displaystyle\int \frac{dx}{x(a+bx)} = -\frac{1}{a}\log\frac{a+bx}{x}.$

38. $\displaystyle\int \frac{dx}{x(a+bx)^2} = \frac{1}{a(a+bx)} - \frac{1}{a^2}\log\frac{a+bx}{x}.$

39. $\displaystyle\int \frac{dx}{x(a+bx)^3} = \frac{1}{a^3}\left[\frac{1}{2}\left(\frac{2a+bx}{a+bx}\right)^2 - \log\frac{a+bx}{x}\right].$

40. $\displaystyle\int \frac{dx}{x^2(a+bx)} = -\frac{1}{ax} + \frac{b}{a^2}\log\frac{a+bx}{x}.$

41. $\displaystyle\int \frac{dx}{x^3(a+bx)} = \frac{2bx-a}{2a^2x^2} + \frac{b^2}{a^3}\log\frac{x}{a+bx}.$

42. $\displaystyle\int \frac{dx}{x^2(a+bx)^2} = -\frac{a+2bx}{a^2x(a+bx)} + \frac{2b}{a^3}\log\frac{a+bx}{x}.$

5.4.3 FORMS CONTAINING $c^2 \pm x^2$ AND $x^2 - c^2$

43. $\displaystyle\int \frac{dx}{c^2+x^2} = \frac{1}{c}\tan^{-1}\frac{x}{c}.$

44. $\displaystyle\int \frac{dx}{c^2-x^2} = \frac{1}{2c}\log\frac{c+x}{c-x},\quad c^2 > x^2.$

45. $\displaystyle\int \frac{dx}{x^2-c^2} = \frac{1}{2c}\log\frac{x-c}{x+c},\quad x^2 > c^2.$

46. $\displaystyle\int \frac{x}{c^2\pm x^2}\,dx = \pm\frac{1}{2}\log(c^2\pm x^2).$

47. $\displaystyle\int \frac{x}{(c^2\pm x^2)^{n+1}}\,dx = \mp\frac{1}{2n(c^2\pm x^2)^n},\quad n\neq 0.$

48. $\displaystyle\int \frac{dx}{(c^2\pm x^2)^n} = \frac{1}{2c^2(n-1)}\left[\frac{x}{(c^2\pm x^2)^{n-1}} + (2n-3)\int\frac{dx}{(c^2\pm x^2)^{n-1}}\right].$

49. $\displaystyle\int \frac{dx}{(x^2-c^2)^n} = \frac{1}{2c^2(n-1)}\left[-\frac{x}{(x^2-c^2)^{n-1}} - (2n-3)\int\frac{dx}{(x^2-c^2)^{n-1}}\right].$

50. $\displaystyle\int \frac{x}{x^2-c^2}\,dx = \frac{1}{2}\log\left(x^2-c^2\right).$

51. $\displaystyle\int \frac{x}{(x^2-c^2)^{n+1}}\,dx = -\frac{1}{2n(x^2-c^2)^n}.$

5.4.4 FORMS CONTAINING $a+bx$ AND $c+dx$

$u = a+bx, \quad v = c+dx, \quad$ and $\; k = ad-bc. \quad$ (If $k=0$, then $v = (c/a)u$.)

52. $\displaystyle\int \frac{dx}{uv} = \frac{1}{k}\log\left(\frac{v}{u}\right).$

53. $\displaystyle\int \frac{x}{uv}\,dx = \frac{1}{k}\left(\frac{a}{b}\log u - \frac{c}{d}\log v\right).$

54. $\displaystyle\int \frac{dx}{u^2v} = \frac{1}{k}\left(\frac{1}{u}+\frac{d}{k}\log\frac{v}{u}\right).$

55. $\displaystyle\int \frac{x}{u^2v}\,dx = -\frac{a}{bku} - \frac{c}{k^2}\log\frac{v}{u}.$

56. $\displaystyle\int \frac{x^2}{u^2v}\,dx = \frac{a^2}{b^2ku} + \frac{1}{k^2}\left(\frac{c^2}{d}\log v + \frac{a(k-bc)}{b^2}\log u\right).$

57. $\displaystyle\int \frac{dx}{u^nv^m} = \frac{1}{k(m-1)}\left[\frac{-1}{u^{n-1}v^{m-1}} - b(m+n-2)\int\frac{dx}{u^nv^{m-1}}\right].$

58. $\displaystyle\int \frac{u}{v}\,dx = \frac{bx}{d} + \frac{k}{d^2}\log v.$

59. $$\int \frac{u^m}{v^n}\,dx = \begin{cases} -\dfrac{1}{k(n-1)}\left[\dfrac{u^{m+1}}{v^{n-1}} + b(n-m-2)\displaystyle\int\frac{u^m}{v^{n-1}}\,dx\right], \\ \text{or} \\ -\dfrac{1}{d(n-m-1)}\left[\dfrac{u^m}{v^{n-1}} + mk\displaystyle\int\frac{u^{m-1}}{v^n}\,dx\right], \\ \text{or} \\ -\dfrac{1}{d(n-1)}\left[\dfrac{u^m}{v^{n-1}} - mb\displaystyle\int\frac{u^{m-1}}{v^{n-1}}\,dx\right]. \end{cases}$$

5.4.5 FORMS CONTAINING $a+bx^n$

60. $\displaystyle\int \frac{dx}{a+bx^2} = \frac{1}{\sqrt{ab}}\tan^{-1}\frac{x\sqrt{ab}}{a}, \quad ab>0.$

61. $$\int \frac{dx}{a+bx^2} = \begin{cases} \dfrac{1}{2\sqrt{-ab}}\log\dfrac{a+x\sqrt{-ab}}{a-x\sqrt{-ab}}, & ab<0, \\ \text{or} \\ \dfrac{1}{\sqrt{-ab}}\tanh^{-1}\dfrac{x\sqrt{-ab}}{a}, & ab<0. \end{cases}$$

62. $\displaystyle\int \frac{dx}{a^2+b^2x^2}\,dx = \frac{1}{ab}\tan^{-1}\frac{bx}{a}.$

63. $\displaystyle\int \frac{x}{a+bx^2}\,dx = \frac{1}{2b}\log\left(a+bx^2\right).$

64. $\displaystyle\int \frac{x^2}{a+bx^2}\,dx = \frac{x}{b} - \frac{a}{b}\int \frac{dx}{a+bx^2}.$

65. $\displaystyle\int \frac{dx}{(a+bx^2)^2} = \frac{x}{2a(a+bx^2)} + \frac{1}{2a}\int \frac{dx}{a+bx^2}.$

66. $\displaystyle\int \frac{dx}{a^2-b^2x^2} = \frac{1}{2ab}\log\frac{a+bx}{a-bx}.$

67. $\displaystyle\int \frac{dx}{(a+bx^2)^{m+1}} =$

$$\begin{cases} \dfrac{1}{2ma}\dfrac{x}{(a+bx^2)^m} + \dfrac{2m-1}{2ma}\displaystyle\int \dfrac{dx}{(a+bx^2)^m}, \\ \text{or} \\ \dfrac{(2m)!}{(m!)^2}\left[\dfrac{x}{2a}\displaystyle\sum_{r=1}^{m}\dfrac{r!(r-1)!}{(4a)^{m-r}(2r)!(a+bx^2)^r} + \dfrac{1}{(4a)^m}\displaystyle\int\dfrac{dx}{a+bx^2}\right]. \end{cases}$$

68. $\displaystyle\int \frac{x\,dx}{(a+bx^2)^{m+1}} = -\frac{1}{2bm(a+bx^2)^m}, \quad m \neq 0.$

69. $\displaystyle\int \frac{x^2\,dx}{(a+bx^2)^{m+1}} = -\frac{x}{2mb(a+bx^2)^m} + \frac{1}{2mb}\int \frac{dx}{(a+bx^2)^m}, \quad m \neq 0.$

70. $\displaystyle\int \frac{dx}{x(a+bx^2)} = \frac{1}{2a}\log\frac{x^2}{a+bx^2}.$

71. $\displaystyle\int \frac{dx}{x^2(a+bx^2)} = -\frac{1}{ax} - \frac{b}{a}\int \frac{dx}{a+bx^2}.$

72. $$\int \frac{dx}{x\,(a+bx^2)^{m+1}} = \begin{cases} \dfrac{1}{2am(a+bx^2)^m} + \dfrac{1}{a}\displaystyle\int \dfrac{dx}{x\,(a+bx^2)^m}, \\ \text{or} \\ \dfrac{1}{2a^{m+1}}\left[\displaystyle\sum_{r=1}^{m}\dfrac{a^r}{r(a+bx^2)^r} + \log\dfrac{x^2}{a+bx^2}\right]. \end{cases}$$

73. $\displaystyle\int \frac{dx}{x^2\,(a+bx^2)^{m+1}} = \frac{1}{a}\int \frac{dx}{x^2(a+bx^2)^m} - \frac{b}{a}\int \frac{dx}{(a+bx^2)^{m+1}}.$

74. $\displaystyle\int \frac{dx}{a+bx^3} = \frac{k}{3a}\left[\frac{1}{2}\log\frac{(k+x)^3}{a+bx^3} + \sqrt{3}\tan^{-1}\frac{2x-k}{k\sqrt{3}}\right], \quad k = \sqrt[3]{\tfrac{a}{b}}.$

75. $\displaystyle\int \frac{x\,dx}{a+bx^3} = \frac{1}{3bk}\left[\frac{1}{2}\log\frac{a+bx^3}{(k+x)^3} + \sqrt{3}\tan^{-1}\frac{2x-k}{k\sqrt{3}}\right], \quad k = \sqrt[3]{\tfrac{a}{b}}.$

76. $\displaystyle\int \frac{x^2\,dx}{a+bx^3} = \frac{1}{3b}\log a + bx^3.$

77. $\displaystyle\int \frac{dx}{a+bx^4} =$

$$\begin{cases} \dfrac{k}{2a}\left[\dfrac{1}{2}\log\dfrac{x^2+2kx+2k^2}{x^2-2kx+2k^2} + \tan^{-1}\dfrac{2kx}{2k^2-x^2}\right], & ab>0,\ k=\left(\dfrac{a}{4b}\right)^{1/4}, \\ \text{or} \\ \dfrac{k}{2a}\left[\dfrac{1}{2}\log\dfrac{x+k}{x-k} + \tan^{-1}\dfrac{x}{k}\right], & ab<0,\ k=\left(-\dfrac{a}{b}\right)^{1/4}. \end{cases}$$

78. $\displaystyle\int \frac{x}{a+bx^4}\,dx = \frac{1}{2bk}\tan^{-1}\frac{x^2}{k}, \quad ab>0,\ k=\sqrt{\tfrac{a}{b}}.$

79. $\displaystyle\int \frac{x}{a+bx^4}\,dx = \frac{1}{4bk}\log\frac{x^2-k}{x^2+k}, \quad ab<0,\ k=\sqrt{-\tfrac{a}{b}}.$

80. $\displaystyle\int \frac{x^2}{a+bx^4}\,dx = \frac{1}{4bk}\left[\frac{1}{2}\log\frac{x^2-2kx+2k^2}{x^2+2kx+2k^2}+\tan^{-1}\frac{2kx}{2k^2-x^2}\right],$

$ab>0,\ k=\left(\frac{a}{4b}\right)^{1/4}.$

81. $\displaystyle\int \frac{x^2\,dx}{a+bx^4} = \frac{1}{4bk}\left[\log\frac{x-k}{x+k}+2\tan^{-1}\frac{x}{k}\right],\quad ab<0,\ k=\sqrt[4]{-\tfrac{a}{b}}.$

82. $\displaystyle\int \frac{x^3\,dx}{a+bx^4} = \frac{1}{4b}\log\left(a+bx^4\right).$

83. $\displaystyle\int \frac{dx}{x(a+bx^n)} = \frac{1}{an}\log\frac{x^n}{a+bx^n},\quad n\neq 0.$

84. $\displaystyle\int \frac{dx}{(a+bx^n)^{m+1}} = \frac{1}{a}\int\frac{dx}{(a+bx^n)^m}-\frac{b}{a}\int\frac{x^n\,dx}{(a+bx^n)^{m+1}}.$

85. $\displaystyle\int \frac{x^m\,dx}{(a+bx^n)^{p+1}} = \frac{1}{b}\int\frac{x^{m-n}\,dx}{(a+bx^n)^p}-\frac{a}{b}\int\frac{x^{m-n}\,dx}{(a+bx^n)^{p+1}}.$

86. $\displaystyle\int \frac{dx}{x^m(a+bx^n)^{p+1}} = \frac{1}{a}\int\frac{dx}{x^m(a+bx^n)^p}-\frac{b}{a}\int\frac{dx}{x^{m-n}(a+bx^n)^{p+1}}.$

87. $\displaystyle\int x^m(a+bx^n)^p\,dx =$

$$\begin{cases}\dfrac{1}{b(np+m+1)}\left[x^{m-n+1}(a+bx^n)^{p+1}-a(m-n+1)\displaystyle\int x^{m-n}(a+bx^n)^p\,dx\right],\\ \text{or}\\ \dfrac{1}{np+m+1}\left[x^{m+1}(a+bx^n)^p+anp\displaystyle\int x^m(a+bx^n)^{p-1}\,dx\right],\\ \text{or}\\ \dfrac{1}{a(m+1)}\left[x^{m+1}(a+bx^n)^{p+1}-b(m+1+np+n)\displaystyle\int x^{m+n}(a+bx^n)^p\,dx\right],\\ \text{or}\\ \dfrac{1}{an(p+1)}\left[-x^{m+1}(a+bx^n)^{p+1}+(m+1+np+n)\displaystyle\int x^m(a+bx^n)^{p+1}\,dx\right].\end{cases}$$

5.4.6 FORMS CONTAINING $c^3 \pm x^3$

88. $\displaystyle\int \frac{dx}{c^3\pm x^3} = \pm\frac{1}{6c^2}\log\left(\frac{(c\pm x)^3}{c^3\pm x^3}\right)+\frac{1}{c^2\sqrt{3}}\tan^{-1}\frac{2x\mp c}{c\sqrt{3}}.$

89. $\displaystyle\int \frac{dx}{(c^3\pm x^3)^2} = \frac{x}{3c^3(c^3\pm x^3)}+\frac{2}{3c^3}\int\frac{dx}{c^3\pm x^3}.$

90. $\displaystyle\int \frac{dx}{(c^3\pm x^3)^{n+1}} = \frac{1}{3nc^3}\left[\frac{x}{(c^3\pm x^3)^n}+(3n-1)\int\frac{dx}{(c^3\pm x^3)^n}\right],\quad n\neq 0.$

91. $\displaystyle\int \frac{x\,dx}{c^3\pm x^3} = \frac{1}{6c}\log\frac{c^3\pm x^3}{(c\pm x)^3}\pm\frac{1}{c\sqrt{3}}\tan^{-1}\frac{2x\mp c}{c\sqrt{3}}.$

92. $\displaystyle\int \frac{x\,dx}{(c^3\pm x^3)^2} = \frac{x^2}{3c^3(c^3\pm x^3)}+\frac{1}{3c^3}\int\frac{x\,dx}{c^3\pm x^3}.$

93. $\displaystyle\int \frac{x\,dx}{(c^3\pm x^3)^{n+1}} = \frac{1}{3nc^3}\left[\frac{x^2}{(c^3\pm x^3)^n}+(3n-2)\int\frac{x\,dx}{(c^3\pm x^3)^n}\right],\quad n\neq 0.$

94. $\displaystyle\int \frac{x^2\,dx}{c^3 \pm x^3} = \pm\frac{1}{3}\log\left(c^3 \pm x^3\right).$

95. $\displaystyle\int \frac{x^2\,dx}{(c^3 \pm x^3)^{n+1}} = \mp\frac{1}{3n(c^3 \pm x^3)^n}, \quad n \neq 0.$

96. $\displaystyle\int \frac{dx}{x(c^3 \pm x^3)} = \frac{1}{3c^3}\log\frac{x^3}{c^3 \pm x^3}.$

97. $\displaystyle\int \frac{dx}{x(c^3 \pm x^3)^2} = \frac{1}{3c^3(c^3 \pm x^3)} + \frac{1}{3c^6}\log\frac{x^3}{c^3 \pm x^3}.$

98. $\displaystyle\int \frac{dx}{x(c^3 \pm x^3)^{n+1}} = \frac{1}{3nc^3(c^3 \pm x^3)^n} + \frac{1}{c^3}\int \frac{dx}{x(c^3 \pm x^3)^n}, \quad n \neq 0.$

99. $\displaystyle\int \frac{dx}{x^2(c^3 \pm x^3)} = -\frac{1}{c^3 x} \mp \frac{1}{c^3}\int \frac{x\,dx}{(c^3 \pm x^3)}.$

100. $\displaystyle\int \frac{dx}{x^2(c^3 \pm x^3)^{n+1}} = \frac{1}{c^3}\int \frac{dx}{x^2(c^3 \pm x^3)^n} \mp \frac{1}{c^3}\int \frac{x\,dx}{(c^3 \pm x^3)^{n+1}}.$

5.4.7 FORMS CONTAINING $c^4 \pm x^4$

101. $\displaystyle\int \frac{dx}{c^4 + x^4} = \frac{1}{2c^3\sqrt{2}}\left[\frac{1}{2}\log\left(\frac{x^2 + cx\sqrt{2} + c^2}{x^2 - cx\sqrt{2} + c^2}\right) + \tan^{-1}\frac{cx\sqrt{2}}{c^2 - x^2}\right].$

102. $\displaystyle\int \frac{dx}{c^4 - x^4} = \frac{1}{2c^3}\left[\frac{1}{2}\log\frac{c+x}{c-x} + \tan^{-1}\frac{x}{c}\right].$

103. $\displaystyle\int \frac{x\,dx}{c^4 + x^4} = \frac{1}{2c^2}\tan^{-1}\frac{x^2}{c^2}.$

104. $\displaystyle\int \frac{x\,dx}{c^4 - x^4} = \frac{1}{4c^2}\log\frac{c^2 + x^2}{c^2 - x^2}.$

105. $\displaystyle\int \frac{x^2\,dx}{c^4 + x^4} = \frac{1}{2c\sqrt{2}}\left[\frac{1}{2}\log\left(\frac{x^2 - cx\sqrt{2} + c^2}{x^2 + cx\sqrt{2} + c^2}\right) + \tan^{-1}\frac{cx\sqrt{2}}{c^2 - x^2}\right].$

106. $\displaystyle\int \frac{x^2\,dx}{c^4 - x^4} = \frac{1}{2c}\left[\frac{1}{2}\log\frac{c+x}{c-x} - \tan^{-1}\frac{x}{c}\right].$

107. $\displaystyle\int \frac{x^3\,dx}{c^4 \pm x^4} = \pm\frac{1}{4}\log\left(c^4 \pm x^4\right).$

5.4.8 FORMS CONTAINING $a + bx + cx^2$

$$X = a + bx + cx^2 \quad \text{and} \quad q = 4ac - b^2.$$

If $q = 0$, then $X = c\left(x + \frac{b}{2c}\right)^2$ and other formulae should be used.

108. $$\int \frac{dx}{X} = \begin{cases} \dfrac{2}{\sqrt{q}} \tan^{-1} \dfrac{2cx+b}{\sqrt{q}}, & q > 0, \\ \text{or} & \\ \dfrac{-2}{\sqrt{-q}} \tanh^{-1} \dfrac{2cx+b}{\sqrt{-q}}, & q < 0, \\ \text{or} & \\ \dfrac{1}{\sqrt{-q}} \log \dfrac{2cx+b-\sqrt{-q}}{2cx+b+\sqrt{-q}}, & q < 0. \end{cases}$$

109. $$\int \frac{dx}{X^2} = \frac{2cx+b}{qX} + \frac{2c}{q} \int \frac{dx}{X}.$$

110. $$\int \frac{dx}{X^3} = \frac{2cx+b}{q}\left(\frac{1}{2X^2} + \frac{3c}{qX}\right) + \frac{6c^2}{q^2} \int \frac{dx}{X}.$$

111. $$\int \frac{dx}{X^{n+1}} = \begin{cases} \dfrac{2cx+b}{nqX^n} + \dfrac{2(2n-1)c}{qn} \displaystyle\int \frac{dx}{X^n}, \\ \text{or} \\ \dfrac{(2n)!}{(n!)^2}\left(\dfrac{c}{q}\right)^n \left[\dfrac{2cx+b}{q} \displaystyle\sum_{r=1}^{n} \left(\frac{q}{cX}\right)^r \left(\frac{(r-1)!r!}{(2r)!}\right) + \int \frac{dx}{X}\right]. \end{cases}$$

112. $$\int \frac{x\,dx}{X} = \frac{1}{2c} \log X - \frac{b}{2c} \int \frac{dx}{X}.$$

113. $$\int \frac{x\,dx}{X^2} = -\frac{bx+2a}{qX} - \frac{b}{q} \int \frac{dx}{X}.$$

114. $$\int \frac{x\,dx}{X^{n+1}} = -\frac{2a+bx}{nqX^n} - \frac{b(2n-1)}{nq} \int \frac{dx}{X^n}, \quad n \neq 0.$$

115. $$\int \frac{x^2\,dx}{X} = \frac{x}{c} - \frac{b}{2c^2} \log X + \frac{b^2-2ac}{2c^2} \int \frac{dx}{X}.$$

116. $$\int \frac{x^2\,dx}{X^2} = \frac{(b^2-2ac)x+ab}{cqX} + \frac{2a}{q} \int \frac{dx}{X}.$$

117. $$\int \frac{x^m\,dx}{X^{n+1}} = -\frac{x^{m-1}}{(2n-m+1)cX^n} - \frac{n-m+1}{2n-m+1}\frac{b}{c} \int \frac{x^{m-1}}{X^{n+1}}\,dx + \frac{m-1}{2n-m+1}\frac{a}{c} \int \frac{x^{m-2}}{X^{n+1}}\,dx.$$

118. $$\int \frac{dx}{xX} = \frac{1}{2a} \log \frac{x^2}{X} - \frac{b}{2a} \int \frac{dx}{X}.$$

119. $$\int \frac{dx}{x^2X} = \frac{b}{2a^2} \log \frac{X}{x^2} - \frac{1}{ax} + \left(\frac{b^2}{2a^2} - \frac{c}{a}\right) \int \frac{dx}{X}.$$

120. $$\int \frac{dx}{xX^n} = \frac{1}{2a(n-1)X^{n-1}} - \frac{b}{2a} \int \frac{dx}{X^n} + \frac{1}{a} \int \frac{dx}{xX^{n-1}}, \quad n \neq 1.$$

121. $$\int \frac{dx}{x^m X^{n+1}} = -\frac{1}{(m-1)ax^{m-1}X^n} - \frac{n+m-1}{m-1}\frac{b}{a}\int \frac{dx}{x^{m-1}X^{n+1}} - \frac{2n+m-1}{m-1}\frac{c}{a}\int \frac{dx}{x^{m-2}X^{n+1}}.$$

5.4.9 FORMS CONTAINING $\sqrt{a+bx}$

122. $$\int \sqrt{a+bx}\,dx = \frac{2}{3b}\sqrt{(a+bx)^3}.$$

123. $$\int x\sqrt{a+bx}\,dx = -\frac{2(2a-3bx)}{15b^2}\sqrt{(a+bx)^3}.$$

124. $$\int x^2\sqrt{a+bx}\,dx = \frac{2(8a^2-12abx+15b^2x^2)}{105b^3}\sqrt{(a+bx)^3}.$$

125. $$\int x^m\sqrt{a+bx}\,dx = \begin{cases} \dfrac{2}{b(2m+3)}\left[x^m\sqrt{(a+bx)^3} - ma\displaystyle\int x^{m-1}\sqrt{a+bx}\,dx\right], \\ \text{or} \\ \dfrac{2}{b^{m+1}}\sqrt{a+bx}\displaystyle\sum_{r=0}^{m}\frac{m!(-a)^{m-r}}{r!(m-r)!(2r+3)}(a+bx)^{r+1}. \end{cases}$$

126. $$\int \frac{\sqrt{a+bx}}{x}\,dx = 2\sqrt{a+bx} + a\int \frac{dx}{x\sqrt{a+bx}}.$$

127. $$\int \frac{\sqrt{a+bx}}{x^m}\,dx = -\frac{1}{(m-1)a}\left[\frac{\sqrt{(a+bx)^3}}{x^{m-1}} + \frac{(2m-5)b}{2}\int \frac{\sqrt{a+bx}}{x^{m-1}}\,dx\right].$$

128. $$\int \frac{dx}{\sqrt{a+bx}} = \frac{2\sqrt{a+bx}}{b}.$$

129. $$\int \frac{x\,dx}{\sqrt{a+bx}} = -\frac{2(2a-bx)}{3b^2}\sqrt{a+bx}.$$

130. $$\int \frac{x^2\,dx}{\sqrt{a+bx}} = \frac{2(8a^2-4abx+3b^2x^2)}{15b^3}\sqrt{a+bx}.$$

131. $$\int \frac{x^m\,dx}{\sqrt{a+bx}} = \begin{cases} \dfrac{2}{(2m+1)b}\left[x^m\sqrt{a+bx} - ma\displaystyle\int \frac{x^{m-1}}{\sqrt{a+bx}}\,dx\right], \\ \text{or} \\ \dfrac{2(-a)^m\sqrt{a+bx}}{b^{m+1}}\displaystyle\sum_{r=0}^{m}\frac{(-1)^r m!(a+bx)^r}{(2r+1)r!(m-r)!a^r}. \end{cases}$$

132. $$\int \frac{dx}{x\sqrt{a+bx}} = \begin{cases} \dfrac{2}{\sqrt{-a}}\tan^{-1}\sqrt{\dfrac{a+bx}{-a}}, & a<0, \\ \text{or} \\ \dfrac{1}{\sqrt{a}}\log\left(\dfrac{\sqrt{a+bx}-\sqrt{a}}{\sqrt{a+bx}+\sqrt{a}}\right), & a>0. \end{cases}$$

133. $$\int \frac{dx}{x^2\sqrt{a+bx}} = -\frac{\sqrt{a+bx}}{ax} - \frac{b}{2a}\int \frac{dx}{x\sqrt{a+bx}}.$$

134. $\displaystyle\int \frac{dx}{x^n\sqrt{a+bx}} =$

$$\begin{cases} -\dfrac{\sqrt{a+bx}}{(n-1)ax^{n-1}} - \dfrac{(2n-3)b}{(2n-2)a}\displaystyle\int \frac{dx}{x^{n-1}\sqrt{a+bx}}, \\ \text{or} \\ \dfrac{(2n-2)!}{[(n-1)!]^2}\left[-\dfrac{\sqrt{a+bx}}{a}\displaystyle\sum_{r=1}^{n-1}\frac{r!(r-1)!}{x^r(2r)!}\left(-\frac{b}{4a}\right)^{n-r-1} + \left(-\frac{b}{4a}\right)^{n-1}\int\frac{dx}{x\sqrt{a+bx}}\right]. \end{cases}$$

135. $\displaystyle\int (a+bx)^{\pm n/2}\,dx = \frac{2(a+bx)^{(2\pm n)/2}}{b(2\pm n)}.$

136. $\displaystyle\int x(a+bx)^{\pm n/2}\,dx = \frac{2}{b^2}\left[\frac{(a+bx)^{(4\pm n)/2}}{4\pm n} - \frac{a(a+bx)^{(2\pm n)/2}}{2\pm n}\right].$

137. $\displaystyle\int \frac{dx}{x(a+bx)^{n/2}} = \frac{1}{a}\int\frac{dx}{x(a+bx)^{(n-2)/2}} - \frac{b}{a}\int\frac{dx}{(a+bx)^{n/2}}.$

138. $\displaystyle\int \frac{(a+bx)^{n/2}}{x}\,dx = b\int (a+bx)^{(n-2)/2}\,dx + a\int\frac{(a+bx)^{(n-2)/2}}{x}\,dx.$

5.4.10 FORMS CONTAINING $\sqrt{a+bx}$ AND $\sqrt{c+dx}$

$$u = a+bx, \qquad v = c+dx, \qquad k = ad-bc.$$

If $k=0$, then $v = \frac{c}{a}u$, and other formulae should be used.

139. $$\int\frac{dx}{\sqrt{uv}} = \begin{cases} \dfrac{2}{\sqrt{bd}}\tanh^{-1}\dfrac{\sqrt{bduv}}{bv}, & bd>0,\ k<0, \\ \dfrac{2}{\sqrt{bd}}\tanh^{-1}\dfrac{\sqrt{bduv}}{du}, & bd>0,\ k>0, \\ \dfrac{1}{\sqrt{bd}}\log\dfrac{(bv+\sqrt{bduv})^2}{v}, & bd>0, \\ \dfrac{2}{\sqrt{-bd}}\tan^{-1}\dfrac{\sqrt{-bduv}}{bv}, & bd<0, \\ -\dfrac{1}{\sqrt{-bd}}\sin^{-1}\left(\dfrac{2bdx+ad+bc}{|k|}\right), & bd<0. \end{cases}$$

140. $\displaystyle\int\sqrt{uv}\,dx = \frac{k+2bv}{4bd}\sqrt{uv} - \frac{k^2}{8bd}\int\frac{dx}{\sqrt{uv}}.$

141. $$\int\frac{dx}{v\sqrt{u}} = \begin{cases} \dfrac{1}{\sqrt{kd}}\log\dfrac{d\sqrt{u}-\sqrt{kd}}{d\sqrt{u}+\sqrt{kd}}, & kd>0, \\ \text{or} \\ \dfrac{1}{\sqrt{kd}}\log\dfrac{\left(d\sqrt{u}-\sqrt{kd}\right)^2}{v}, & kd>0, \\ \text{or} \\ \dfrac{2}{\sqrt{-kd}}\tan^{-1}\dfrac{d\sqrt{u}}{\sqrt{-kd}}, & kd<0. \end{cases}$$

142. $\displaystyle\int\frac{x\,dx}{\sqrt{uv}} = \frac{\sqrt{uv}}{bd} - \frac{ad+bc}{2bd}\int\frac{dx}{\sqrt{uv}}.$

143. $\displaystyle\int \frac{dx}{v\sqrt{uv}} = -\frac{2\sqrt{uv}}{kv}.$

144. $\displaystyle\int \frac{v\,dx}{\sqrt{uv}} = \frac{\sqrt{uv}}{b} - \frac{k}{2b}\int \frac{dx}{\sqrt{uv}}.$

145. $\displaystyle\int \sqrt{\frac{v}{u}}\,dx = \frac{v}{|v|}\int \frac{v\,dx}{\sqrt{uv}}.$

146. $\displaystyle\int v^m\sqrt{u}\,dx = \frac{1}{(2m+3)d}\left(2v^{m+1}\sqrt{u} + k\int \frac{v^m\,dx}{\sqrt{u}}\right).$

147. $\displaystyle\int \frac{dx}{v^m\sqrt{u}} = -\frac{1}{(m-1)k}\left(\frac{\sqrt{u}}{v^{m-1}} + \left(m-\frac{3}{2}\right)b\int \frac{dx}{v^{m-1}\sqrt{u}}\right), \quad m \neq 1.$

148. $\displaystyle\int \frac{v^m}{\sqrt{u}}\,dx = \begin{cases} \dfrac{2}{b(2m+1)}\left(v^m\sqrt{u} - mk\displaystyle\int \frac{v^{m-1}}{\sqrt{u}}\,dx\right), \\ \text{or} \\ \dfrac{2(m!)^2\sqrt{u}}{b(2m+1)!}\displaystyle\sum_{r=0}^{m}\left(-\frac{4k}{b}\right)^{m-r}\frac{(2r)!}{(r!)^2}v^r. \end{cases}$

5.4.11 FORMS CONTAINING $\sqrt{x^2 \pm a^2}$

149. $\displaystyle\int \sqrt{x^2 \pm a^2}\,dx = \frac{1}{2}\left[x\sqrt{x^2 \pm a^2} \pm a^2 \log\left(x + \sqrt{x^2 \pm a^2}\right)\right].$

150. $\displaystyle\int \frac{dx}{\sqrt{x^2 \pm a^2}} = \log\left(x + \sqrt{x^2 \pm a^2}\right).$

151. $\displaystyle\int \frac{dx}{x\sqrt{x^2 - a^2}} = \frac{1}{|a|}\sec^{-1}\frac{x}{a}.$

152. $\displaystyle\int \frac{dx}{x\sqrt{x^2 + a^2}} = -\frac{1}{a}\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$

153. $\displaystyle\int \frac{\sqrt{x^2 + a^2}}{x}\,dx = \sqrt{x^2 + a^2} - a\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$

154. $\displaystyle\int \frac{\sqrt{x^2 - a^2}}{x}\,dx = \sqrt{x^2 - a^2} - |a|\sec^{-1}\frac{x}{a}.$

155. $\displaystyle\int \frac{x}{\sqrt{x^2 \pm a^2}}\,dx = \sqrt{x^2 \pm a^2}.$

156. $\displaystyle\int x\sqrt{x^2 \pm a^2}\,dx = \frac{1}{3}\sqrt{(x^2 \pm a^2)^3}.$

157. $\displaystyle\int \sqrt{(x^2 \pm a^2)^3}\,dx = \frac{1}{4}\left[x\sqrt{(x^2 \pm a^2)^3} \pm \frac{3a^2x}{2}\sqrt{x^2 \pm a^2} + \frac{3a^4}{2}\log\left(x + \sqrt{x^2 \pm a^2}\right)\right].$

158. $\displaystyle\int \frac{dx}{\sqrt{(x^2 \pm a^2)^3}} = \frac{\pm x}{a^2\sqrt{x^2 \pm a^2}}.$

159. $\displaystyle\int \frac{x}{\sqrt{(x^2 \pm a^2)^3}}\,dx = \frac{-1}{\sqrt{x^2 \pm a^2}}.$

160. $\displaystyle\int x\sqrt{(x^2 \pm a^2)^3}\,dx = \frac{1}{5}\sqrt{(x^2 \pm a^2)^5}.$

161. $\displaystyle\int x^2\sqrt{x^2 \pm a^2}\,dx = \frac{x}{4}\sqrt{(x^2 \pm a^2)^3} \mp \frac{a^2}{8}x\sqrt{x^2 \pm a^2} - \frac{a^4}{8}\log\left(x + \sqrt{x^2 \pm a^2}\right).$

162. $\displaystyle\int x^3\sqrt{x^2 + a^2}\,dx = \frac{1}{15}(3x^2 - 2a^2)\sqrt{(x^2 + a^2)^3}.$

163. $\displaystyle\int x^3\sqrt{x^2 - a^2}\,dx = \frac{1}{5}\sqrt{(x^2 - a^2)^5} + \frac{a^2}{3}\sqrt{(x^2 - a^2)^3}.$

164. $\displaystyle\int \frac{x^2}{\sqrt{x^2 \pm a^2}}\,dx = \frac{x}{2}\sqrt{x^2 \pm a^2} \mp \frac{a^2}{2}\log\left(x + \sqrt{x^2 \pm a^2}\right).$

165. $\displaystyle\int \frac{x^3}{\sqrt{x^2 \pm a^2}}\,dx = \frac{1}{3}\sqrt{(x^2 \pm a^2)^3} \mp a^2\sqrt{x^2 \pm a^2}.$

166. $\displaystyle\int \frac{dx}{x^2\sqrt{x^2 \pm a^2}}\,dx = \mp\frac{\sqrt{x^2 \pm a^2}}{a^2x}.$

167. $\displaystyle\int \frac{dx}{x^3\sqrt{x^2 + a^2}}\,dx = -\frac{\sqrt{x^2 + a^2}}{2a^2x^2} + \frac{1}{2a^3}\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$

168. $\displaystyle\int \frac{dx}{x^3\sqrt{x^2 - a^2}}\,dx = \frac{\sqrt{x^2 - a^2}}{2a^2x^2} + \frac{1}{2|a|^3}\sec^{-1}\frac{x}{a}.$

169. $\displaystyle\int x^2\sqrt{(x^2 \pm a^2)^3}\,dx = \frac{x}{6}\sqrt{(x^2 \pm a^2)^5} \mp \frac{a^2x}{24}\sqrt{(x^2 \pm a^2)^3} - \frac{a^4x}{16}\sqrt{x^2 \pm a^2} \mp \frac{a^6}{16}\log\left(x + \sqrt{x^2 \pm a^2}\right).$

170. $\displaystyle\int x^3\sqrt{(x^2 \pm a^2)^3}\,dx = \frac{1}{7}\sqrt{(x^2 \pm a^2)^7} \mp \frac{a^2}{5}\sqrt{(x^2 \pm a^2)^5}.$

171. $\displaystyle\int \frac{\sqrt{x^2 \pm a^2}}{x^2}\,dx = -\frac{\sqrt{x^2 \pm a^2}}{x} + \log\left(x + \sqrt{x^2 \pm a^2}\right).$

172. $\displaystyle\int \frac{\sqrt{x^2 + a^2}}{x^3}\,dx = -\frac{\sqrt{x^2 + a^2}}{2x^2} - \frac{1}{2a}\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$

173. $\displaystyle\int \frac{\sqrt{x^2 - a^2}}{x^3}\,dx = -\frac{\sqrt{x^2 - a^2}}{2x^2} + \frac{1}{2\,|a|}\sec^{-1}\frac{x}{a}.$

174. $\displaystyle\int \frac{\sqrt{x^2 \pm a^2}}{x^4}\,dx = \mp\frac{\sqrt{(x^2 \pm a^2)^3}}{3a^2x^3}.$

175. $\displaystyle\int \frac{x^2\,dx}{\sqrt{(x^2 \pm a^2)^3}} = -\frac{x}{\sqrt{x^2 \pm a^2}} + \log\left(x + \sqrt{x^2 \pm a^2}\right).$

176. $\displaystyle\int \frac{x^3\,dx}{\sqrt{(x^2 \pm a^2)^3}} = \sqrt{x^2 \pm a^2} \pm \frac{a^2}{\sqrt{x^2 \pm a^2}}.$

177. $\displaystyle\int \frac{dx}{x\sqrt{(x^2 + a^2)^3}} = \frac{1}{a^2\sqrt{x^2 + a^2}} - \frac{1}{a^3}\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$

178. $\displaystyle\int \frac{dx}{x\sqrt{(x^2 - a^2)^3}} = -\frac{1}{a^2\sqrt{x^2 - a^2}} - \frac{1}{|a^3|}\sec^{-1}\frac{x}{a}.$

179. $\displaystyle\int \frac{dx}{x^2\sqrt{(x^2 \pm a^2)^3}} = -\frac{1}{a^4}\left[\frac{\sqrt{x^2 \pm a^2}}{x} + \frac{x}{\sqrt{x^2 \pm a^2}}\right].$

180.
$$\int \frac{dx}{x^3\sqrt{(x^2 + a^2)^3}} = -\frac{1}{2a^2x^2\sqrt{x^2 + a^2}} - \frac{3}{2a^4\sqrt{x^2 + a^2}} + \frac{3}{2a^5}\log\left(\frac{a + \sqrt{x^2 + a^2}}{x}\right).$$

181. $\displaystyle\int \frac{dx}{x^3\sqrt{(x^2-a^2)^3}} = \frac{1}{2a^2x^2\sqrt{x^2-a^2}} - \frac{3}{2a^4\sqrt{x^2-a^2}} - \frac{3}{2\,|a^5|}\sec^{-1}\frac{x}{a}.$

182. $\displaystyle\int \frac{x^m\,dx}{\sqrt{x^2\pm a^2}} = \frac{1}{m}x^{m-1}\sqrt{x^2\pm a^2} \mp \frac{m-1}{m}a^2\int \frac{x^{m-2}}{\sqrt{x^2\pm a^2}}\,dx.$

183. $\displaystyle\int \frac{x^{2m}\,dx}{\sqrt{x^2\pm a^2}} = \frac{(2m)!}{2^{2m}(m!)^2}\left[\sqrt{x^2\pm a^2}\sum_{r=1}^{m}\frac{r!(r-1)!}{(2r)!}(\mp a^2)^{m-r}(2x)^{2r-1} + (\mp a^2)^m\log\left(x+\sqrt{x^2\pm a^2}\right)\right].$

184. $\displaystyle\int \frac{x^{2m+1}\,dx}{\sqrt{x^2\pm a^2}} = \sqrt{x^2\pm a^2}\sum_{r=0}^{m}\frac{(2r)!(m!)^2}{(2m+1)!(r!)^2}(\mp 4a^2)^{m-r}x^{2r}.$

185. $\displaystyle\int \frac{dx}{x^m\sqrt{x^2\pm a^2}} = \mp\frac{\sqrt{x^2\pm a^2}}{(m-1)a^2x^{m-1}} \mp \frac{(m-2)}{(m-1)a^2}\int \frac{dx}{x^{m-2}\sqrt{x^2\pm a^2}}.$

186. $\displaystyle\int \frac{dx}{x^{2m}\sqrt{x^2\pm a^2}} = \sqrt{x^2\pm a^2}\sum_{r=0}^{m-1}\frac{(m-1)!m!(2r)!2^{2m-2r-1}}{(r!)^2(2m)!(\mp a^2)^{m-r}x^{2r+1}}.$

187. $\displaystyle\int \frac{dx}{x^{2m+1}\sqrt{x^2+a^2}} = \frac{(2m)!}{(m!)^2}\left[\frac{\sqrt{x^2+a^2}}{a^2}\sum_{r=1}^{m}(-1)^{m-r+1}\frac{r!(r-1)!}{2(2r)!(4a^2)^{m-r}x^{2r}} + \frac{(-1)^{m+1}}{2^{2m}a^{2m+1}}\log\left(\frac{\sqrt{x^2+a^2}+a}{x}\right)\right].$

188. $\displaystyle\int \frac{dx}{x^{2m+1}\sqrt{x^2-a^2}} = \frac{(2m)!}{(m!)^2}\left[\frac{\sqrt{x^2-a^2}}{a^2}\sum_{r=1}^{m}\frac{r!(r-1)!}{2(2r)!(4a^2)^{m-r}x^{2r}} + \frac{1}{2^{2m}\,|a|^{2m+1}}\sec^{-1}\frac{x}{a}\right].$

189. $\displaystyle\int \frac{dx}{(x-a)\sqrt{x^2-a^2}} = -\frac{\sqrt{x^2-a^2}}{a(x-a)}.$

190. $\displaystyle\int \frac{dx}{(x+a)\sqrt{x^2-a^2}} = \frac{\sqrt{x^2-a^2}}{a(x+a)}.$

5.4.12 FORMS CONTAINING $\sqrt{a^2-x^2}$

191. $\displaystyle\int \sqrt{a^2-x^2}\,dx = \frac{1}{2}\left(x\sqrt{a^2-x^2} + a^2\sin^{-1}\frac{x}{|a|}\right).$

192. $\displaystyle\int \frac{dx}{\sqrt{a^2-x^2}} = \sin^{-1}\frac{x}{|a|} = -\cos^{-1}\frac{x}{|a|}.$

193. $\displaystyle\int \frac{dx}{x\sqrt{a^2-x^2}} = -\frac{1}{a}\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$

194. $\displaystyle\int \frac{\sqrt{a^2-x^2}}{x}\,dx = \sqrt{a^2-x^2} - a\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$

195. $\displaystyle\int \frac{x}{\sqrt{a^2-x^2}}\,dx = -\sqrt{a^2-x^2}.$

196. $\displaystyle\int x\sqrt{a^2-x^2}\,dx = -\frac{1}{3}\sqrt{(a^2-x^2)^3}.$

197. $\int \sqrt{(a^2-x^2)^3}\,dx = \frac{1}{4}\left(x\sqrt{(a^2-x^2)^3} + \frac{3a^2x}{2}\sqrt{a^2-x^2} + \frac{3a^4}{2}\sin^{-1}\frac{x}{|a|}\right).$

198. $\int \frac{dx}{\sqrt{(a^2-x^2)^3}} = \frac{x}{a^2\sqrt{a^2-x^2}}.$

199. $\int \frac{x}{\sqrt{(a^2-x^2)^3}}\,dx = \frac{1}{\sqrt{a^2-x^2}}.$

200. $\int x\sqrt{(a^2-x^2)^3}\,dx = -\frac{1}{5}\sqrt{(a^2-x^2)^5}.$

201. $\int x^2\sqrt{a^2-x^2}\,dx = -\frac{x}{4}\sqrt{(a^2-x^2)^3} + \frac{a^2}{8}\left(x\sqrt{a^2-x^2} + a^2\sin^{-1}\frac{x}{|a|}\right).$

202. $\int x^3\sqrt{a^2-x^2}\,dx = \left(-\frac{1}{5}x^2 - \frac{2}{15}a^2\right)\sqrt{(a^2-x^2)^3}.$

203. $\int x^2\sqrt{(a^2-x^2)^3}\,dx = -\frac{1}{6}x\sqrt{(a^2-x^2)^5} + \frac{a^2x}{24}\sqrt{(a^2-x^2)^3} + \frac{a^4x}{16}\sqrt{a^2-x^2} + \frac{a^6}{16}\sin^{-1}\frac{x}{|a|}.$

204. $\int x^3\sqrt{(a^2-x^2)^3}\,dx = \frac{1}{7}\sqrt{(a^2-x^2)^7} - \frac{a^2}{5}\sqrt{(a^2-x^2)^5}.$

205. $\int \frac{x^2}{\sqrt{a^2-x^2}}\,dx = -\frac{x}{2}\sqrt{a^2-x^2} + \frac{a^2}{2}\sin^{-1}\frac{x}{|a|}.$

206. $\int \frac{dx}{x^2\sqrt{a^2-x^2}} = -\frac{\sqrt{a^2-x^2}}{a^2x}.$

207. $\int \frac{\sqrt{a^2-x^2}}{x^2}\,dx = -\frac{\sqrt{a^2-x^2}}{x} - \sin^{-1}\frac{x}{|a|}.$

208. $\int \frac{\sqrt{a^2-x^2}}{x^3}\,dx = -\frac{\sqrt{a^2-x^2}}{2x^2} + \frac{1}{2a}\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$

209. $\int \frac{\sqrt{a^2-x^2}}{x^4}\,dx = -\frac{\sqrt{(a^2-x^2)^3}}{3a^2x^3}.$

210. $\int \frac{x^2\,dx}{\sqrt{(a^2-x^2)^3}} = \frac{x}{\sqrt{a^2-x^2}} - \sin^{-1}\frac{x}{|a|}.$

211. $\int \frac{x^3\,dx}{\sqrt{a^2-x^2}} = -\frac{2}{3}\sqrt{(a^2-x^2)^3} - x^2\sqrt{a^2-x^2}.$

212. $\int \frac{x^3\,dx}{\sqrt{(a^2-x^2)^3}} = 2\sqrt{a^2-x^2} + \frac{x^2}{\sqrt{a^2-x^2}} = \frac{a^2}{\sqrt{a^2-x^2}} + \sqrt{a^2-x^2}.$

213. $\int \frac{dx}{x^3\sqrt{a^2-x^2}} = -\frac{\sqrt{a^2-x^2}}{2a^2x^2} - \frac{1}{2a^3}\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$

214. $\int \frac{dx}{x\sqrt{(a^2-x^2)^3}} = \frac{1}{a^2\sqrt{a^2-x^2}} - \frac{1}{a^3}\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$

215. $\int \frac{dx}{x^2\sqrt{(a^2-x^2)^3}} = \frac{1}{a^4}\left(-\frac{\sqrt{a^2-x^2}}{x} + \frac{x}{\sqrt{a^2-x^2}}\right).$

216.
$$\int \frac{dx}{x^3\sqrt{(a^2-x^2)^3}} = -\frac{1}{2a^2x^2\sqrt{a^2-x^2}} + \frac{3}{2a^4\sqrt{a^2-x^2}} - \frac{3}{2a^5}\log\left(\frac{a+\sqrt{a^2-x^2}}{x}\right).$$

217. $\displaystyle\int \frac{x^m}{\sqrt{a^2-x^2}}\,dx = -\frac{x^{m-1}\sqrt{a^2-x^2}}{m} + \frac{(m-1)a^2}{m}\int \frac{x^{m-2}}{\sqrt{a^2-x^2}}\,dx.$

218. $\displaystyle\int \frac{x^{2m}}{\sqrt{a^2-x^2}}\,dx = \frac{(2m)!}{(m!)^2}\left[-\sqrt{a^2-x^2}\sum_{r=1}^{m}\frac{r!(r-1)!}{2^{2m-2r+1}(2r)!}a^{2m-2r}x^{2r-1} + \frac{a^{2m}}{2^{2m}}\sin^{-1}\frac{x}{|a|}\right].$

219. $\displaystyle\int \frac{x^{2m+1}}{\sqrt{a^2-x^2}}\,dx = -\sqrt{a^2-x^2}\sum_{r=0}^{m}\frac{(2r)!(m!)^2}{(2m+1)!(r!)^2}(4a^2)^{m-r}x^{2r}.$

220. $\displaystyle\int \frac{dx}{x^m\sqrt{a^2-x^2}} = -\frac{\sqrt{a^2-x^2}}{(m-1)a^2x^{m-1}} + \frac{(m-2)}{(m-1)a^2}\int \frac{dx}{x^{m-2}\sqrt{a^2-x^2}}.$

221. $\displaystyle\int \frac{dx}{x^{2m}\sqrt{a^2-x^2}} = -\sqrt{a^2-x^2}\sum_{r=0}^{m-1}\frac{(m-1)!m!(2r)!2^{2m-2r-1}}{(r!)^2(2m)!a^{2m-2r}x^{2r+1}}.$

222. $\displaystyle\int \frac{dx}{x^{2m+1}\sqrt{a^2-x^2}} = \frac{(2m)!}{(m!)^2}\left[-\frac{\sqrt{a^2-x^2}}{a^2}\sum_{r=1}^{m}\frac{r!(r-1)!}{2(2r)!(4a^2)^{m-r}x^{2r}} + \frac{1}{2^{2m}a^{2m+1}}\log\left(\frac{a-\sqrt{a^2-x^2}}{x}\right)\right].$

223. $\displaystyle\int \frac{dx}{(b^2-x^2)\sqrt{a^2-x^2}} = \begin{cases} \dfrac{1}{2b\sqrt{a^2-b^2}}\log\left(\dfrac{(b\sqrt{a^2-x^2}+x\sqrt{a^2-b^2})^2}{b^2-x^2}\right), & a^2>b^2, \\ \text{or} & \\ \dfrac{1}{b\sqrt{b^2-a^2}}\tan^{-1}\dfrac{x\sqrt{b^2-a^2}}{b\sqrt{a^2-x^2}}, & b^2>a^2. \end{cases}$

224. $\displaystyle\int \frac{dx}{(b^2+x^2)\sqrt{a^2-x^2}} = \frac{1}{b\sqrt{a^2+b^2}}\tan^{-1}\frac{x\sqrt{a^2+b^2}}{b\sqrt{a^2-x^2}}.$

225. $\displaystyle\int \frac{\sqrt{a^2-x^2}}{b^2+x^2}\,dx = \frac{\sqrt{a^2+b^2}}{|b|}\sin^{-1}\frac{x\sqrt{a^2+b^2}}{|a|\sqrt{x^2+b^2}} - \sin^{-1}\frac{x}{|a|},\quad b^2>a^2.$

5.4.13 FORMS CONTAINING $\sqrt{a+bx+cx^2}$

$X = a+bx+cx^2$, $q = 4ac-b^2$, and $k = 4c/q$.
If $q=0$, then $\sqrt{X} = \sqrt{c}\left|x+\frac{b}{2c}\right|$.

226. $\displaystyle\int \frac{dx}{\sqrt{X}} = \begin{cases} \dfrac{1}{\sqrt{c}}\log\left(\dfrac{2\sqrt{cX}+2cx+b}{\sqrt{q}}\right), & c>0, \\ \text{or} & \\ \dfrac{1}{\sqrt{c}}\sinh^{-1}\dfrac{2cx+b}{\sqrt{q}}, & c>0, \\ \text{or} & \\ -\dfrac{1}{\sqrt{-c}}\sin^{-1}\dfrac{2cx+b}{\sqrt{-q}}, & c<0. \end{cases}$

227. $\displaystyle\int \frac{dx}{X\sqrt{X}} = \frac{2(2cx+b)}{q\sqrt{X}}.$

228. $\displaystyle\int \frac{dx}{X^2\sqrt{X}} = \frac{2(2cx+b)}{3q\sqrt{X}}\left(\frac{1}{X}+2k\right).$

229. $$\int \frac{dx}{X^n\sqrt{X}} = \begin{cases} \dfrac{2(2cx+b)\sqrt{X}}{(2n-1)qX^n} + \dfrac{2k(n-1)}{2n-1}\displaystyle\int \frac{dx}{X^{n-1}\sqrt{X}}, \\ \quad \text{or} \\ \dfrac{(2cx+b)(n!)(n-1)!4^n k^{n-1}}{q(2n)!\sqrt{X}} \displaystyle\sum_{r=0}^{n-1} \frac{(2r)!}{(4kX)^r(r!)^2}. \end{cases}$$

230. $\displaystyle\int \sqrt{X}\,dx = \frac{(2cx+b)\sqrt{X}}{4c} + \frac{1}{2k}\int \frac{dx}{\sqrt{X}}.$

231. $\displaystyle\int X\sqrt{X}\,dx = \frac{(2cx+b)\sqrt{X}}{8c}\left(X+\frac{3}{2k}\right) + \frac{3}{8k^2}\int \frac{dx}{\sqrt{X}}.$

232. $\displaystyle\int X^2\sqrt{X}\,dx = \frac{(2cx+b)\sqrt{X}}{12c}\left(X^2+\frac{5X}{4k}+\frac{15}{8k^2}\right) + \frac{5}{16k^3}\int \frac{dx}{\sqrt{X}}.$

233. $\displaystyle\int X^n\sqrt{X}\,dx =$
$$\begin{cases} \dfrac{(2cx+b)X^n\sqrt{X}}{4(n+1)c} + \dfrac{2n+1}{2(n+1)k}\displaystyle\int X^{n-1}\sqrt{X}\,dx, \\ \quad \text{or} \\ \dfrac{(2n+2)!}{\left[(n+1)!\right]^2(4k)^{n+1}}\left[\dfrac{k(2cx+b)\sqrt{X}}{c}\displaystyle\sum_{r=0}^{n}\frac{r!(r+1)!(4kX)^r}{(2r+2)!} + \int \frac{dx}{\sqrt{X}}\right]. \end{cases}$$

234. $\displaystyle\int \frac{x\,dx}{\sqrt{X}} = \frac{\sqrt{X}}{c} - \frac{b}{2c}\int \frac{dx}{\sqrt{X}}.$

235. $\displaystyle\int \frac{x\,dx}{X\sqrt{X}} = -\frac{2(bx+2a)}{q\sqrt{X}}.$

236. $\displaystyle\int \frac{x\,dx}{X^n\sqrt{X}} = -\frac{\sqrt{X}}{(2n-1)cX^n} - \frac{b}{2c}\int \frac{dx}{X^n\sqrt{X}}.$

237. $\displaystyle\int \frac{x^2\,dx}{\sqrt{X}} = \left(\frac{x}{2c}-\frac{3b}{4c^2}\right)\sqrt{X} + \frac{3b^2-4ac}{8c^2}\int \frac{dx}{\sqrt{X}}.$

238. $\displaystyle\int \frac{x^2\,dx}{X\sqrt{X}} = \frac{(2b^2-4ac)x+2ab}{cq\sqrt{X}} + \frac{1}{c}\int \frac{dx}{\sqrt{X}}.$

239. $\displaystyle\int \frac{x^2\,dx}{X^n\sqrt{X}} = \frac{(2b^2-4ac)x+2ab}{(2n-1)cqX^{n-1}\sqrt{X}} + \frac{4ac+(2n-3)b^2}{(2n-1)cq}\int \frac{dx}{X^{n-1}\sqrt{X}}.$

240. $\displaystyle\int \frac{x^3\,dx}{\sqrt{X}} = \left(\frac{x^2}{3c}-\frac{5bx}{12c^2}+\frac{5b^2}{8c^3}-\frac{2a}{3c^2}\right)\sqrt{X} + \left(\frac{3ab}{4c^2}-\frac{5b^3}{16c^3}\right)\int \frac{dx}{\sqrt{X}}.$

241. $\displaystyle\int \frac{x^n\,dx}{\sqrt{X}} = \frac{1}{nc}x^{n-1}\sqrt{X} - \frac{(2n-1)b}{2nc}\int \frac{x^{n-1}\,dx}{\sqrt{X}} - \frac{(n-1)a}{nc}\int \frac{x^{n-2}\,dx}{\sqrt{X}}.$

242. $\displaystyle\int x\sqrt{X}\,dx = \frac{X\sqrt{X}}{3c} - \frac{b(2cx+b)}{8c^2}\sqrt{X} - \frac{b}{4ck}\int \frac{dx}{\sqrt{X}}.$

243. $\displaystyle\int xX\sqrt{X}\,dx = \frac{X^2\sqrt{X}}{5c} - \frac{b}{2c}\int X\sqrt{X}\,dx.$

244. $\displaystyle\int xX^n\sqrt{X}\,dx = \frac{X^{n+1}\sqrt{X}}{(2n+3)c} - \frac{b}{2c}\int X^n\sqrt{X}\,dx.$

245. $\int x^2\sqrt{X}\,dx = \left(x - \frac{5b}{6c}\right)\frac{X\sqrt{X}}{4c} + \frac{5b^2-4ac}{16c^2}\int\sqrt{X}\,dx.$

246. $$\int \frac{dx}{x\sqrt{X}} = \begin{cases} \frac{1}{\sqrt{-a}}\sin^{-1}\left(\frac{bx+2a}{|x|\sqrt{-q}}\right), & a<0, \\ \text{or} & \\ -\frac{2\sqrt{X}}{bx}, & a=0, \\ \text{or} & \\ -\frac{1}{\sqrt{a}}\log\left(\frac{2\sqrt{aX}+bx+2a}{x}\right), & a>0. \end{cases}$$

247. $\int \frac{dx}{x^2\sqrt{X}} = -\frac{\sqrt{X}}{ax} - \frac{b}{2a}\int\frac{dx}{x\sqrt{X}}.$

248. $\int \frac{\sqrt{X}}{x}\,dx = \sqrt{X} + \frac{b}{2}\int\frac{dx}{\sqrt{X}} + a\int\frac{dx}{x\sqrt{X}}.$

249. $\int \frac{\sqrt{X}}{x^2}\,dx = -\frac{\sqrt{X}}{x} + \frac{b}{2}\int\frac{dx}{x\sqrt{X}} + c\int\frac{dx}{\sqrt{X}}.$

5.4.14 FORMS CONTAINING $\sqrt{2ax - x^2}$

250. $\int \sqrt{2ax-x^2}\,dx = \frac{1}{2}\left[(x-a)\sqrt{2ax-x^2} + a^2\sin^{-1}\frac{x-a}{|a|}\right].$

251. $$\int \frac{dx}{\sqrt{2ax-x^2}} = \begin{cases} \cos^{-1}\left(\frac{a-x}{|a|}\right), \\ \text{or} \\ \sin^{-1}\left(\frac{x-a}{|a|}\right). \end{cases}$$

252. $\int x^n\sqrt{2ax-x^2}\,dx =$

$$\begin{cases} -\frac{x^{n-1}\sqrt{(2ax-x^2)^3}}{n+2} + \frac{(2n+1)a}{n+2}\int x^{n-1}\sqrt{2ax-x^2}\,dx, \\ \text{or} \\ \sqrt{2ax-x^2}\left[\frac{x^{n+1}}{n+2} - \sum_{r=0}^{n}\frac{(2n+1)!(r!)^2a^{n-r+1}}{2^{n-r}(2r+1)!(n+2)!n!}x^r\right] + \frac{(2n+1)!a^{n+2}}{2^n n!(n+2)!}\sin^{-1}\left(\frac{x-a}{|a|}\right) \end{cases}$$

253. $\int \frac{\sqrt{2ax-x^2}}{x^n}\,dx = \frac{\sqrt{(2ax-x^2)^3}}{(3-2n)ax^n} + \frac{n-3}{(2n-3)a}\int\frac{\sqrt{2ax-x^2}}{x^{n-1}}\,dx.$

254. $\int \frac{x^n\,dx}{\sqrt{2ax-x^2}} =$

$$\begin{cases} -\frac{x^{n-1}\sqrt{2ax-x^2}}{n} + \frac{a(2n-1)}{n}\int\frac{x^{n-1}}{\sqrt{2ax-x^2}}\,dx, \\ \text{or} \\ -\sqrt{2ax-x^2}\sum_{r=1}^{n}\frac{(2n)!r!(r-1)!a^{n-r}}{2^{n-r}(2r)!(n!)^2}x^{r-1} + \frac{(2n)!a^n}{2^n(n!)^2}\sin^{-1}\left(\frac{x-a}{|a|}\right). \end{cases}$$

255. $$\int \frac{dx}{x^n\sqrt{2ax-x^2}} = \begin{cases} \dfrac{\sqrt{2ax-x^2}}{a(1-2n)x^n} + \dfrac{n-1}{(2n-1)a}\displaystyle\int \frac{dx}{x^{n-1}\sqrt{2ax-x^2}}, \\ \text{or} \\ -\sqrt{2ax-x^2}\displaystyle\sum_{r=0}^{n-1} \frac{2^{n-r}(n-1)!n!(2r)!}{(2n)!(r!)^2 a^{n-r}x^{r+1}}. \end{cases}$$

256. $$\int \frac{dx}{\sqrt{(2ax-x^2)^3}} = \frac{x-a}{a^2\sqrt{2ax-x^2}}.$$

257. $$\int \frac{x\,dx}{\sqrt{(2ax-x^2)^3}} = \frac{x}{a\sqrt{2ax-x^2}}.$$

5.4.15 MISCELLANEOUS ALGEBRAIC FORMS

258. $$\int \frac{dx}{\sqrt{2ax+x^2}} = \log\left(x+a+\sqrt{2ax+x^2}\right).$$

259. $$\int \sqrt{ax^2+c}\,dx = \begin{cases} \dfrac{x}{2}\sqrt{ax^2+c} + \dfrac{c}{2\sqrt{-a}}\sin^{-1}\left(x\sqrt{-\dfrac{a}{c}}\right), & a<0, \\ \text{or} \\ \dfrac{x}{2}\sqrt{ax^2+c} + \dfrac{c}{2\sqrt{a}}\log\left(x\sqrt{a}+\sqrt{ax^2+c}\right), & a>0. \end{cases}$$

260. $$\int \sqrt{\frac{1+x}{1-x}}\,dx = \sin^{-1}x - \sqrt{1-x^2}.$$

261. $$\int \frac{dx}{x\sqrt{ax^n+c}} = \begin{cases} \dfrac{1}{n\sqrt{c}}\log\dfrac{\sqrt{ax^n+c}-\sqrt{c}}{\sqrt{ax^n+c}+\sqrt{c}}, \\ \text{or} \\ \dfrac{2}{n\sqrt{c}}\log\dfrac{\sqrt{ax^n+c}-\sqrt{c}}{\sqrt{x^n}}, & c>0, \\ \text{or} \\ \dfrac{2}{n\sqrt{-c}}\sec^{-1}\sqrt{-\dfrac{ax^n}{c}}, & c<0. \end{cases}$$

262. $$\int \frac{dx}{\sqrt{ax^2+c}} = \begin{cases} \dfrac{1}{\sqrt{-a}}\sin^{-1}\left(x\sqrt{-\dfrac{a}{c}}\right), & a<0, \\ \text{or} \\ \dfrac{1}{\sqrt{a}}\log\left(x\sqrt{a}+\sqrt{ax^2+c}\right), & a>0. \end{cases}$$

263. $$\int (ax^2+c)^{m+1/2}\,dx = $$
$$\begin{cases} \dfrac{x(ax^2+c)^{m+1/2}}{2(m+1)} + \dfrac{(2m+1)c}{2(m+1)}\displaystyle\int (ax^2+c)^{m-1/2}\,dx, \\ \text{or} \\ x\sqrt{ax^2+c}\displaystyle\sum_{r=0}^{m} \frac{(2m+1)!(r!)^2c^{m-r}}{2^{2m-2r+1}m!(m+1)!(2r+1)!}(ax^2+c)^r \\ \qquad + \dfrac{(2m+1)!c^{m+1}}{2^{2m+1}m!(m+1)!}\displaystyle\int \frac{dx}{\sqrt{ax^2+c}}. \end{cases}$$

264. $$\int x(ax^2+c)^{m+1/2}\,dx = \frac{(ax^2+c)^{m+3/2}}{(2m+3)a}.$$

265. $$\int \frac{(ax^2+c)^{m+1/2}}{x}\,dx = \begin{cases} \dfrac{(ax^2+c)^{m+1/2}}{2m+1} + c\displaystyle\int \frac{(ax^2+c)^{m-1/2}}{x}\,dx, \\ \text{or} \\ \sqrt{ax^2+c}\displaystyle\sum_{r=0}^{m} \frac{c^{m-r}(ax^2+c)^r}{2r+1} + c^{m+1}\int \frac{dx}{x\sqrt{ax^2+c}}. \end{cases}$$

266. $$\int \frac{dx}{(ax^2+c)^{m+1/2}} = \begin{cases} \dfrac{x}{(2m-1)c(ax^2+c)^{m-1/2}} + \dfrac{2m-2}{(2m-1)c}\displaystyle\int \frac{dx}{(ax^2+c)^{m-1/2}}, \\ \text{or} \\ \dfrac{x}{\sqrt{ax^2+c}}\displaystyle\sum_{r=0}^{m-1} \frac{2^{2m-2r-1}(m-1)!m!(2r)!}{(2m)!(r!)^2c^{m-r}(ax^2+c)^r}. \end{cases}$$

267. $$\int \frac{dx}{x^m\sqrt{ax^2+c}} = -\frac{\sqrt{ax^2+c}}{(m-1)cx^{m-1}} - \frac{(m-2)a}{(m-1)c}\int \frac{dx}{x^{m-2}\sqrt{ax^2+c}}, \quad m \neq 1.$$

268. $$\int \frac{1+x^2}{(1-x^2)\sqrt{1+x^4}}\,dx = \frac{1}{\sqrt{2}}\log\left(\frac{x\sqrt{2}+\sqrt{1+x^4}}{1-x^2}\right).$$

269. $$\int \frac{1-x^2}{(1+x^2)\sqrt{1+x^4}}\,dx = \frac{1}{\sqrt{2}}\tan^{-1}\frac{x\sqrt{2}}{\sqrt{1+x^4}}.$$

270. $$\int \frac{dx}{x\sqrt{x^n+a^2}} = -\frac{2}{na}\log\left(\frac{a+\sqrt{x^n+a^2}}{\sqrt{x^n}}\right).$$

271. $$\int \frac{dx}{x\sqrt{x^n-a^2}} = -\frac{2}{na}\sin^{-1}\frac{a}{\sqrt{x^n}}.$$

272. $$\int \sqrt{\frac{x}{a^3-x^3}}\,dx = \frac{2}{3}\sin^{-1}\left(\frac{x}{a}\right)^{3/2}.$$

5.4.16 FORMS INVOLVING TRIGONOMETRIC FUNCTIONS

273. $$\int \sin ax\,dx = -\frac{1}{a}\cos ax.$$

274. $$\int \cos ax\,dx = \frac{1}{a}\sin ax.$$

275. $$\int \tan ax\,dx = -\frac{1}{a}\log\cos ax = \frac{1}{a}\log\sec ax.$$

276. $$\int \cot ax\,dx = \frac{1}{a}\log\sin ax = -\frac{1}{a}\log\csc ax.$$

277. $$\int \sec ax\,dx = \frac{1}{a}\log(\sec ax + \tan ax) = \frac{1}{a}\log\tan\left(\frac{\pi}{4}+\frac{ax}{2}\right).$$

278. $$\int \csc ax\,dx = \frac{1}{a}\log(\csc ax - \cot ax) = \frac{1}{a}\log\tan\frac{ax}{2}.$$

279. $$\int \sin^2 ax\,dx = \frac{x}{2} - \frac{1}{2a}\cos ax\sin ax = \frac{x}{2} - \frac{1}{4a}\sin 2ax.$$

280. $\int \sin^3 ax\,dx = -\frac{1}{3a}(\cos ax)(\sin^2 ax + 2).$

281. $\int \sin^4 ax\,dx = \frac{3x}{8} - \frac{\sin 2ax}{4a} + \frac{\sin 4ax}{32a}.$

282. $\int \sin^n ax\,dx = -\frac{\sin^{n-1} ax \cos ax}{na} + \frac{n-1}{n}\int \sin^{n-2} ax\,dx.$

283. $\int \sin^{2m} ax\,dx = -\frac{\cos ax}{a}\sum_{r=0}^{m-1}\frac{(2m)!(r!)^2}{2^{2m-2r}(2r+1)!(m!)^2}\sin^{2r+1} ax + \frac{(2m)!}{2^{2m}(m!)^2}x.$

284. $\int \sin^{2m+1} ax\,dx = -\frac{\cos ax}{a}\sum_{r=0}^{m-1}\frac{2^{2m-2r}(m!)^2(2r)!}{(2m+1)!(r!)^2}\sin^{2r} ax.$

285. $\int \cos^2 ax\,dx = \frac{1}{2}x + \frac{1}{2a}\sin ax \cos ax = \frac{1}{2}x + \frac{1}{4a}\sin 2ax$

286. $\int \cos^3 ax\,dx = \frac{1}{3a}\sin ax(\cos^2 ax + 2).$

287. $\int \cos^4 ax\,dx = \frac{3}{8}x + \frac{\sin 2ax}{4a} + \frac{\sin 4ax}{32a}.$

288. $\int \cos^n ax\,dx = \frac{1}{na}\cos^{n-1} ax \sin ax + \frac{n-1}{n}\int \cos^{n-2} ax\,dx.$

289. $\int \cos^{2m} ax\,dx = \frac{\sin ax}{a}\sum_{r=0}^{m-1}\frac{(2m)!(r!)^2}{2^{2m-2r}(2r+1)!(m!)^2}\cos^{2r+1} ax + \frac{(2m)!}{2^{2m}(m!)^2}x.$

290. $\int \cos^{2m+1} ax\,dx = \frac{\sin ax}{a}\sum_{r=0}^{m}\frac{2^{2m-2r}(m!)^2(2r)!}{(2m+1)!(r!)^2}\cos^{2r} ax.$

291. $\int \frac{dx}{\sin^2 ax} = \int \operatorname{cosec}^2 ax\,dx = -\frac{1}{a}\cot ax.$

292. $\int \frac{dx}{\sin^m ax} = \int \operatorname{cosec}^m ax\,dx = -\frac{1}{a(m-1)}\frac{\cos ax}{\sin^{m-1} ax} + \frac{m-2}{m-1}\int \frac{dx}{\sin^{m-2} ax}.$

293. $\int \frac{dx}{\sin^{2m} ax} = \int \operatorname{cosec}^{2m} ax\,dx = -\frac{1}{a}\cos ax\sum_{r=0}^{m-1}\frac{2^{2m-2r-1}(m-1)!m!(2r)!}{(2m)!(r!)^2\sin^{2r+1} ax}.$

294. $\int \frac{dx}{\sin^{2m+1} ax} = \int \operatorname{cosec}^{2m+1} ax\,dx =$

$$-\frac{1}{a}\cos ax\sum_{r=0}^{m-1}\frac{(2m)!(r!)^2}{2^{2m-2r}(2r+1)!(m!)^2\sin^{2r+2} ax} + \frac{1}{a}\frac{(2m)!}{2^{2m}(m!)^2}\log\tan\frac{ax}{2}.$$

295. $\int \frac{dx}{\cos^2 ax} = \int \sec^2 ax\,dx = \frac{1}{a}\tan ax.$

296. $\int \frac{dx}{\cos^m ax} = \int \sec^m ax\,dx = \frac{1}{a(m-1)}\frac{\sin ax}{\cos^{m-1} ax} + \frac{m-2}{m-1}\int \frac{dx}{\cos^{m-2} ax}.$

297. $\int \frac{dx}{\cos^{2m} ax} = \int \sec^{2m} ax\,dx = \frac{1}{a}\sin ax\sum_{r=0}^{m-1}\frac{2^{2m-2r-1}(m-1)!m!(2r)!}{(2m)!(r!)^2\cos^{2r+1} ax}.$

298.

$$\int \frac{dx}{\cos^{2m+1} ax} = \int \sec^{2m+1} ax \, dx = \frac{1}{a} \sin ax \sum_{r=0}^{m-1} \frac{(2m)!(r!)^2}{2^{2m-2r}(m!)^2(2r+1)! \cos^{2r+2} ax} + \frac{1}{a} \frac{(2m)!}{2^{2m}(m!)^2} \log(\sec ax + \tan ax).$$

299. $\int (\sin mx)(\sin nx) \, dx = \frac{\sin(m-n)x}{2(m-n)} - \frac{\sin(m+n)x}{2(m+n)}, \quad m^2 \neq n^2.$

300. $\int (\cos mx)(\cos nx) \, dx = \frac{\sin(m-n)x}{2(m-n)} + \frac{\sin(m+n)x}{2(m+n)}, \quad m^2 \neq n^2.$

301. $\int (\sin ax)(\cos ax) \, dx = \frac{1}{2a} \sin^2 ax.$

302. $\int (\sin mx)(\cos nx) \, dx = -\frac{\cos(m-n)x}{2(m-n)} - \frac{\cos(m+n)x}{2(m+n)}, \quad m^2 \neq n^2.$

303. $\int (\sin^2 ax)(\cos^2 ax) \, dx = -\frac{1}{32a} \sin 4ax + \frac{x}{8}.$

304. $\int (\sin ax)(\cos^m ax) \, dx = -\frac{\cos^{m+1} ax}{(m+1)a}.$

305. $\int (\sin^m ax)(\cos ax) \, dx = \frac{\sin^{m+1} ax}{(m+1)a}.$

306. $\int (\cos^m ax)(\sin^n ax) \, dx =$

$$\begin{cases} \dfrac{\cos^{m-1} ax \sin^{n+1} ax}{(m+n)a} + \dfrac{m-1}{m+n} \displaystyle\int (\cos^{m-2} ax)(\sin^n ax) \, dx, \\ \text{or} \\ -\dfrac{\cos^{m+1} ax \sin^{n-1} ax}{(m+n)a} + \dfrac{n-1}{m+n} \displaystyle\int (\cos^m ax)(\sin^{n-2} ax) \, dx. \end{cases}$$

307. $$\int \frac{\cos^m ax}{\sin^n ax} \, dx = \begin{cases} -\dfrac{\cos^{m+1} ax}{a(n-1) \sin^{n-1} ax} - \dfrac{m-n+2}{n-1} \displaystyle\int \frac{\cos^m ax}{\sin^{n-2} ax} \, dx, \\ \text{or} \\ \dfrac{\cos^{m-1} ax}{a(m-n) \sin^{n-1} ax} + \dfrac{m-1}{m-n} \displaystyle\int \frac{\cos^{m-2} ax}{\sin^n ax} \, dx. \end{cases}$$

308. $$\int \frac{\sin^m ax}{\cos^n ax} \, dx = \begin{cases} \dfrac{\sin^{m+1} ax}{a(n-1) \cos^{n-1} ax} - \dfrac{m-n+2}{n-1} \displaystyle\int \frac{\sin^m ax}{\cos^{n-2} ax} \, dx, \\ \text{or} \\ -\dfrac{\sin^{m-1} ax}{a(m-n) \cos^{n-1} ax} + \dfrac{m-1}{m-n} \displaystyle\int \frac{\sin^{m-2} ax}{\cos^n ax} \, dx. \end{cases}$$

309. $\int \frac{\sin ax}{\cos^2 ax} \, dx = \frac{1}{a \cos ax} = \frac{\sec ax}{a}.$

310. $\int \frac{\sin^2 ax}{\cos ax} \, dx = -\frac{1}{a} \sin ax + \frac{1}{a} \log \tan\left(\frac{\pi}{4} + \frac{ax}{2}\right).$

311. $\int \frac{\cos ax}{\sin^2 ax} \, dx = -\frac{\csc ax}{a} = -\frac{1}{a \sin ax}.$

312. $\int \frac{dx}{(\sin ax)(\cos ax)} = \frac{1}{a} \log \tan ax.$

313. $$\int \frac{dx}{(\sin ax)(\cos^2 ax)} = \frac{1}{a}\left(\sec ax + \log\tan\frac{ax}{2}\right).$$

314. $$\int \frac{dx}{(\sin ax)(\cos^n ax)} = \frac{1}{a(n-1)\cos^{n-1} ax} + \int \frac{dx}{(\sin ax)(\cos^{n-2} ax)}.$$

315. $$\int \frac{dx}{(\sin^2 ax)(\cos ax)} = -\frac{1}{a}\csc ax + \frac{1}{a}\log\tan\left(\frac{\pi}{4} + \frac{ax}{2}\right).$$

316. $$\int \frac{dx}{(\sin^2 ax)(\cos^2 ax)} = -\frac{2}{a}\cot 2ax.$$

317. $$\int \frac{dx}{\sin^m ax\cos^n ax} = \begin{cases} -\dfrac{1}{a(m-1)\sin^{m-1} ax\cos^{n-1} ax} + \dfrac{m+n-2}{m-1}\displaystyle\int \frac{dx}{\sin^{m-2} ax\cos^n ax}, \\ \text{or} \\ \dfrac{1}{a(n-1)\sin^{m-1} ax\cos^{n-1} ax} + \dfrac{m+n-2}{n-1}\displaystyle\int \frac{dx}{\sin^m ax\cos^{n-2} ax}. \end{cases}$$

318. $$\int \sin(a+bx)\,dx = -\frac{1}{b}\cos(a+bx).$$

319. $$\int \cos(a+bx)\,dx = \frac{1}{b}\sin(a+bx).$$

320. $$\int \frac{dx}{1 \pm \sin ax} = \mp\frac{1}{a}\tan\left(\frac{\pi}{4} \mp \frac{ax}{2}\right).$$

321. $$\int \frac{dx}{1+\cos ax} = \frac{1}{a}\tan\frac{ax}{2}.$$

322. $$\int \frac{dx}{1-\cos ax} = -\frac{1}{a}\cot\frac{ax}{2}.$$

323. $$\int \frac{dx}{a+b\sin x} = \begin{cases} \dfrac{2}{\sqrt{a^2-b^2}}\tan^{-1}\left(\dfrac{a\tan\frac{x}{2}+b}{\sqrt{a^2-b^2}}\right), \\ \text{or} \\ \dfrac{1}{\sqrt{b^2-a^2}}\log\left(\dfrac{a\tan\frac{x}{2}+b-\sqrt{b^2-a^2}}{a\tan\frac{x}{2}+b+\sqrt{b^2-a^2}}\right). \end{cases}$$

324. $$\int \frac{dx}{a+b\cos x} = \begin{cases} \dfrac{2}{\sqrt{a^2-b^2}}\tan^{-1}\dfrac{\sqrt{a^2-b^2}\tan\frac{x}{2}}{a+b}, \\ \text{or} \\ \dfrac{1}{\sqrt{b^2-a^2}}\log\left(\dfrac{\sqrt{b^2-a^2}\tan\frac{x}{2}+a+b}{\sqrt{b^2-a^2}\tan\frac{x}{2}-a-b}\right). \end{cases}$$

325. $$\int \frac{dx}{a+b\sin x+c\cos x} =$$
$$\begin{cases} \dfrac{1}{\sqrt{b^2+c^2-a^2}}\log\left(\dfrac{b-\sqrt{b^2+c^2-a^2}+(a-c)\tan\frac{x}{2}}{b+\sqrt{b^2+c^2-a^2}+(a-c)\tan\frac{x}{2}}\right), & a \neq c,\ a^2 < b^2+c^2, \\ \text{or} \\ \dfrac{2}{\sqrt{a^2-b^2-c^2}}\tan^{-1}\dfrac{b+(a-c)\tan\frac{x}{2}}{\sqrt{a^2-b^2-c^2}}, & a^2 > b^2+c^2, \\ \text{or} \\ \dfrac{1}{a}\left[\dfrac{a-(b+c)\sin x-(b-c)\sin x}{a-(b+c)\sin x+(b-c)\sin x}\right]. & a^2 = b^2+c^2. \end{cases}$$

326. $\displaystyle\int \frac{\sin^2 x}{a+b\cos^2 x}\,dx = \frac{1}{b}\sqrt{\frac{a+b}{a}}\tan^{-1}\left(\sqrt{\frac{a}{a+b}}\tan x\right) - \frac{x}{b}, \quad ab>0,\ |a|>|b|.$

327. $\displaystyle\int \frac{dx}{a^2\cos^2 x + b^2\sin^2 x} = \frac{1}{ab}\tan^{-1}\left(\frac{b\tan x}{a}\right).$

328. $\displaystyle\int \frac{\cos^2 cx}{a^2+b^2\sin^2 cx}\,dx = \frac{\sqrt{a^2+b^2}}{ab^2c}\tan^{-1}\frac{\sqrt{a^2+b^2}\tan cx}{a} - \frac{x}{b^2}.$

329. $\displaystyle\int \frac{\sin cx\cos cx}{a\cos^2 cx + b\sin^2 cx}\,dx = \frac{1}{2c(b-a)}\log\left(a\cos^2 cx + b\sin^2 cx\right), \quad a\neq b.$

330. $\displaystyle\int \frac{\cos cx}{a\cos cx + b\sin cx}\,dx =$

$$\int \frac{dx}{a+b\tan cx} = \frac{1}{c(a^2+b^2)}\left[acx + b\log\left(a\cos cx + b\sin cx\right)\right].$$

331. $\displaystyle\int \frac{\sin cx}{a\cos cx + b\sin cx}\,dx =$

$$\int \frac{dx}{b+a\cot cx} = \frac{1}{c(a^2+b^2)}\left[bcx - a\log\left(a\cos cx + b\sin cx\right)\right].$$

332. $\displaystyle\int \frac{dx}{a\cos^2 x + 2b\cos x\sin x + c\sin^2 x} =$

$$\begin{cases} \dfrac{1}{2\sqrt{b^2-ac}}\log\left(\dfrac{c\tan x + b - \sqrt{b^2-ac}}{c\tan x + b + \sqrt{b^2-ac}}\right), & b^2>ac, \\ \text{or} & \\ \dfrac{1}{\sqrt{ac-b^2}}\tan^{-1}\left(\dfrac{c\tan x + b}{\sqrt{ac-b^2}}\right), & b^2<ac, \\ \text{or} & \\ -\dfrac{1}{c\tan x + b}, & b^2=ac. \end{cases}$$

333. $\displaystyle\int \frac{\sin ax}{1\pm\sin ax}\,dx = \pm x + \frac{1}{a}\tan\left(\frac{\pi}{4}\mp\frac{ax}{2}\right).$

334. $\displaystyle\int \frac{dx}{(\sin ax)(1\pm\sin ax)} = \frac{1}{a}\tan\left(\frac{\pi}{4}\mp\frac{ax}{2}\right) + \frac{1}{a}\log\tan\frac{ax}{2}.$

335. $\displaystyle\int \frac{dx}{(1+\sin ax)^2} = -\frac{1}{2a}\tan\left(\frac{\pi}{4}-\frac{ax}{2}\right) - \frac{1}{6a}\tan^3\left(\frac{\pi}{4}-\frac{ax}{2}\right).$

336. $\displaystyle\int \frac{dx}{(1-\sin ax)^2} = \frac{1}{2a}\cot\left(\frac{\pi}{4}-\frac{ax}{2}\right) + \frac{1}{6a}\cot^3\left(\frac{\pi}{4}-\frac{ax}{2}\right).$

337. $\displaystyle\int \frac{\sin ax}{(1+\sin ax)^2}\,dx = -\frac{1}{2a}\tan\left(\frac{\pi}{4}-\frac{ax}{2}\right) + \frac{1}{6a}\tan^3\left(\frac{\pi}{4}-\frac{ax}{2}\right).$

338. $\displaystyle\int \frac{\sin ax}{(1-\sin ax)^2}\,dx = -\frac{1}{2a}\cot\left(\frac{\pi}{4}-\frac{ax}{2}\right) + \frac{1}{6a}\cot^3\left(\frac{\pi}{4}-\frac{ax}{2}\right).$

339. $\displaystyle\int \frac{\sin x}{a+b\sin x}\,dx = \frac{x}{b} - \frac{a}{b}\int\frac{dx}{a+b\sin x}.$

340. $\displaystyle\int \frac{dx}{(\sin x)(a+b\sin x)} = \frac{1}{a}\log\tan\frac{x}{2} - \frac{b}{a}\int\frac{dx}{a+b\sin x}.$

341. $\displaystyle\int \frac{dx}{(a+b\sin x)^2} = \begin{cases} \dfrac{b\cos x}{(a^2-b^2)(a+b\sin x)} + \dfrac{a}{a^2-b^2}\displaystyle\int\frac{dx}{a+b\sin x}, \\ \text{or} \\ \dfrac{a\cos x}{(b^2-a^2)(a+b\sin x)} + \dfrac{b}{b^2-a^2}\displaystyle\int\frac{dx}{a+b\sin x}. \end{cases}$

342. $\displaystyle\int \frac{dx}{a^2+b^2\sin^2 cx} = \frac{1}{ac\sqrt{a^2+b^2}}\tan^{-1}\left(\frac{\sqrt{a^2+b^2}\tan cx}{a}\right).$

343. $\displaystyle\int \frac{dx}{a^2-b^2\sin^2 cx} = \begin{cases} \dfrac{1}{ac\sqrt{a^2-b^2}}\tan^{-1}\left(\dfrac{\sqrt{a^2-b^2}\tan cx}{a}\right), & a^2>b^2, \\ \text{or} & \\ \dfrac{1}{2ac\sqrt{b^2-a^2}}\log\left(\dfrac{\sqrt{b^2-a^2}\tan cx+a}{\sqrt{b^2-a^2}\tan cx-a}\right), & a^2<b^2. \end{cases}$

344. $\displaystyle\int \frac{\cos ax}{1+\cos ax}\,dx = x-\frac{1}{a}\tan\frac{ax}{2}.$

345. $\displaystyle\int \frac{\cos ax}{1-\cos ax}\,dx = -x-\frac{1}{a}\cot\frac{ax}{2}.$

346. $\displaystyle\int \frac{dx}{(\cos ax)(1+\cos ax)} = \frac{1}{a}\log\tan\left(\frac{\pi}{4}+\frac{ax}{2}\right)-\frac{1}{a}\tan\frac{ax}{2}.$

347. $\displaystyle\int \frac{dx}{(\cos ax)(1-\cos ax)} = \frac{1}{a}\log\tan\left(\frac{\pi}{4}+\frac{ax}{2}\right)-\frac{1}{a}\cot\frac{ax}{2}.$

348. $\displaystyle\int \frac{dx}{(1+\cos ax)^2} = \frac{1}{2a}\tan\frac{ax}{2}+\frac{1}{6a}\tan^3\frac{ax}{2}.$

349. $\displaystyle\int \frac{dx}{(1-\cos ax)^2} = -\frac{1}{2a}\cot\frac{ax}{2}-\frac{1}{6a}\cot^3\frac{ax}{2}.$

350. $\displaystyle\int \frac{\cos ax}{(1+\cos ax)^2}\,dx = \frac{1}{2a}\tan\frac{ax}{2}-\frac{1}{6a}\tan^3\frac{ax}{2}.$

351. $\displaystyle\int \frac{\cos ax}{(1-\cos ax)^2}\,dx = \frac{1}{2a}\cot\frac{ax}{2}-\frac{1}{6a}\cot^3\frac{ax}{2}.$

352. $\displaystyle\int \frac{\cos x}{a+b\cos x}\,dx = \frac{x}{b}-\frac{a}{b}\int\frac{dx}{a+b\cos x}.$

353. $\displaystyle\int \frac{dx}{(\cos x)(a+b\cos x)} = \frac{1}{a}\log\tan\left(\frac{x}{2}+\frac{\pi}{4}\right)-\frac{b}{a}\int\frac{dx}{a+b\cos x}.$

354. $\displaystyle\int \frac{dx}{(a+b\cos x)^2} = \frac{b\sin x}{(b^2-a^2)(a+b\cos x)}-\frac{a}{b^2-a^2}\int\frac{dx}{a+b\cos x}.$

355. $\displaystyle\int \frac{\cos x}{(a+b\cos x)^2}\,dx = \frac{a\sin x}{(a^2-b^2)(a+b\cos x)}-\frac{b}{a^2-b^2}\int\frac{dx}{a+b\cos x}.$

356. $\displaystyle\int \frac{dx}{a^2+b^2-2ab\cos cx} = \frac{2}{c(a^2-b^2)}\tan^{-1}\left(\frac{a+b}{a-b}\tan\frac{cx}{2}\right).$

357. $\displaystyle\int \frac{dx}{a^2+b^2\cos^2 cx} = \frac{1}{ac\sqrt{a^2+b^2}}\tan^{-1}\frac{a\tan cx}{\sqrt{a^2+b^2}}.$

358. $\displaystyle\int \frac{dx}{a^2-b^2\cos^2 cx} = \begin{cases} \dfrac{1}{ac\sqrt{a^2-b^2}}\tan^{-1}\left(\dfrac{a\tan cx}{\sqrt{a^2-b^2}}\right), & a^2>b^2, \\ \text{or} & \\ \dfrac{1}{2ac\sqrt{b^2-a^2}}\log\left(\dfrac{a\tan cx-\sqrt{b^2-a^2}}{a\tan cx+\sqrt{b^2-a^2}}\right), & b^2>a^2. \end{cases}$

359. $\displaystyle\int \frac{\sin ax}{1\pm\cos ax}\,dx = \mp\frac{1}{a}\log(1\pm\cos ax).$

360. $\displaystyle\int \frac{\cos ax}{1\pm\sin ax}\,dx = \pm\frac{1}{a}\log(1\pm\sin ax).$

361. $\displaystyle\int \frac{dx}{(\sin ax)(1 \pm \cos ax)} = \pm\frac{1}{2a(1 \pm \cos ax)} + \frac{1}{2a}\log\tan\frac{ax}{2}.$

362. $\displaystyle\int \frac{dx}{(\cos ax)(1 \pm \sin ax)} = \mp\frac{1}{2a(1 \pm \sin ax)} + \frac{1}{2a}\log\tan\left(\frac{ax}{2} + \frac{\pi}{4}\right).$

363. $\displaystyle\int \frac{\sin ax}{(\cos ax)(1 \pm \cos ax)}\,dx = \frac{1}{a}\log\left(\sec ax \pm 1\right).$

364. $\displaystyle\int \frac{\cos ax}{(\sin ax)(1 \pm \sin ax)}\,dx = -\frac{1}{a}\log\left(\csc ax \pm 1\right).$

365. $\displaystyle\int \frac{\sin ax}{(\cos ax)(1 \pm \sin ax)}\,dx = \frac{1}{2a(1 \pm \sin ax)} \pm \frac{1}{2a}\log\tan\left(\frac{ax}{2} + \frac{\pi}{4}\right).$

366. $\displaystyle\int \frac{\cos ax}{(\sin ax)(1 \pm \cos ax)}\,dx = -\frac{1}{2a(1 \pm \cos ax)} \pm \frac{1}{2a}\log\tan\frac{ax}{2}.$

367. $\displaystyle\int \frac{dx}{\sin ax \pm \cos ax} = \frac{1}{a\sqrt{2}}\log\tan\left(\frac{ax}{2} \pm \frac{\pi}{8}\right).$

368. $\displaystyle\int \frac{dx}{(\sin ax \pm \cos ax)^2} = \frac{1}{2a}\tan\left(ax \mp \frac{\pi}{4}\right).$

369. $\displaystyle\int \frac{dx}{1 + \cos ax \pm \sin ax} = \pm\frac{1}{a}\log\left(1 \pm \tan\frac{ax}{2}\right).$

370. $\displaystyle\int \frac{dx}{a^2\cos^2 cx - b^2\sin^2 cx} = \frac{1}{2abc}\log\left(\frac{b\tan cx + a}{b\tan cx - a}\right).$

371. $\displaystyle\int x\sin ax\,dx = \frac{1}{a^2}\sin ax - \frac{x}{a}\cos ax.$

372. $\displaystyle\int x^2\sin ax\,dx = \frac{2x}{a^2}\sin ax + \frac{2 - a^2x^2}{a^3}\cos ax.$

373. $\displaystyle\int x^3\sin ax\,dx = \frac{3a^2x^2 - 6}{a^4}\sin ax + \frac{6x - a^2x^3}{a^3}\cos ax.$

374. $\displaystyle\int x^m\sin ax\,dx =$

$$\begin{cases} -\dfrac{1}{a}x^m\cos ax + \dfrac{m}{a}\displaystyle\int x^{m-1}\cos ax\,dx, \\ \text{or} \\ \cos ax\displaystyle\sum_{r=0}^{\lfloor \frac{m}{2}\rfloor}\frac{(-1)^{r+1}m!}{(m-2r)!}\frac{x^{m-2r}}{a^{2r+1}} + \sin ax\sum_{r=0}^{\lfloor \frac{m-1}{2}\rfloor}\frac{(-1)^r m!}{(m-2r-1)!}\frac{x^{m-2r-1}}{a^{2r+2}}. \end{cases}$$

375. $\displaystyle\int x\cos ax\,dx = \frac{1}{a^2}\cos ax + \frac{x}{a}\sin ax.$

376. $\displaystyle\int x^2\cos ax\,dx = \frac{2x}{a^2}\cos ax + \frac{a^2x^2 - 2}{a^3}\sin ax.$

377. $\displaystyle\int x^3\cos ax\,dx = \frac{3a^2x^2 - 6}{a^4}\cos ax + \frac{a^2x^3 - 6x}{a^3}\sin ax.$

378. $\displaystyle\int x^m \cos ax\,dx =$

$$\begin{cases} \dfrac{x^m}{a}\sin ax - \dfrac{m}{a}\displaystyle\int x^{m-1}\sin ax\,dx, \\ \quad \text{or} \\ \sin ax \displaystyle\sum_{r=0}^{\lfloor \frac{m}{2} \rfloor} \dfrac{(-1)^r m!}{(m-2r)!}\dfrac{x^{m-2r}}{a^{2r+1}} + \cos ax \displaystyle\sum_{r=0}^{\lfloor \frac{m-1}{2} \rfloor} \dfrac{(-1)^r m!}{(m-2r-1)!}\dfrac{x^{m-2r-1}}{a^{2r+2}}. \end{cases}$$

379. $\displaystyle\int \frac{\sin ax}{x}\,dx = \sum_{n=0}^{\infty}(-1)^n \frac{(ax)^{2n+1}}{(2n+1)(2n+1)!}.$

380. $\displaystyle\int \frac{\cos ax}{x}\,dx = \sum_{n=0}^{\infty}(-1)^n \frac{(ax)^{2n}}{(2n)(2n)!}.$

381. $\displaystyle\int x\sin^2 ax\,dx = \frac{x^2}{4} - \frac{x}{4a}\sin 2ax - \frac{1}{8a^2}\cos 2ax.$

382. $\displaystyle\int x^2\sin^2 ax\,dx = \frac{x^3}{6} - \left(\frac{x^2}{4a} - \frac{1}{8a^3}\right)\sin 2ax - \frac{x}{4a^2}\cos 2ax.$

383. $\displaystyle\int x\sin^3 ax\,dx = \frac{x}{12a}\cos 3ax - \frac{1}{36a^2}\sin 3ax - \frac{3x}{4a}\cos ax + \frac{3}{4a^2}\sin ax.$

384. $\displaystyle\int x\cos^2 ax\,dx = \frac{x^2}{4} + \frac{x}{4a}\sin 2ax + \frac{1}{8a^2}\cos 2ax.$

385. $\displaystyle\int x^2\cos^2 ax\,dx = \frac{x^3}{6} + \left(\frac{x^2}{4a} - \frac{1}{8a^3}\right)\sin 2ax + \frac{x}{4a^2}\cos 2ax.$

386. $\displaystyle\int x\cos^3 ax\,dx = \frac{x}{12a}\sin 3ax + \frac{1}{36a^2}\cos 3ax + \frac{3x}{4a}\sin ax + \frac{3}{4a^2}\cos ax.$

387. $\displaystyle\int \frac{\sin ax}{x^m}\,dx = \frac{\sin ax}{(1-m)x^{m-1}} + \frac{a}{m-1}\int \frac{\cos ax}{x^{m-1}}\,dx.$

388. $\displaystyle\int \frac{\cos ax}{x^m}\,dx = \frac{\cos ax}{(1-m)x^{m-1}} + \frac{a}{1-m}\int \frac{\sin ax}{x^{m-1}}\,dx.$

389. $\displaystyle\int \frac{x}{1\pm\sin ax}\,dx = \mp\frac{x\cos ax}{a(1\pm\sin ax)} + \frac{1}{a^2}\log(1\pm\sin ax).$

390. $\displaystyle\int \frac{x}{1+\cos ax}\,dx = \frac{x}{a}\tan\frac{ax}{2} + \frac{2}{a^2}\log\cos\frac{ax}{2}.$

391. $\displaystyle\int \frac{x}{1-\cos ax}\,dx = -\frac{x}{a}\cot\frac{ax}{2} + \frac{2}{a^2}\log\sin\frac{ax}{2}.$

392. $\displaystyle\int \frac{x+\sin x}{1+\cos x}\,dx = x\tan\frac{x}{2}.$

393. $\displaystyle\int \frac{x-\sin x}{1-\cos x}\,dx = -x\cot\frac{x}{2}.$

394. $\displaystyle\int \sqrt{1-\cos ax}\,dx = -\frac{2\sin ax}{a\sqrt{1-\cos ax}} = -\frac{2\sqrt{2}}{a}\cos\frac{ax}{2}.$

395. $\displaystyle\int \sqrt{1+\cos ax}\,dx = \frac{2\sin ax}{a\sqrt{1+\cos ax}} = \frac{2\sqrt{2}}{a}\sin\frac{ax}{2}.$

For the following six integrals, each k represents an integer.

396. $$\int \sqrt{1+\sin x}\,dx = \begin{cases} 2\left(\sin\frac{x}{2} - \cos\frac{x}{2}\right), & (8k-1)\frac{\pi}{2} < x \le (8k+3)\frac{\pi}{2}, \\ \text{or} & \\ -2\left(\sin\frac{x}{2} - \cos\frac{x}{2}\right), & (8k+3)\frac{\pi}{2} < x \le (8k+7)\frac{\pi}{2}. \end{cases}$$

397. $$\int \sqrt{1-\sin x}\,dx = \begin{cases} 2\left(\sin\frac{x}{2} + \cos\frac{x}{2}\right), & (8k-3)\frac{\pi}{2} < x \le (8k+1)\frac{\pi}{2}, \\ \text{or} & \\ -2\left(\sin\frac{x}{2} + \cos\frac{x}{2}\right), & (8k+1)\frac{\pi}{2} < x \le (8k+5)\frac{\pi}{2}. \end{cases}$$

398. $$\int \frac{dx}{\sqrt{1-\cos x}} = \begin{cases} \sqrt{2}\log\tan\frac{x}{4}, & 4k\pi < x \le (4k+2)\pi, \\ \text{or} & \\ -\sqrt{2}\log\tan\frac{x}{4} & (4k+2)\pi < x \le (4k+4)\pi. \end{cases}$$

399. $$\int \frac{dx}{\sqrt{1+\cos x}} = \begin{cases} \sqrt{2}\log\tan\left(\frac{x+\pi}{4}\right), & (4k-1)\pi < x \le (4k+1)\pi, \\ \text{or} & \\ -\sqrt{2}\log\tan\left(\frac{x+\pi}{4}\right), & (4k+1)\pi < x \le (4k+3)\pi. \end{cases}$$

400. $$\int \frac{dx}{\sqrt{1-\sin x}} = \begin{cases} \sqrt{2}\log\tan\left(\frac{x}{4} - \frac{\pi}{8}\right), & (8k+1)\frac{\pi}{2} < x \le (8k+5)\frac{\pi}{2}, \\ \text{or} & \\ -\sqrt{2}\log\tan\left(\frac{x}{4} - \frac{\pi}{8}\right), & (8k+5)\frac{\pi}{2} < x \le (8k+9)\frac{\pi}{2}. \end{cases}$$

401. $$\int \frac{dx}{\sqrt{1+\sin x}} = \begin{cases} \sqrt{2}\log\tan\left(\frac{x}{4} + \frac{\pi}{8}\right), & (8k-1)\frac{\pi}{2} < x \le (8k+3)\frac{\pi}{2}, \\ \text{or} & \\ -\sqrt{2}\log\tan\left(\frac{x}{4} + \frac{\pi}{8}\right), & (8k+3)\frac{\pi}{2} < x \le (8k+7)\frac{\pi}{2}. \end{cases}$$

402. $$\int \tan^2 ax\,dx = \frac{1}{a}\tan ax - x.$$

403. $$\int \tan^3 ax\,dx = \frac{1}{2a}\tan^2 ax + \frac{1}{a}\log\cos ax.$$

404. $$\int \tan^4 ax\,dx = \frac{1}{3a}\tan^3 ax - \frac{1}{a}\tan ax + x.$$

405. $$\int \tan^n ax\,dx = \frac{1}{a(n-1)}\tan^{n-1} ax - \int \tan^{n-2} ax\,dx.$$

406. $$\int \cot^2 ax\,dx = -\frac{1}{a}\cot ax - x.$$

407. $$\int \cot^3 ax\,dx = -\frac{1}{2a}\cot^2 ax - \frac{1}{a}\log\sin ax.$$

408. $$\int \cot^4 ax\,dx = -\frac{1}{3a}\cot^3 ax + \frac{1}{a}\cot ax + x.$$

409. $$\int \cot^n ax\,dx = -\frac{1}{a(n-1)}\cot^{n-1} ax - \int \cot^{n-2} ax\,dx.$$

410. $$\int \frac{x}{\sin^2 ax}\,dx = \int x\csc^2 ax\,dx = -\frac{x\cot ax}{a} + \frac{1}{a^2}\log\sin ax$$

411. $$\int \frac{x}{\sin^n ax}\,dx = \int x \csc^n ax\,dx = -\frac{x\cos ax}{a(n-1)\sin^{n-1} ax} - \frac{1}{a^2(n-1)(n-2)\sin^{n-2} ax} + \frac{n-2}{n-1}\int \frac{x}{\sin^{n-2} ax}\,dx.$$

412. $$\int \frac{x}{\cos^2 ax}\,dx = \int x\sec^2 ax\,dx = \frac{x}{a}\tan ax + \frac{1}{a^2}\log\cos ax.$$

413. $$\int \frac{x}{\cos^n ax}\,dx = \int x\sec^n ax\,dx = \frac{x\sin ax}{a(n-1)\cos^{n-1} ax} - \frac{1}{a^2(n-1)(n-2)\cos^{n-2} ax} + \frac{n-2}{n-1}\int \frac{x}{\cos^{n-2} ax}\,dx.$$

414. $$\int \frac{\sin ax}{\sqrt{1+b^2\sin^2 ax}}\,dx = -\frac{1}{ab}\sin^{-1}\frac{b\cos ax}{\sqrt{1+b^2}}.$$

415. $$\int \frac{\sin ax}{\sqrt{1-b^2\sin^2 ax}}\,dx = -\frac{1}{ab}\log\left(b\cos ax + \sqrt{1-b^2\sin^2 ax}\right).$$

416. $$\int (\sin ax)\sqrt{1+b^2\sin^2 ax}\,dx = -\frac{\cos ax}{2a}\sqrt{1+b^2\sin^2 ax} - \frac{1+b^2}{2ab}\sin^{-1}\frac{b\cos ax}{\sqrt{1+b^2}}.$$

417. $$\int (\sin ax)\sqrt{1-b^2\sin^2 ax}\,dx = -\frac{\cos ax}{2a}\sqrt{1-b^2\sin^2 ax} - \frac{1-b^2}{2ab}\log\left(b\cos ax + \sqrt{1-b^2\sin^2 ax}\right).$$

418. $$\int \frac{\cos ax}{\sqrt{1+b^2\sin^2 ax}}\,dx = \frac{1}{ab}\log\left(b\sin ax + \sqrt{1+b^2\sin^2 ax}\right).$$

419. $$\int \frac{\cos ax}{\sqrt{1-b^2\sin^2 ax}}\,dx = \frac{1}{ab}\sin^{-1}(b\sin ax).$$

420. $$\int (\cos ax)\sqrt{1+b^2\sin^2 ax}\,dx = \frac{\sin ax}{2a}\sqrt{1+b^2\sin^2 ax} + \frac{1}{2ab}\log\left(b\sin ax + \sqrt{1+b^2\sin^2 ax}\right).$$

421. $$\int (\cos ax)\sqrt{1-b^2\sin^2 ax}\,dx = \frac{\sin ax}{2a}\sqrt{1-b^2\sin^2 ax} + \frac{1}{2ab}\sin^{-1}(b\sin ax).$$

For the following integral, k represents an integer and $a > |b|$

422. $$\int \frac{dx}{\sqrt{a+b\tan^2 cx}} = \begin{cases} \dfrac{1}{c\sqrt{a-b}}\sin^{-1}\left(\sqrt{\dfrac{a-b}{a}}\sin cx\right), & (4k-1)\frac{\pi}{2} < x \le (4k+1)\frac{\pi}{2}, \\ \text{or} & \\ \dfrac{-1}{c\sqrt{a-b}}\sin^{-1}\left(\sqrt{\dfrac{a-b}{a}}\sin cx\right), & (4k+1)\frac{\pi}{2} < x \le (4k+3)\frac{\pi}{2}. \end{cases}$$

423. $$\int \cos^n x\,dx = \frac{1}{2^{n-1}}\sum_{k=0}^{\frac{n}{2}-1}\binom{n}{k}\frac{\sin[(n-2k)x]}{(n-2k)} + \frac{1}{2^n}\binom{n}{\frac{n}{2}}x,\quad n \text{ is an even integer.}$$

424. $$\int \cos^n x\,dx = \frac{1}{2^{n-1}}\sum_{k=0}^{\frac{n-1}{2}}\binom{n}{k}\frac{\sin[(n-2k)x]}{(n-2k)},\quad n \text{ is an odd integer.}$$

425. $\int \sin^n x\,dx = \frac{1}{2^{n-1}} \sum_{k=0}^{\frac{n}{2}-1} \binom{n}{k} \frac{\sin\left(\left[(n-2k)(\frac{\pi}{2}-x)\right]\right)}{(2k-n)} + \frac{1}{2^n}\binom{n}{\frac{n}{2}} x,$
n is an even integer.

426. $\int \sin^n x\,dx = \frac{1}{2^{n-1}} \sum_{k=0}^{\frac{n-1}{2}} \binom{n}{k} \frac{\sin\left(\left[(n-2k)(\frac{\pi}{2}-x)\right]\right)}{(2k-n)},$ n is an odd integer.

5.4.17 FORMS INVOLVING INVERSE TRIGONOMETRIC FUNCTIONS

427. $\int \sin^{-1} ax\,dx = x\sin^{-1} ax + \frac{\sqrt{1-a^2x^2}}{a}.$

428. $\int \cos^{-1} ax\,dx = x\cos^{-1} ax - \frac{\sqrt{1-a^2x^2}}{a}.$

429. $\int \tan^{-1} ax\,dx = x\tan^{-1} ax - \frac{1}{2a}\log(1+a^2x^2).$

430. $\int \cot^{-1} ax\,dx = x\cot^{-1} ax + \frac{1}{2a}\log(1+a^2x^2).$

431. $\int \sec^{-1} ax\,dx = x\sec^{-1} ax - \frac{1}{a}\log\left(ax+\sqrt{a^2x^2-1}\right).$

432. $\int \csc^{-1} ax\,dx = x\csc^{-1} ax + \frac{1}{a}\log\left(ax+\sqrt{a^2x^2-1}\right).$

433. $\int \left(\sin^{-1}\frac{x}{a}\right)dx = x\sin^{-1}\frac{x}{a} + \sqrt{a^2-x^2}, \quad a>0.$

434. $\int \left(\cos^{-1}\frac{x}{a}\right)dx = x\cos^{-1}\frac{x}{a} - \sqrt{a^2-x^2}, \quad a>0.$

435. $\int \left(\tan^{-1}\frac{x}{a}\right)dx = x\tan^{-1}\frac{x}{a} - \frac{a}{2}\log(a^2+x^2).$

436. $\int \left(\cot^{-1}\frac{x}{a}\right)dx = x\cot^{-1}\frac{x}{a} + \frac{a}{2}\log(a^2+x^2).$

437. $\int x\sin^{-1}(ax)\,dx = \frac{1}{4a^2}\left((2a^2x^2-1)\sin^{-1}(ax) + ax\sqrt{1-a^2x^2}\right).$

438. $\int x\cos^{-1}(ax)\,dx = \frac{1}{4a^2}\left((2a^2x^2-1)\cos^{-1}(ax) - ax\sqrt{1-a^2x^2}\right).$

439. $\int x^n\sin^{-1}(ax)\,dx = \frac{x^{n+1}}{n+1}\sin^{-1}(ax) - \frac{a}{n+1}\int \frac{x^{n+1}}{\sqrt{1-a^2x^2}}\,dx, \quad n\neq -1.$

440. $\int x^n\cos^{-1} ax\,dx = \frac{x^{n+1}}{n+1}\cos^{-1}(ax) + \frac{a}{n+1}\int \frac{x^{n+1}}{\sqrt{1-a^2x^2}}\,dx, \quad n\neq -1.$

441. $\int x\tan^{-1}(ax)\,dx = \frac{1+a^2x^2}{2a^2}\tan^{-1}(ax) - \frac{x}{2a}.$

442. $\int x^n\tan^{-1}(ax)\,dx = \frac{x^{n+1}}{n+1}\tan^{-1}(ax) - \frac{a}{n+1}\int \frac{x^{n+1}}{1+a^2x^2}\,dx.$

443. $\int x\cot^{-1}(ax)\,dx = \frac{1+a^2x^2}{2a^2}\cot^{-1}(ax) + \frac{x}{2a}.$

444. $\displaystyle\int x^n \cot^{-1}(ax)\,dx = \frac{x^{n+1}}{n+1}\cot^{-1}(ax) + \frac{a}{n+1}\int \frac{x^{n+1}}{1+a^2x^2}\,dx.$

445. $\displaystyle\int \frac{\sin^{-1}(ax)}{x^2}\,dx = a\log\left(\frac{1-\sqrt{1-a^2x^2}}{x}\right) - \frac{\sin^{-1}(ax)}{x}.$

446. $\displaystyle\int \frac{\cos^{-1}(ax)}{x^2}\,dx = -\frac{1}{x}\cos^{-1}(ax) + a\log\left(\frac{1+\sqrt{1-a^2x^2}}{x}\right).$

447. $\displaystyle\int \frac{\tan^{-1}(ax)}{x^2}\,dx = -\frac{1}{x}\tan^{-1}(ax) - \frac{a}{2}\log\left(\frac{1+a^2x^2}{x^2}\right).$

448. $\displaystyle\int \frac{\cot^{-1}(ax)}{x^2}\,dx = -\frac{1}{x}\cot^{-1}(ax) - \frac{a}{2}\log\left(\frac{x^2}{1+a^2x^2}\right).$

449. $\displaystyle\int (\sin^{-1}(ax))^2\,dx = x(\sin^{-1}(ax))^2 - 2x + \frac{2\sqrt{1-a^2x^2}}{a}\sin^{-1}(ax).$

450. $\displaystyle\int (\cos^{-1}(ax))^2\,dx = x(\cos^{-1}(ax))^2 - 2x - \frac{2\sqrt{1-a^2x^2}}{a}\cos^{-1}(ax).$

451. $\displaystyle\int (\sin^{-1}(ax))^n\,dx =$

$$\begin{cases} x(\sin^{-1}(ax))^n + \dfrac{n\sqrt{1-a^2x^2}}{a}(\sin^{-1}(ax))^{n-1} - n(n-1)\displaystyle\int (\sin^{-1}(ax))^{n-2}\,dx, \\ \text{or} \\ \displaystyle\sum_{r=0}^{\lfloor \frac{n}{2}\rfloor} \frac{(-1)^r n!}{(n-2r)!}x(\sin^{-1}ax)^{n-2r} + \sum_{r=0}^{\lfloor \frac{n-1}{2}\rfloor}(-1)^r\frac{n!\sqrt{1-a^2x^2}}{(n-2r-1)!a}(\sin^{-1}ax)^{n-2r-1}. \end{cases}$$

452. $\displaystyle\int (\cos^{-1}(ax))^n\,dx =$

$$\begin{cases} x(\cos^{-1}(ax))^n - \dfrac{n\sqrt{1-a^2x^2}}{a}(\cos^{-1}(ax))^{n-1} - n(n-1)\displaystyle\int (\cos^{-1}(ax))^{n-2}\,dx, \\ \text{or} \\ \displaystyle\sum_{r=0}^{\lfloor \frac{n}{2}\rfloor} \frac{(-1)^r n!}{(n-2r)!}x(\cos^{-1}ax)^{n-2r} - \sum_{r=0}^{\lfloor \frac{n-1}{2}\rfloor}(-1)^r\frac{n!\sqrt{1-a^2x^2}}{(n-2r-1)!a}(\cos^{-1}ax)^{n-2r-1}. \end{cases}$$

453. $\displaystyle\int \frac{\sin^{-1}ax}{\sqrt{1-a^2x^2}}\,dx = \frac{1}{2a}\left(\sin^{-1}ax\right)^2.$

454. $\displaystyle\int \frac{x^n\sin^{-1}ax}{\sqrt{1-a^2x^2}}\,dx = -\frac{x^{n-1}}{na^2}\sqrt{1-a^2x^2}\sin^{-1}ax + \frac{x^n}{n^2a} + \frac{n-1}{na^2}\int \frac{x^{n-2}\sin^{-1}ax}{\sqrt{1-a^2x^2}}\,dx.$

455. $\displaystyle\int \frac{\cos^{-1}ax}{\sqrt{1-a^2x^2}}\,dx = -\frac{1}{2a}\left(\cos^{-1}ax\right)^2.$

456. $\displaystyle\int \frac{x^n\cos^{-1}ax}{\sqrt{1-a^2x^2}}\,dx = -\frac{x^{n-1}}{na^2}\sqrt{1-a^2x^2}\cos^{-1}ax - \frac{x^n}{n^2a} + \frac{n-1}{na^2}\int \frac{x^{n-2}\cos^{-1}ax}{\sqrt{1-a^2x^2}}\,dx.$

457. $\displaystyle\int \frac{\tan^{-1}ax}{1+a^2x^2}\,dx = \frac{1}{2a}\left(\tan^{-1}ax\right)^2.$

458. $\displaystyle\int \frac{\cot^{-1} ax}{1+a^2x^2}\,dx = -\frac{1}{2a}\left(\cot^{-1} ax\right)^2.$

459. $\displaystyle\int x\sec^{-1} ax\,dx = \frac{x^2}{2}\sec^{-1} ax - \frac{1}{2a^2}\sqrt{a^2x^2-1}.$

460. $\displaystyle\int x^n\sec^{-1} ax\,dx = \frac{x^{n+1}}{n+1}\sec^{-1} ax - \frac{1}{n+1}\int \frac{x^n}{\sqrt{a^2x^2-1}}\,dx.$

461. $\displaystyle\int \frac{\sec^{-1} ax}{x^2}\,dx = -\frac{\sec^{-1} ax}{x} + \frac{\sqrt{a^2x^2-1}}{x}.$

462. $\displaystyle\int x\csc^{-1} ax\,dx = \frac{x^2}{2}\csc^{-1} ax + \frac{1}{2a^2}\sqrt{a^2x^2-1}.$

463. $\displaystyle\int x^n\csc^{-1} ax\,dx = \frac{x^{n+1}}{n+1}\csc^{-1} ax + \frac{1}{n+1}\int \frac{x^n}{\sqrt{a^2x^2-1}}\,dx.$

464. $\displaystyle\int \frac{\csc^{-1} ax}{x^2}\,dx = -\frac{\csc^{-1} ax}{x} - \frac{\sqrt{a^2x^2-1}}{x}.$

5.4.18 LOGARITHMIC FORMS

465. $\displaystyle\int \log x\,dx = x\log x - x.$

466. $\displaystyle\int x\log x\,dx = \frac{x^2}{2}\log x - \frac{x^2}{4}.$

467. $\displaystyle\int x^2\log x\,dx = \frac{x^3}{3}\log x - \frac{x^3}{9}.$

468. $\displaystyle\int x^n\log x\,dx = \frac{x^{n+1}}{n+1}\log x - \frac{x^{n+1}}{(n+1)^2}.$

469. $\displaystyle\int (\log x)^2\,dx = x(\log x)^2 - 2x\log x + 2x.$

470. $$\int (\log x)^n\,dx = \begin{cases} x(\log x)^n - n\displaystyle\int (\log x)^{n-1}\,dx, & n \neq -1, \\ \text{or} & \\ (-1)^n n!x\displaystyle\sum_{r=0}^{n}\frac{(-\log x)^r}{r!}, & n \neq -1. \end{cases}$$

471. $\displaystyle\int \frac{(\log x)^n}{x}\,dx = \frac{1}{n+1}(\log x)^{n+1}, \quad n \neq -1.$

472. $\displaystyle\int \frac{dx}{\log x} = \log(\log x) + \log x + \frac{(\log x)^2}{2\cdot 2!} + \frac{(\log x)^3}{3\cdot 3!} + \cdots.$

473. $\displaystyle\int \frac{dx}{x\log x} = \log(\log x).$

474. $\displaystyle\int \frac{dx}{x(\log x)^n} = \frac{1}{(1-n)(\log x)^{n-1}}, \quad n \neq 1.$

475. $\displaystyle\int \frac{x^m\,dx}{(\log x)^n} = \frac{x^{m+1}}{(1-n)(\log x)^{n-1}} + \frac{m+1}{n-1}\int \frac{x^m\,dx}{(\log x)^{n-1}}, \quad n \neq 1.$

476. $$\int x^m (\log x)^n \, dx = \begin{cases} \dfrac{x^{m+1}(\log x)^n}{m+1} - \dfrac{n}{m+1}\displaystyle\int x^m (\log x)^{n-1}\, dx, \\ \text{or} \\ (-1)^n \dfrac{n!}{m+1} x^{m+1} \displaystyle\sum_{r=0}^{n} \dfrac{(-\log x)^r}{r!(m+1)^{n-r}}. \end{cases}$$

477. $$\int x^p \cos(b\log x)\, dx = \frac{x^{p+1}}{(p+1)^2+b^2}\left[b\sin(b\log x) + (p+1)\cos(b\log x)\right].$$

478. $$\int x^p \sin(b\log x)\, dx = \frac{x^{p+1}}{(p+1)^2+b^2}\left[(p+1)\sin(b\log x) - b\cos(b\log x)\right].$$

479. $$\int \log(ax+b)\, dx = \frac{ax+b}{a}\log(ax+b) - x.$$

480. $$\int \frac{\log(ax+b)}{x^2}\, dx = \frac{a}{b}\log x - \frac{ax+b}{bx}\log(ax+b).$$

481. $$\int x^m \log(ax+b)\, dx = \frac{1}{m+1}\left[x^{m+1} - \left(-\frac{b}{a}\right)^{m+1}\right]\log(ax+b) - \frac{1}{m+1}\left(-\frac{b}{a}\right)^{m+1}\sum_{r=1}^{m+1}\frac{1}{r}\left(-\frac{ax}{b}\right)^r.$$

482. $$\int \frac{\log(ax+b)}{x^m}\, dx = -\frac{1}{m-1}\frac{\log(ax+b)}{x^{m-1}} + \frac{1}{m-1}\left(-\frac{a}{b}\right)^{m-1}\log\frac{ax+b}{x} + \frac{1}{m-1}\left(-\frac{a}{b}\right)^{m-1}\sum_{r=1}^{m-2}\frac{1}{r}\left(-\frac{b}{ax}\right)^r, \quad m > 2.$$

483. $$\int \log\frac{x+a}{x-a}\, dx = (x+a)\log(x+a) - (x-a)\log(x-a).$$

484. $$\int x^m \log\frac{x+a}{x-a}\, dx = \frac{x^{m+1} - (-a)^{m+1}}{m+1}\log(x+a) - \frac{x^{m+1} - a^{m+1}}{m+1}\log(x-a) + \frac{2a^{m+1}}{m+1}\sum_{r=1}^{\lfloor\frac{m+1}{2}\rfloor}\frac{1}{m-2r+2}\left(\frac{x}{a}\right)^{m-2r+2}.$$

485. $$\int \frac{1}{x^2}\log\frac{x+a}{x-a}\, dx = \frac{1}{x}\log\frac{x-a}{x+a} - \frac{1}{a}\log\frac{x^2-a^2}{x^2}.$$

For the following two integrals, $X = a + bx + cx^2$.

486. $$\int \log X\, dx =$$
$$\begin{cases} \left(x + \dfrac{b}{2c}\right)\log X - 2x + \dfrac{\sqrt{4ac-b^2}}{c}\tan^{-1}\dfrac{2cx+b}{\sqrt{4ac-b^2}}, & b^2 - 4ac < 0, \\ \text{or} \\ \left(x + \dfrac{b}{2c}\right)\log X - 2x + \dfrac{\sqrt{b^2-4ac}}{c}\tanh^{-1}\dfrac{2cx+b}{\sqrt{b^2-4ac}}, & b^2 - 4ac > 0. \end{cases}$$

487. $$\int x^n \log X\, dx = \frac{x^{n+1}}{n+1}\log X - \frac{2c}{n+1}\int \frac{x^{n+2}}{X}\, dx - \frac{b}{n+1}\int \frac{x^{n+1}}{X}\, dx, \quad n \neq -1.$$

488. $$\int \log(x^2+a^2)\, dx = x\log(x^2+a^2) - 2x + 2a\tan^{-1}\frac{x}{a}.$$

489. $$\int \log(x^2-a^2)\, dx = x\log(x^2-a^2) - 2x + a\log\frac{x+a}{x-a}.$$

490. $\int x \log\left(x^2+a^2\right)dx = \frac{1}{2}\left(x^2+a^2\right)\log\left(x^2+a^2\right) - \frac{1}{2}x^2.$

491. $\int \log\left(x+\sqrt{x^2\pm a^2}\right)dx = x\log\left(x+\sqrt{x^2\pm a^2}\right) - \sqrt{x^2\pm a^2}.$

492. $\int x\log\left(x+\sqrt{x^2\pm a^2}\right)dx = \left(\frac{x^2}{2}\pm\frac{a^2}{4}\right)\log\left(x+\sqrt{x^2\pm a^2}\right) - \frac{x\sqrt{x^2\pm a^2}}{4}.$

493. $$\int x^m\log\left(x+\sqrt{x^2\pm a^2}\right)dx = \frac{x^{m+1}}{m+1}\log\left(x+\sqrt{x^2\pm a^2}\right) - \frac{1}{m+1}\int\frac{x^{m+1}}{\sqrt{x^2\pm a^2}}\,dx.$$

494. $\int \frac{\log\left(x+\sqrt{x^2+a^2}\right)}{x^2}dx = -\frac{\log\left(x+\sqrt{x^2+a^2}\right)}{x} - \frac{1}{a}\log\frac{a+\sqrt{x^2+a^2}}{x}.$

495. $\int \frac{\log\left(x+\sqrt{x^2-a^2}\right)}{x^2}dx = -\frac{\log\left(x+\sqrt{x^2-a^2}\right)}{x} + \frac{1}{|a|}\sec^{-1}\frac{x}{a}.$

496. $$\int x^n\log\left(x^2-a^2\right)dx = \frac{1}{n+1}\left[x^{n+1}\log\left(x^2-a^2\right) - a^{n+1}\log(x-a) - (-a)^{n+1}\log(x+a) - 2\sum_{r=0}^{\lfloor\frac{n}{2}\rfloor}\frac{a^{2r}x^{n-2r+1}}{n-2r+1}\right].$$

5.4.19 EXPONENTIAL FORMS

497. $\int e^x\,dx = e^x.$

498. $\int e^{-x}\,dx = -e^{-x}.$

499. $\int e^{ax}\,dx = \frac{e^{ax}}{a}.$

500. $\int xe^{ax}\,dx = \frac{e^{ax}}{a^2}(ax-1).$

501. $$\int x^m e^{ax}\,dx = \begin{cases} \dfrac{x^m e^{ax}}{a} - \dfrac{m}{a}\displaystyle\int x^{m-1}e^{ax}\,dx, \\ \text{or} \\ e^{ax}\displaystyle\sum_{r=0}^{m}(-1)^r\frac{m!x^{m-r}}{(m-r)!a^{r+1}}. \end{cases}$$

502. $\int \frac{e^{ax}}{x}dx = \log x + \frac{ax}{1!} + \frac{a^2x^2}{2\cdot 2!} + \frac{a^3x^3}{3\cdot 3!} + \dots.$

503. $\int \frac{e^{ax}}{x^m}dx = \frac{1}{1-m}\frac{e^{ax}}{x^{m-1}} + \frac{a}{m-1}\int\frac{e^{ax}}{x^{m-1}}\,dx, \quad m\neq 1.$

504. $\int e^{ax}\log x\,dx = \frac{e^{ax}\log x}{a} - \frac{1}{a}\int\frac{e^{ax}}{x}\,dx.$

505. $\int \frac{dx}{1+e^x} = x - \log\left(1+e^x\right) = \log\frac{e^x}{1+e^x}$

506. $\displaystyle\int \frac{dx}{a+be^{px}} = \frac{x}{a} - \frac{1}{ap}\log\left(a+be^{px}\right).$

507. $\displaystyle\int \frac{dx}{ae^{mx}+be^{-mx}} = \frac{1}{m\sqrt{ab}}\tan^{-1}\left(e^{mx}\sqrt{\frac{a}{b}}\right), \quad a>0,\ b>0.$

508. $$\int \frac{dx}{ae^{mx}-be^{-mx}} = \begin{cases} \dfrac{1}{2m\sqrt{ab}}\log\left(\dfrac{\sqrt{a}e^{mx}-\sqrt{b}}{\sqrt{a}e^{mx}+\sqrt{b}}\right), & a>0,\ b>0, \\ \text{or} & \\ \dfrac{-1}{m\sqrt{ab}}\tanh^{-1}\left(\sqrt{\dfrac{a}{b}}e^{mx}\right), & a>0,\ b>0. \end{cases}$$

509. $\displaystyle\int \left(a^{x}-a^{-x}\right)\,dx = \frac{a^{x}+a^{-x}}{\log a}.$

510. $\displaystyle\int \frac{e^{ax}}{b+ce^{ax}}\,dx = \frac{1}{ac}\log\left(b+ce^{ax}\right).$

511. $\displaystyle\int \frac{xe^{ax}}{(1+ax)^2}\,dx = \frac{e^{ax}}{a^2(1+ax)}.$

512. $\displaystyle\int xe^{-x^2}\,dx = -\frac{1}{2}e^{-x^2}.$

513. $\displaystyle\int e^{ax}\sin(bx)\,dx = \frac{e^{ax}\left[a\sin(bx)-b\cos(bx)\right]}{a^2+b^2}.$

514. $$\int e^{ax}\sin(bx)\sin(cx)\,dx = \frac{e^{ax}\left[(b-c)\sin(b-c)x+a\cos(b-c)x\right]}{2\left[a^2+(b-c)^2\right]} - \frac{e^{ax}\left[(b+c)\sin(b+c)x+a\cos(b+c)x\right]}{2\left[a^2+(b+c)^2\right]}.$$

515. $$\int e^{ax}\sin(bx)\cos(cx)\,dx = \frac{e^{ax}\left[a\sin(b-c)x-(b-c)\cos(b-c)x\right]}{2\left[a^2+(b-c)^2\right]} + \frac{e^{ax}\left[a\sin(b+c)x-(b+c)\cos(b+c)x\right]}{2\left[a^2+(b+c)^2\right]}.$$

516. $\displaystyle\int e^{ax}\sin(bx)\sin(bx+c)\,dx = \frac{e^{ax}\cos c}{2a} - \frac{e^{ax}\left[a\cos 2bx+c+2b\sin 2bx+c\right]}{2\left[a^2+4b^2\right]}.$

517. $\displaystyle\int e^{ax}\sin(bx)\cos(bx+c)\,dx = -\frac{e^{ax}\sin c}{2a} + \frac{e^{ax}\left[a\sin 2bx+c-2b\cos 2bx+c\right]}{2\left[a^2+4b^2\right]}.$

518. $\displaystyle\int e^{ax}\cos(bx)\,dx = \frac{e^{ax}}{a^2+b^2}\left[a\cos(bx)+b\sin(bx)\right].$

519. $$\int e^{ax}\cos(bx)\cos(cx)\,dx = \frac{e^{ax}\left[(b-c)\sin(b-c)x+a\cos(b-c)x\right]}{2\left[a^2+(b-c)^2\right]} + \frac{e^{ax}\left[(b+c)\sin(b+c)x+a\cos(b+c)x\right]}{2\left[a^2+(b+c)^2\right]}.$$

520. $\displaystyle\int e^{ax}\cos(bx)\cos(bx+c)\,dx = \frac{e^{ax}\cos c}{2a} + \frac{e^{ax}\left[a\cos 2bx+c+2b\sin 2bx+c\right]}{2\left[a^2+4b^2\right]}.$

521. $\displaystyle\int e^{ax}\cos(bx)\sin(bx+c)\,dx = \frac{e^{ax}\sin c}{2a} + \frac{e^{ax}\left[a\sin 2bx+c-2b\cos 2bx+c\right]}{2\left[a^2+4b^2\right]}.$

522. $$\int e^{ax}\sin^{n}(bx)\,dx = \frac{1}{a^2+n^2b^2}\left[(a\sin(bx)-nb\cos(bx))e^{ax}\sin^{n-1}(bx) + n(n-1)b^2\int e^{ax}\sin^{n-2}(bx)\,dx\right].$$

523. $$\int e^{ax}\cos^n(bx)\,dx = \frac{1}{a^2+n^2b^2}\Big[(a\cos(bx)+nb\sin(bx))e^{ax}\cos^{n-1}(bx) + n(n-1)b^2\int e^{ax}\cos^{n-2}(bx)\,dx\Big].$$

524. $$\int x^m e^x \sin x\,dx = \frac{1}{2}x^m e^x(\sin x - \cos x) - \frac{m}{2}\int x^{m-1}e^x\sin x\,dx + \frac{m}{2}\int x^{m-1}e^x\cos x\,dx.$$

525. $$\int x^m e^{ax}\sin bx\,dx = x^m e^{ax}\frac{a\sin(bx)-b\cos(bx)}{a^2+b^2} - \frac{m}{a^2+b^2}\int x^{m-1}e^{ax}(a\sin(bx)-b\cos(bx))\,dx.$$

526. $$\int x^m e^x \cos x\,dx = \frac{1}{2}x^m e^x(\sin x + \cos x) - \frac{m}{2}\int x^{m-1}e^x\sin x\,dx - \frac{m}{2}\int x^{m-1}e^x\cos x\,dx.$$

527. $$\int x^m e^{ax}\cos bx\,dx = x^m e^{ax}\frac{a\cos(bx)+b\sin(bx)}{a^2+b^2} - \frac{m}{a^2+b^2}\int x^{m-1}e^{ax}(a\cos(bx)+b\sin(bx))\,dx.$$

528. $$\int e^{ax}\cos^m x\sin^n x\,dx =$$

$$\begin{cases} \dfrac{e^{ax}(\cos^{m-1}x)(\sin^n x)\,[a\cos x+(m+n)\sin x]}{(m+n)^2+a^2} - \dfrac{na}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^{m-1}x)(\sin^{n-1}x)\,dx + \dfrac{(m-1)(m+n)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^{m-2}x)(\sin^n x)\,dx, \\ \text{or} \\ \dfrac{e^{ax}(\cos^m x)(\sin^{n-1}x)\,[a\sin x-(m+n)\cos x]}{(m+n)^2+a^2} + \dfrac{ma}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^{m-1}x)(\sin^{n-1}x)\,dx + \dfrac{(n-1)(m+n)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^m x)(\sin^{n-2}x)\,dx, \\ \text{or} \\ \dfrac{e^{ax}(\cos^{m-1}x)(\sin^{n-1}x)\,\left[a\sin x\cos x+m\sin^2 x-n\cos^2 x\right]}{(m+n)^2+a^2} + \dfrac{m(m-1)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^{m-2}x)(\sin^n x)\,dx + \dfrac{n(n-1)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^m x)(\sin^{n-2}x)\,dx, \\ \text{or} \\ \dfrac{e^{ax}(\cos^{m-1}x)(\sin^{n-1}x)\,\left[a\sin x\cos x+m\sin^2 x-n\cos^2 x\right]}{(m+n)^2+a^2} + \dfrac{m(m-1)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^{m-2}x)(\sin^{n-2}x)\,dx + \dfrac{(n-m)(n+m-1)}{(m+n)^2+a^2}\displaystyle\int e^{ax}(\cos^m x)(\sin^{n-2}x)\,dx. \end{cases}$$

529. $$\int xe^{ax}\sin(bx)\,dx = \frac{xe^{ax}}{a^2+b^2}\left[a\sin(bx) - b\cos(bx)\right] - \frac{e^{ax}}{(a^2+b^2)^2}\left[\left(a^2-b^2\right)\sin bx - 2ab\cos(bx)\right].$$

530. $$\int xe^{ax}\cos(bx)\,dx = \frac{xe^{ax}}{a^2+b^2}\left[a\cos(bx) + b\sin(bx)\right] - \frac{e^{ax}}{(a^2+b^2)^2}\left[\left(a^2-b^2\right)\cos bx + 2ab\sin(bx)\right].$$

531. $$\int \frac{e^{ax}}{\sin^n x}\,dx = -\frac{e^{ax}\left[a\sin x + (n-2)\cos x\right]}{(n-1)(n-2)\sin^{n-1}x} + \frac{a^2+(n-2)^2}{(n-1)(n-2)}\int \frac{e^{ax}}{\sin^{n-2}x}\,dx.$$

532. $$\int \frac{e^{ax}}{\cos^n x}\,dx = -\frac{e^{ax}\left[a\cos x - (n-2)\sin x\right]}{(n-1)(n-2)\cos^{n-1}x} + \frac{a^2+(n-2)^2}{(n-1)(n-2)}\int \frac{e^{ax}}{\cos^{n-2}x}\,dx.$$

533. $$\int e^{ax}\tan^n x\,dx = e^{ax}\frac{\tan^{n-1}x}{n-1} - \frac{a}{n-1}\int e^{ax}\tan^{n-1}x\,dx - \int e^{ax}\tan^{n-2}x\,dx.$$

5.4.20 HYPERBOLIC FORMS

534. $$\int \sinh x\,dx = \cosh x.$$

535. $$\int \cosh x\,dx = \sinh x.$$

536. $$\int \tanh x\,dx = \log\cosh x.$$

537. $$\int \coth x\,dx = \log\sinh x.$$

538. $$\int \operatorname{sech} x\,dx = \tan^{-1}(\sinh x).$$

539. $$\int \operatorname{csch} x\,dx = \log\tanh\left(\frac{x}{2}\right).$$

540. $$\int x\sinh x\,dx = x\cosh x - \sinh x.$$

541. $$\int x^n\sinh x\,dx = x^n\cosh x - n\int x^{n-1}(\cosh x)\,dx.$$

542. $$\int x\cosh x\,dx = x\sinh x - \cosh x.$$

543. $$\int x^n\cosh x\,dx = x^n\sinh x - n\int x^{n-1}(\sinh x)\,dx.$$

544. $$\int \operatorname{sech} x\tanh x\,dx = -\operatorname{sech} x.$$

545. $$\int \operatorname{csch} x\coth x\,dx = -\operatorname{csch} x.$$

546. $$\int \sinh^2 x\,dx = \frac{\sinh 2x}{4} - \frac{x}{2}.$$

547. $\int \sinh^m x \cosh^n x\,dx =$

$$\begin{cases} \dfrac{1}{m+n}\sinh^{m+1} x\cosh^{n-1} x + \dfrac{n-1}{m+n}\displaystyle\int \sinh^m x\cosh^{n-2} x\,dx, & m+n\neq 0, \\ \text{or} & \\ \dfrac{1}{m+n}\sinh^{m-1} x\cosh^{n+1} x - \dfrac{m-1}{m+n}\displaystyle\int \sinh^{m-2} x\cosh^{n} x\,dx, & m+n\neq 0. \end{cases}$$

548. $\displaystyle\int \frac{dx}{(\sinh^m x)(\cosh^n x)} =$

$$\begin{cases} -\dfrac{1}{(m-1)(\sinh^{m-1} x)(\cosh^{n-1} x)} - \dfrac{m+n-2}{m-1}\displaystyle\int \frac{dx}{(\sinh^{m-2} x)(\cosh^n x)}\,dx, & m\neq 1, \\ \text{or} & \\ \dfrac{1}{(n-1)(\sinh^{m-1} x)(\cosh^{n-1} x)} + \dfrac{m+n-2}{n-1}\displaystyle\int \frac{dx}{(\sinh^{m} x)(\cosh^{n-2} x)}\,dx, & n\neq 1. \end{cases}$$

549. $\int \tanh^2 x\,dx = x - \tanh x.$

550. $\displaystyle\int \tanh^n x\,dx = -\frac{\tanh^{n-1} x}{n-1} + \int (\tanh^{n-2} x)\,dx, \quad n\neq 1.$

551. $\int \operatorname{sech}^2 x\,dx = \tanh x.$

552. $\displaystyle\int \cosh^2 x\,dx = \frac{\sinh 2x}{4} + \frac{x}{2}.$

553. $\int \coth^2 x\,dx = x - \coth x.$

554. $\displaystyle\int \coth^n x\,dx = -\frac{\coth^{n-1} x}{n-1} + \int \coth^{n-2} x\,dx, \quad n\neq 1.$

555. $\int \operatorname{csch}^2 x\,dx = -\coth x.$

556. $\displaystyle\int (\sinh mx)(\sinh nx)\,dx = \frac{\sinh(m+n)x}{2(m+n)} - \frac{\sinh(m-n)x}{2(m-n)}, \quad m^2\neq n^2.$

557. $\displaystyle\int (\cosh mx)(\cosh nx)\,dx = \frac{\sinh(m+n)x}{2(m+n)} + \frac{\sinh(m-n)x}{2(m-n)}, \quad m^2\neq n^2.$

558. $\displaystyle\int (\sinh mx)(\cosh nx)\,dx = \frac{\cosh(m+n)x}{2(m+n)} + \frac{\cosh(m-n)x}{2(m-n)}, \quad m^2\neq n^2.$

559. $\displaystyle\int \left(\sinh^{-1}\frac{x}{a}\right)dx = x\sinh^{-1}\frac{x}{a} - \sqrt{x^2+a^2}, \quad a>0.$

560. $\displaystyle\int x\left(\sinh^{-1}\frac{x}{a}\right)dx = \left(\frac{x^2}{2}+\frac{a^2}{4}\right)\sinh^{-1}\frac{x}{a} - \frac{x}{4}\sqrt{x^2+a^2}, \quad a>0.$

561. $\displaystyle\int x^n\sinh^{-1} x\,dx = \frac{x^{n+1}}{n+1}\sinh^{-1} x - \frac{1}{n+1}\int \frac{x^{n+1}}{\sqrt{1+x^2}}\,dx, \quad n\neq -1.$

562. $$\int^z \cosh^{-1}\frac{x}{a}\,dx = \begin{cases} z\cosh^{-1}\dfrac{z}{a} - \sqrt{z^2-a^2}, & \cosh^{-1}\frac{z}{a} > 0, \\ \text{or} & \\ z\cosh^{-1}\dfrac{z}{a} + \sqrt{z^2-a^2}, & \cosh^{-1}\frac{z}{a} < 0,\ a>0. \end{cases}$$

563. $\int x\left(\cosh^{-1}\frac{x}{a}\right)\,dx = \left(\frac{x^2}{2}-\frac{a^2}{4}\right)\cosh^{-1}\frac{x}{a} - \frac{x}{4}\sqrt{x^2-a^2}.$

564. $\int x^n \cosh^{-1} x\,dx = \frac{x^{n+1}}{n+1}\cosh^{-1} x - \frac{1}{n+1}\int \frac{x^{n+1}}{\sqrt{x^2-1}}\,dx, \quad n \neq -1.$

565. $\int \left(\tanh^{-1}\frac{x}{a}\right)\,dx = x\tanh^{-1}\frac{x}{a} + \frac{a}{2}\log\left(a^2-x^2\right), \quad \left|\frac{x}{a}\right| < 1.$

566. $\int \left(\coth^{-1}\frac{x}{a}\right)\,dx = x\coth^{-1}\frac{x}{a} + \frac{a}{2}\log\left(x^2-a^2\right), \quad \left|\frac{x}{a}\right| > 1.$

567. $\int x\left(\tanh^{-1}\frac{x}{a}\right)\,dx = \frac{x^2-a^2}{2}\tanh^{-1}\frac{x}{a} + \frac{ax}{2}, \quad \left|\frac{x}{a}\right| < 1.$

568. $\int x^n \tanh^{-1} x\,dx = \frac{x^{n+1}}{n+1}\tanh^{-1} x - \frac{1}{n+1}\int \frac{x^{n+1}}{1-x^2}\,dx, \quad n \neq -1.$

569. $\int x\left(\coth^{-1}\frac{x}{a}\right)\,dx = \frac{x^2-a^2}{2}\coth^{-1}\frac{x}{a} + \frac{ax}{2}, \quad \left|\frac{x}{a}\right| > 1.$

570. $\int x^n \coth^{-1} x\,dx = \frac{x^{n+1}}{n+1}\coth^{-1} x + \frac{1}{n+1}\int \frac{x^{n+1}}{x^2-1}\,dx, \quad n \neq -1.$

571. $\int \operatorname{sech}^{-1} x\,dx = x\operatorname{sech}^{-1} x + \sin^{-1} x.$

572. $\int x\operatorname{sech}^{-1} x\,dx = \frac{x^2}{2}\operatorname{sech}^{-1} x - \frac{1}{2}\sqrt{1-x^2}.$

573. $\int x^n \operatorname{sech}^{-1} x\,dx = \frac{x^{n+1}}{n+1}\operatorname{sech}^{-1} x + \frac{1}{n+1}\int \frac{x^n}{\sqrt{1-x^2}}\,dx, \quad n \neq -1.$

574. $\int \operatorname{csch}^{-1} x\,dx = x\operatorname{csch}^{-1} x + \frac{x}{|x|}\sinh^{-1} x.$

575. $\int x\operatorname{csch}^{-1} x\,dx = \frac{x^2}{2}\operatorname{csch}^{-1} x + \frac{1}{2}\frac{x}{|x|}\sqrt{1+x^2}.$

576. $\int x^n \operatorname{csch}^{-1} x\,dx = \frac{x^{n+1}}{n+1}\operatorname{csch}^{-1} x + \frac{1}{n+1}\frac{x}{|x|}\int \frac{x^n}{\sqrt{1+x^2}}\,dx, \quad n \neq -1.$

5.4.21 BESSEL FUNCTIONS

$Z_p(x)$ represents any of the Bessel functions $\{J_p(x), Y_p(x), e^{p\pi i}K_p(x), I_p(x)\}$.

577. $\int x^{p+1} Z_p(x)\,dx = x^{p+1} Z_{p+1}(x).$

578. $\int x^{-p+1} Z_p(x)\,dx = -x^{-p+1} Z_{p-1}(x).$

579. $\int x\left[Z_p(ax)\right]^2\,dx = \frac{x^2}{2}\left[\left[Z_p(ax)\right]^2 - Z_{p-1}(ax)Z_{p+1}(ax)\right].$

580. $\int Z_1(x)\,dx = -Z_0(x).$

581. $\int xZ_0(x)\,dx = xZ_1(x).$

5.5 TABLE OF DEFINITE INTEGRALS

582. $\displaystyle\int_0^\infty x^{n-1}e^{-x}\,dx = \Gamma(n), \qquad \operatorname{Re} n > 0.$

583. $\displaystyle\int_0^\infty x^n p^{-x}\,dx = \frac{n!}{(\log p)^{n+1}}, \quad p > 0,\ n$ is a non-negative integer.

584. $\displaystyle\int_0^\infty x^{n-1}e^{-(a+1)x}\,dx = \frac{\Gamma(n)}{(a+1)^n}, \quad n > 0,\ a > -1.$

585. $\displaystyle\int_0^1 x^m \left(\log\frac{1}{x}\right)^n dx = \frac{\Gamma(n+1)}{(m+1)^{n+1}}, \quad m > -1,\ n > -1.$

586. $\displaystyle\int_0^1 x^{m-1}(1-x)^{n-1}\,dx = \int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}} = \frac{\Gamma(m)\,\Gamma(n)}{\Gamma(m+n)}, \quad n > 0,\ m > 0.$

587. $\displaystyle\int_a^b (x-a)^m(b-x)^n\,dx = (b-a)^{m+n+1}\frac{\Gamma(m+1)\,\Gamma(n+1)}{\Gamma(m+n+2)},$
$m > -1,\ n > -1,\ b > a.$

588. $\displaystyle\int_1^\infty \frac{dx}{x^m} = \frac{1}{m-1}, \quad m > 1.$

589. $\displaystyle\int_0^\infty \frac{dx}{(1+x)x^p} = \pi\csc p\pi, \quad 0 < p < 1.$

590. $\displaystyle\int_0^\infty \frac{dx}{(1-x)x^p} = -\pi\cot p\pi, \quad 0 < p < 1.$

591. $\displaystyle\int_0^1 \frac{x^p}{(1-x)^p}\,dx = p\pi\csc p\pi, \quad |p| < 1.$

592. $\displaystyle\int_0^1 \frac{x^p}{(1-x)^{p+1}}\,dx = \int_0^1 \frac{(1-x)^p}{x^{p+1}}\,dx = -\pi\operatorname{cosec} p\pi, \quad -1 < p < 0.$

593. $\displaystyle\int_0^\infty \frac{x^{p-1}}{1+x}\,dx = \frac{\pi}{\sin p\pi}, \quad 0 < p < 1.$

594. $\displaystyle\int_0^\infty \frac{x^{m-1}}{1+x^n}\,dx = \frac{\pi}{n\sin\frac{m\pi}{n}}, \quad 0 < m < n.$

595. $\displaystyle\int_0^\infty \frac{x^a}{(m+x^b)^c}\,dx = \frac{m^{(a+1-bc)/b}}{b}\,\frac{\Gamma\!\left(\frac{a+1}{b}\right)\Gamma\!\left(c-\frac{a+1}{b}\right)}{\Gamma(c)},$
$a > -1,\ b > 0,\ m > 0,\ c > \frac{a+1}{b}.$

596. $\displaystyle\int_0^\infty \frac{dx}{(1+x)\sqrt{x}} = \pi.$

597. $\displaystyle\int_0^\infty \frac{a}{a^2+x^2}\,dx = \begin{cases} \frac{\pi}{2}, & a > 0, \\ \text{or} & \\ 0, & a = 0, \\ \text{or} & \\ -\frac{\pi}{2}, & a < 0. \end{cases}$

598. $\displaystyle\int_0^a \left(a^2-x^2\right)^{n/2}\,dx = \int_{-a}^a \frac{1}{2}\left(a^2-x^2\right)^{n/2}\,dx = \frac{n!!}{(n+1)!!}\,\frac{\pi}{2}a^{n+1},$
$a > 0,\ n$ is an odd integer.

599. $\displaystyle\int_0^a x^m \left(a^2 - x^2\right)^{n/2}\, dx = \frac{1}{2} a^{m+n+1} \frac{\Gamma\left(\frac{m+1}{2}\right)\Gamma\left(\frac{n+2}{2}\right)}{\Gamma\left(\frac{m+n+3}{2}\right)}, \quad a > 0,\ m > -1,\ n > -2.$

600.

$$\int_0^{\pi/2} \sin^n x\, dx = \int_0^{\pi/2} \cos^n x\, dx = \begin{cases} \dfrac{\sqrt{\pi}}{2} \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n+2}{2}\right)}, & n > -1, \\ \text{or} & \\ \dfrac{(n-1)!!}{n!!} \dfrac{\pi}{2}, & n \neq 0,\ n \text{ is an even integer}, \\ \text{or} & \\ \dfrac{(n-1)!!}{n!!}, & n \neq 1,\ n \text{ is an odd integer}. \end{cases}$$

601. $\displaystyle\int_0^\infty \frac{\sin ax}{x}\, dx = \begin{cases} \dfrac{\pi}{2}, & a > 0, \\ \text{or} & \\ 0, & a = 0, \\ \text{or} & \\ -\dfrac{\pi}{2}, & a < 0. \end{cases}$

602. $\displaystyle\int_0^\infty \frac{\cos x}{x}\, dx = \infty.$

603. $\displaystyle\int_0^\infty \frac{\tan x}{x}\, dx = \frac{\pi}{2}.$

604. $\displaystyle\int_0^\infty \frac{\tan ax}{x}\, dx = \frac{\pi}{2}, \quad a > 0.$

605. $\displaystyle\int_0^\pi \sin(nx)\sin(mx)\, dx = \int_0^\pi \cos(nx)\cos(mx)\, dx = 0,$

$n \neq m$, n is an integer, m is an integer.

606. $\displaystyle\int_0^{\pi/n} \sin(nx)\cos(nx)\, dx = \int_0^\pi \sin(nx)\cos(nx)\, dx = 0,$ n is an integer.

607. $\displaystyle\int_0^\pi \sin ax \cos bx\, dx = \begin{cases} \dfrac{2a}{a^2 - b^2}, & a - b \text{ is an odd integer}. \\ \text{or} & \\ 0, & a - b \text{ is an even integer}. \end{cases}$

608. $\displaystyle\int_0^\infty \frac{\sin x \cos ax}{x}\, dx = \begin{cases} 0, & |a| > 1, \\ \text{or} & \\ \dfrac{\pi}{4}, & |a| = 1, \\ \text{or} & \\ \dfrac{\pi}{2}, & |a| < 1. \end{cases}$

609. $\displaystyle\int_0^\infty \frac{\sin ax \sin bx}{x^2}\, dx = \begin{cases} \dfrac{\pi a}{2}, & 0 < a \leq b, \\ \text{or} & \\ \dfrac{\pi b}{2}, & 0 < b \leq a. \end{cases}$

610. $\displaystyle\int_0^\pi \sin^2 mx\, dx = \int_0^\pi \cos^2 mx\, dx = \frac{\pi}{2},$ m is an integer.

611. $\displaystyle\int_0^\infty \frac{\sin^2 px}{x^2}\, dx = \frac{\pi\,|p|}{2}.$

612. $\int_0^\infty \frac{\sin x}{x^p}\,dx = \frac{\pi}{2\Gamma(p)\sin(p\pi/2)}, \quad 0 < p < 1.$

613. $\int_0^\infty \frac{\cos x}{x^p}\,dx = \frac{\pi}{2\Gamma(p)\cos(p\pi/2)}, \quad 0 < p < 1.$

614. $\int_0^\infty \frac{1-\cos px}{x^2}\,dx = \frac{\pi\,|p|}{2}.$

615. $\int_0^\infty \frac{\sin px\cos qx}{x}\,dx = \begin{cases} 0 & q > p > 0, \\ \text{or} & \\ \frac{\pi}{2}, & p > q > 0, \\ \text{or} & \\ \frac{\pi}{4}, & p = q > 0. \end{cases}$

616. $\int_0^\infty \frac{\cos mx}{x^2+a^2}\,dx = \frac{\pi}{2\,|a|}e^{-|ma|}.$

617. $\int_0^\infty \cos x^2\,dx = \int_0^\infty \sin x^2\,dx = \frac{1}{2}\sqrt{\frac{\pi}{2}}.$

618. $\int_0^\infty \sin(ax^n)\,dx = \frac{1}{na^{1/n}}\Gamma\left(\frac{1}{n}\right)\sin\frac{\pi}{2n}, \quad n > 1.$

619. $\int_0^\infty \cos(ax^n)\,dx = \frac{1}{na^{1/n}}\Gamma\left(\frac{1}{n}\right)\cos\frac{\pi}{2n}, \quad n > 1.$

620. $\int_0^\infty \frac{\sin x}{\sqrt{x}}\,dx = \int_0^\infty \frac{\cos x}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{2}}.$

621. $\int_0^\infty \frac{\sin^3 x}{x}\,dx = \frac{\pi}{4}.$

622. $\int_0^\infty \frac{\sin^3 x}{x^2}\,dx = \frac{3}{4}\log 3.$

623. $\int_0^\infty \frac{\sin^3 x}{x^3}\,dx = \frac{3\pi}{8}.$

624. $\int_0^\infty \frac{\sin^4 x}{x^4}\,dx = \frac{\pi}{3}.$

625. $\int_0^{\pi/2} \frac{dx}{1+a\cos x}\,dx = \frac{\cos^{-1}a}{\sqrt{1-a^2}}, \quad |a| < 1.$

626. $\int_0^{\pi} \frac{dx}{a+b\cos x}\,dx = \frac{\pi}{\sqrt{a^2-b^2}}, \quad a > b \geq 0.$

627. $\int_0^{2\pi} \frac{dx}{1+a\cos x}\,dx = \frac{2\pi}{\sqrt{1-a^2}}, \quad |a| < 1.$

628. $\int_0^\infty \frac{\cos ax - \cos bx}{x}\,dx = \log\left|\frac{b}{a}\right|.$

629. $\int_0^{\pi/2} \frac{dx}{a^2\sin^2 x + b^2\cos^2 x}\,dx = \frac{\pi}{2\,|ab|}.$

630. $\int_0^{\pi/2} \frac{dx}{(a^2\sin^2 x + b^2\cos^2 x)^2}\,dx = \frac{\pi(a^2+b^2)}{4a^3b^3}, \quad a > 0,\ b > 0.$

631. $\displaystyle\int_0^{\pi/2} \sin^{n-1} x \cos^{m-1} x \, dx = \frac{1}{2} B\left(\frac{n}{2}, \frac{m}{2}\right),$
m is a positive integer, n is a positive integer.

632. $\displaystyle\int_0^{\pi/2} \sin^{2n+1} x \, dx = \frac{(2n)!!}{(2n+1)!!},$ n is a positive integer.

633. $\displaystyle\int_0^{\pi/2} \sin^{2n} x \, dx = \frac{(2n-1)!!}{(2n)!!} \frac{\pi}{2},$ n is a positive integer.

634. $\displaystyle\int_0^{\pi/2} \frac{x}{\sin x} \, dx = 2\left(\frac{1}{1^2} - \frac{1}{3^2} + \frac{1}{5^2} - \frac{1}{7^2} + \dots\right).$

635. $\displaystyle\int_0^{\pi/2} \frac{dx}{1 + \tan^m x} \, dx = \frac{\pi}{4},$ m is real.

636. $\displaystyle\int_0^{\pi/2} \sqrt{\cos x} \, dx = \frac{(2\pi)^{3/2}}{(\Gamma(1/4))^2}.$

637. $\displaystyle\int_0^{\pi/2} \tan^h x \, dx = \frac{\pi}{2\cos\left(\frac{h\pi}{2}\right)},$ $0 < h < 1.$

638. $\displaystyle\int_0^{\pi/2} \frac{\tan^{-1} ax - \tan^{-1} bx}{x} \, dx = \frac{\pi}{2} \log \frac{a}{b},$ $a > 0,\ b > 0.$

639. $\displaystyle\int_0^{\infty} e^{-ax} \, dx = \frac{1}{a},$ $a > 0.$

640. $\displaystyle\int_0^{\infty} \frac{e^{-ax} - e^{-bx}}{x} \, dx = \log \frac{b}{a},$ $a > 0,\ b > 0.$

641. $$\int_0^{\infty} x^n e^{-ax} \, dx = \begin{cases} \dfrac{\Gamma(n+1)}{a^{n+1}}, & a > 0,\ n > -1, \\ \text{or} & \\ \dfrac{n!}{a^{n+1}}, & a > 0,\ n \text{ is a positive integer.} \end{cases}$$

642. $\displaystyle\int_0^{\infty} x^n e^{-ax^p} \, dx = \frac{\Gamma((n+1)/p)}{pa^{(n+1)/p}},$ $a > 0,\ p > 0,\ n > -1.$

643. $\displaystyle\int_0^{\infty} e^{-a^2 x^2} \, dx = \frac{1}{2a}\sqrt{\pi},$ $a > 0.$

644. $\displaystyle\int_0^{b} e^{-ax^2} \, dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \operatorname{erf}\left(b\sqrt{a}\right),$ $a > 0.$

645. $\displaystyle\int_b^{\infty} e^{-ax^2} \, dx = \frac{1}{2}\sqrt{\frac{\pi}{a}} \operatorname{erfc}\left(b\sqrt{a}\right),$ $a > 0.$

646. $\displaystyle\int_0^{\infty} x e^{-x^2} \, dx = \frac{1}{2}.$

647. $\displaystyle\int_0^{\infty} x^2 e^{-x^2} \, dx = \frac{\sqrt{\pi}}{4}.$

648. $\displaystyle\int_0^{\infty} x^{2n} e^{-ax^2} \, dx = \frac{(2n-1)!!}{2(2a)^n}\sqrt{\frac{\pi}{a}},$ $a > 0,\ n > 0.$

649. $\displaystyle\int_0^{\infty} x^{2n+1} e^{-ax^2} \, dx = \frac{n!}{2a^{n+1}},$ $a > 0,\ n > -1.$

650. $\int_0^1 x^m e^{-ax}\,dx = \frac{m!}{a^{m+1}}\left[1 - e^{-a}\sum_{r=0}^{m}\frac{a^r}{r!}\right].$

651. $\int_0^\infty e^{\left(-x^2-a^2/x^2\right)}\,dx = \frac{e^{-2|a|}\sqrt{\pi}}{2}.$

652. $\int_0^\infty e^{\left(-ax^2-b/x^2\right)}\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a}}e^{-2\sqrt{ab}}, \quad a>0,\ b>0.$

653. $\int_0^\infty \sqrt{x}e^{-ax}\,dx = \frac{1}{2a}\sqrt{\frac{\pi}{a}}, \quad a>0.$

654. $\int_0^\infty \frac{e^{-ax}}{\sqrt{x}}\,dx = \sqrt{\frac{\pi}{a}}, \quad a>0.$

655. $\int_0^\infty e^{-ax}\cos mx\,dx = \frac{a}{a^2+m^2}, \quad a>0.$

656. $\int_0^\infty e^{-ax}\cos(bx+c)\,dx = \frac{a\cos c - b\sin c}{a^2+b^2}, \quad a>0.$

657. $\int_0^\infty e^{-ax}\sin mx\,dx = \frac{m}{a^2+m^2}, \quad a>0.$

658. $\int_0^\infty e^{-ax}\sin(bx+c)\,dx = \frac{b\cos c + a\sin c}{a^2+b^2}, \quad a>0.$

659. $\int_0^\infty xe^{-ax}\sin bx\,dx = \frac{2ab}{(a^2+b^2)^2}, \quad a>0.$

660. $\int_0^\infty xe^{-ax}\cos bx\,dx = \frac{a^2-b^2}{(a^2+b^2)^2}, \quad a>0.$

661. $\int_0^\infty x^n e^{-ax}\sin bx\,dx = \frac{n!\left[(a+ib)^{n+1}-(a-ib)^{n+1}\right]}{2i(a^2+b^2)^{n+1}}, \quad a>0.$

662. $\int_0^\infty x^n e^{-ax}\cos bx\,dx = \frac{n!\left[(a-ib)^{n+1}+(a+ib)^{n+1}\right]}{2(a^2+b^2)^{n+1}}, \quad a>0,\ n>-1.$

663. $\int_0^\infty \frac{e^{-ax}\sin x}{x}\,dx = \cot^{-1} a, \quad a>0.$

664. $\int_0^\infty e^{-a^2x^2}\cos bx\,dx = \frac{\sqrt{\pi}}{2|a|}\exp^{-b^2/(4a^2)}, \quad ab>0.$

665. $\int_0^\infty e^{-x\cos\phi}x^{b-1}\sin(x\sin\phi)\,dx = \Gamma(b)\sin(b\phi), \quad b>0,\ -\frac{\pi}{2}<\phi<\frac{\pi}{2}.$

666. $\int_0^\infty e^{-x\cos\phi}x^{b-1}\cos(x\sin\phi)\,dx = \Gamma(b)\cos(b\phi), \quad b>0,\ -\frac{\pi}{2}<\phi<\frac{\pi}{2}.$

667. $\int_0^\infty x^{b-1}\cos x\,dx = \Gamma(b)\cos\left(\frac{b\pi}{2}\right), \quad 0<b<1.$

668. $\int_0^\infty x^{b-1}\sin x\,dx = \Gamma(b)\sin\left(\frac{b\pi}{2}\right), \quad 0<b<1.$

669. $\int_0^1 (\log x)^n\,dx = (-1)^n n!, \quad n>-1.$

670. $\int_0^1 \sqrt{\log\frac{1}{x}}\,dx = \frac{\sqrt{\pi}}{2}.$

671. $\displaystyle\int_0^1 \left(\log \frac{1}{x}\right)^n dx = n!.$

672. $\displaystyle\int_0^1 x \log(1-x)\, dx = -\frac{3}{4}.$

673. $\displaystyle\int_0^1 x \log(1+x)\, dx = \frac{1}{4}.$

674. $\displaystyle\int_0^1 x^m (\log x)^n\, dx = \frac{(-1)^n \Gamma(n+1)}{(m+1)^{m+1}}, \quad m > -1,$ n is a non-negative integer.

675. $\displaystyle\int_0^1 \frac{\log x}{1+x}\, dx = -\frac{\pi^2}{12}.$

676. $\displaystyle\int_0^1 \frac{\log x}{1-x}\, dx = -\frac{\pi^2}{6}.$

677. $\displaystyle\int_0^1 \frac{\log(1+x)}{x}\, dx = \frac{\pi^2}{12}.$

678. $\displaystyle\int_0^1 \frac{\log(1-x)}{x}\, dx = -\frac{\pi^2}{6}.$

679. $\displaystyle\int_0^1 (\log x) \log(1+x)\, dx = 2 - 2\log 2 - \frac{\pi^2}{12}.$

680. $\displaystyle\int_0^1 (\log x) \log(1-x)\, dx = 2 - \frac{\pi^2}{6}.$

681. $\displaystyle\int_0^1 \frac{\log x}{1-x^2}\, dx = -\frac{\pi^2}{8}.$

682. $\displaystyle\int_0^1 \log\left(\frac{1+x}{1-x}\right) \frac{dx}{x} = \frac{\pi^2}{4}.$

683. $\displaystyle\int_0^1 \frac{\log x}{\sqrt{1-x^2}}\, dx = -\frac{\pi}{2} \log 2.$

684. $\displaystyle\int_0^1 x^m \left[\log\left(\frac{1}{x}\right)\right]^n dx = \frac{\Gamma(n+1)}{(m+1)^{n+1}}, \quad m > -1,\ n > -1.$

685. $\displaystyle\int_0^1 \frac{x^p - x^q}{\log x}\, dx = \log\left(\frac{p+1}{q+1}\right), \quad p > -1,\ q > -1.$

686. $\displaystyle\int_0^1 \frac{dx}{\sqrt{-\log x}} = \sqrt{\pi}.$

687. $\displaystyle\int_0^\infty \log\left(\frac{e^x+1}{e^x-1}\right) dx = \frac{\pi^2}{4}.$

688. $\displaystyle\int_0^{\pi/2} \log \sin x\, dx = \int_0^{\pi/2} \log \cos x\, dx = -\frac{\pi}{2} \log 2.$

689. $\displaystyle\int_0^{\pi/2} \log \sec x\, dx = \int_0^{\pi/2} \log \operatorname{cosec} x\, dx = \frac{\pi}{2} \log 2.$

690. $\displaystyle\int_0^{\pi} x \log \sin x\, dx = -\frac{\pi^2}{2} \log 2.$

691. $\displaystyle\int_0^{\pi/2} (\sin x) \log \sin x\, dx = \log 2 - 1.$

692. $\int_0^{\pi/2} \log \tan x \, dx = 0.$

693. $\int_0^{\pi} \log (a \pm b \cos x) \, dx = \pi \log \left(\frac{a + \sqrt{a^2 - b^2}}{2} \right), \quad a \geq b.$

694. $\int_0^{\pi} \log (a^2 - 2ab \cos x + b^2) \, dx = \begin{cases} 2\pi \log a, & a \geq b > 0, \\ \quad \text{or} & \\ 2\pi \log b, & b \geq a > 0. \end{cases}$

695. $\int_0^{\infty} \frac{\sin ax}{\sinh bx} \, dx = \frac{\pi}{2b} \tanh \frac{a\pi}{2\,|b|}.$

696. $\int_0^{\infty} \frac{\cos ax}{\cosh bx} \, dx = \frac{\pi}{2b} \operatorname{sech} \frac{a\pi}{2b}.$

697. $\int_0^{\infty} \frac{dx}{\cosh ax} = \frac{\pi}{2\,|a|}.$

698. $\int_0^{\infty} \frac{x}{\sinh ax} \, dx = \frac{\pi^2}{4a^2}, \quad a \geq 0.$

699. $\int_0^{\infty} e^{-ax} \cosh (bx) \, dx = \frac{a}{a^2 - b^2}, \quad |b| < a.$

700. $\int_0^{\infty} e^{-ax} \sinh (bx) \, dx = \frac{b}{a^2 - b^2}, \quad |b| < a.$

701. $\int_0^{\infty} \frac{\sinh ax}{e^{bx} + 1} \, dx = \frac{\pi}{2b} \csc \frac{a\pi}{b} - \frac{1}{2a}, \quad b \geq 0.$

702. $\int_0^{\infty} \frac{\sinh ax}{e^{bx} - 1} \, dx = \frac{1}{2a} - \frac{\pi}{2b} \cot \frac{a\pi}{b}, \quad b \geq 0.$

703.

$$\int_0^{\pi/2} \frac{dx}{\sqrt{1 - k^2 \sin^2 x}} = \frac{\pi}{2} \left[1 + \left(\frac{1}{2} \right)^2 k^2 + \left(\frac{1 \cdot 3}{2 \cdot 4} \right)^2 k^4 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \right)^2 k^6 + \dots \right], \quad k^2 < 1.$$

704. $$\int_0^{\pi/2} \frac{dx}{(1 - k^2 \sin^2 x)^{3/2}} = \frac{\pi}{2} \left[1 + \left(\frac{1}{2} \right)^2 3k^2 + \left(\frac{1 \cdot 3}{2 \cdot 4} \right)^2 5k^4 + \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \right)^2 7k^6 + \dots \right], \quad k^2 < 1.$$

705. $$\int_0^{\pi/2} \sqrt{1 - k^2 \sin^2 x} \, dx = \frac{\pi}{2} \left[1 - \left(\frac{1}{2} \right)^2 k^2 - \left(\frac{1 \cdot 3}{2 \cdot 4} \right)^2 \frac{k^4}{3} - \left(\frac{1 \cdot 3 \cdot 5}{2 \cdot 4 \cdot 6} \right)^2 \frac{k^6}{5} - \dots \right], \quad k^2 < 1.$$

706. $\int_0^{\infty} e^{-x} \log x \, dx = -\gamma.$

707. $\int_0^{\infty} e^{-x^2} \log x \, dx = -\frac{\sqrt{\pi}}{4}(\gamma + 2 \log 2).$

708. $\int_0^{\infty} \left(\frac{1}{1 - e^{-x}} - \frac{1}{x} \right) e^{-x} \, dx = \gamma.$

709. $\int_0^{\infty} \frac{1}{x} \left(\frac{1}{1 - e^{-x}} - \frac{1}{x} \right) dx = \gamma.$

5.5.1 SPECIAL FUNCTIONS DEFINED BY INTEGRALS

Not all integrals of elementary functions (sines, cosines, rational functions, and others) can be evaluated in terms of elementary functions. For example, the integral $\int e^{-x^2}\,dx$ is represented by the special function "erf(x)" (see page 475).

The dilogarithm function is defined by $\operatorname{Li}_2(x) = -\int_0^x \ln(1-t)/t\,dt$ (see page 488). All integrals of the form $\int^x P(x,\sqrt{R})\log Q(x,\sqrt{R})\,dx$, where P and Q are rational functions and $R = A + Bx + Cx^2$, can be evaluated in terms of elementary functions and dilogarithms.

All integrals of the form $\int_x R(x,\sqrt{T(x)})\,dx$, where R is a rational function of its arguments and $T(x)$ is a third or fourth order polynomial, can be integrated in terms of elementary functions and elliptic functions (see page 470).

5.6 ORDINARY DIFFERENTIAL EQUATIONS

5.6.1 LINEAR DIFFERENTIAL EQUATIONS

A linear differential equation is one that can be written in the form

$$b_n(x)y^{(n)} + b_{n-1}(x)y^{(n-1)} + \cdots + b_1(x)y' + b_0(x)y = R(x) \tag{5.6.1}$$

or $p(D)y = R(x)$, where D is the differentiation operator ($Dy = dy/dx$), $p(D)$ is a polynomial in D with coefficients $\{b_i\}$ depending on x, and $R(x)$ is an arbitrary function. In this notation, a power of D denotes repeated differentiation, that is, $D^n y = d^n y/dx^n$. For such an equation, the general solution has the form

$$y(x) = y_h(x) + y_p(x) \tag{5.6.2}$$

where $y_h(x)$ is a homogeneous solution and $y_p(x)$ is the particular solution. These functions satisfy $p(D)y_h = 0$ and $p(D)y_p = R(x)$.

5.6.1.1 Vector representation

Equation (5.6.1) can be written in the form $\dfrac{d\mathbf{y}}{dx} = A(x)\mathbf{y} + \mathbf{r}(x)$ where

$$\mathbf{y} = \begin{bmatrix} y \\ y' \\ y'' \\ \vdots \\ y^{(n-1)} \end{bmatrix}, \quad A(x) = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & 0 & & 1 \\ -\frac{b_0}{b_n} & -\frac{b_1}{b_n} & -\frac{b_2}{b_n} & \dots & -\frac{b_{n-1}}{b_n} \end{bmatrix}, \quad \mathbf{r}(x) = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \frac{R}{b_n} \end{bmatrix}.$$

5.6.1.2 Second order linear constant coefficient equation

Consider $ay'' + by' + cy = 0$, where a, b, and c are real constants. Let m_1 and m_2 be the roots of $am^2 + bm + c = 0$. There are three forms of the solution:

1. If m_1 and m_2 are real and distinct, then $y(x) = c_1 e^{m_1 x} + c_2 e^{m_2 x}$
2. If m_1 and m_2 are real and equal, then $y(x) = c_1 e^{m_1 x} + c_2 x e^{m_1 x}$
3. If $m_1 = p + iq$ and $m_2 = p - iq$ (with $p = -b/2a$ and $q = \sqrt{4ac - b^2}/2a$), then $y(x) = e^{px}(c_1 \cos qx + c_2 \sin qx)$

Consider $ay'' + by' + cy = R(x)$, where a, b, and c are real constants. Let m_1 and m_2 be as above.

1. If m_1 and m_2 are real and distinct, then $y(x) = C_1 e^{m_1 x} + C_2 e^{m_2 x} + e^{m_1 x}/(m_1 - m_2) \int^x e^{-m_1 z} R(z)\, dz + e^{m_2 x}/(m_2 - m_1) \int^x e^{-m_2 z} R(z)\, dz$.
2. If m_1 and m_2 are real and equal, then $y(x) = C_1 e^{m_1 x} + C_2 x e^{m_1 x} + x e^{m_1 x} \int^x e^{-m_1 z} R(z)\, dz - e^{m_1 x} \int^x z e^{-m_1 z} R(z)\, dz$.
3. If $m_1 = p + iq$ and $m_2 = p - iq$, then $y(x) = e^{px}(c_1 \cos qx + c_2 \sin qx) + e^{px} \sin qx/q \int^x e^{-pz} R(z) \cos qz\, dz - e^{px} \cos qx/q \int^x e^{-pz} R(z) \sin qz\, dz$.

5.6.1.3 Homogeneous solutions

For the special case of a linear differential equation with constant coefficients (i.e., the $\{b_i\}$ in Equation (5.6.1) are constants), the procedure for finding the homogeneous solution is as follows:

1. Factor the polynomial $p(D)$ into real and complex linear factors, just as if D were a variable instead of an operator.
2. For each non-repeated linear factor of the form $(D - a)$, where a is real, write a term of the form ce^{ax}, where c is an arbitrary constant.
3. For each repeated real linear factor of the form $(D - a)^m$, write the following sum of m terms
$$c_1 e^{ax} + c_2 x e^{ax} + c_3 x^2 e^{ax} + \cdots + c_m x^{m-1} e^{ax} \tag{5.6.3}$$
where the c_i's are arbitrary constants.
4. For each non-repeated complex conjugate pair of factors of the form $(D - a + ib)(D - a - ib)$, write the following two terms
$$c_1 e^{ax} \cos bx + c_2 e^{ax} \sin bx. \tag{5.6.4}$$
5. For each repeated complex conjugate pair of factors of the form $(D - a + ib)^m (D - a - ib)^m$, write the following $2m$ terms
$$c_1 e^{ax} \cos bx + c_2 e^{ax} \sin bx + c_3 x e^{ax} \cos bx + c_4 x e^{ax} \sin bx + \dots$$
$$+ c_{2m-1} x^{m-1} e^{ax} \cos bx + c_{2m} x^{m-1} e^{ax} \sin bx. \tag{5.6.5}$$
6. The sum of all the terms thus written is the homogeneous solution.

EXAMPLE For the linear equation

$$y^{(7)} - 14y^{(6)} + 81y^{(5)} - 252y^{(4)} + 455y^{(3)} - 474y'' + 263y' - 60y = 0,$$

$p(D)$ factors as $p(D) = (D-1)^3(D-(2+i))(D-(2-i))(D-3)(D-4)$. The roots are thus $\{1, 1, 1, 2+i, 2-i, 3, 4\}$. Hence, the homogeneous solution has the form

$$y_h(x) = \left(c_0 + c_1 x + c_2 x^2\right) e^x + \left(c_3 \sin x + c_4 \cos x\right) e^{2x} + c_5 e^{3x} + c_6 e^{4x}$$

where $\{c_0, \dots, c_6\}$ are arbitrary constants.

5.6.1.4 Particular solutions

The following are solutions for some specific ordinary differential equations. These assume that $P(x)$ is a polynomial of degree n and $\{a, b, p, q, r, s\}$ are constants. If you wish to replace "sin" with "cos" in $R(x)$, then use the given result but replace "sin" by "cos," and replace "cos" by "$-$ sin."

1. Particular solutions to $y' - ay = R(x)$

   (a) If $R(x) = e^{rx}$, $\quad y = e^{rx}/(r-a)$

   (b) $R(x) = \sin sx$,
   $$y = -\frac{a \sin sx + s \cos sx}{a^2 + s^2} = -\left(a^2 + s^2\right)^{-1/2} \sin\left(sx + \tan^{-1}\frac{s}{a}\right)$$

   (c) $R(x) = P(x)$,
   $$y = -\frac{1}{a}\left[P(x) + \frac{P'(x)}{a} + \frac{P''(x)}{a^2} + \dots + \frac{P^{(n)}(x)}{a^n}\right]$$

   (d) $R(x) = e^{rx} \sin sx$,
   Replace a by $(a-r)$ in formula (1b) and multiply solution by e^{rx}

   (e) $R(x) = P(x)e^{rx}$,
   Replace a by $(a-r)$ in formula (1c) and multiply solution by e^{rx}

   (f) $R(x) = P(x) \sin sx$,
   $$\begin{aligned} y = &- \sin sx\Big[\frac{a}{a^2+s^2}P(x) + \frac{a^2 - s^2}{(a^2+s^2)^2}P'(x) \\ &\qquad + \dots + \frac{a^k - \binom{k}{2}a^{k-2}s^2 + \binom{k}{4}a^{k-4}s^4 - \dots}{(a^2+s^2)^k}P^{(k-1)}(x) + \dots\Big] \\ &- \cos sx\Big[\frac{s}{a^2+s^2}P(x) + \frac{2as}{(a^2+s^2)^2}P'(x) \\ &\qquad + \dots + \frac{\binom{k}{1}a^{k-1}s - \binom{k}{3}a^{k-3}s^3 + \dots}{(a^2+s^2)^k}P^{(k-1)}(x) + \dots\Big]. \end{aligned}$$

   (g) $R(x) = P(x)e^{rx} \sin sx$,
   Replace a by $(a-r)$ in formula (1f) and multiply solution by e^{rx}

   (h) If $R(x) = e^{ax}$, $\quad y = xe^{ax}$

   (i) If $R(x) = e^{ax} \sin sx$, $\quad y = -e^{ax} \cos sx/s$

   (j) If $R(x) = P(x)e^{ax}$, $\quad y = e^{ax} \int^x P(z)\,dz$

(k) $R(x) = P(x)e^{ax}\sin sx$,

$$y = \frac{e^{ax}\sin sx}{s}\left[\frac{P'(x)}{s} - \frac{P'''(x)}{s^3} + \frac{P^{(5)}(x)}{s^5} + \dots\right] - \frac{e^{ax}\cos sx}{s}\left[P(x) - \frac{P''(x)}{s^2} + \frac{P^{(4)}(x)}{s^4} + \dots\right].$$

2. Particular solutions to $y'' - 2ay' + a^2y = R(x)$

(a) If $R(x) = e^{rx}$, $\quad y = e^{rx}/(r-a)^2$.

(b) If $R(x) = \sin sx$,

$$y = \frac{(a^2 - s^2)\sin sx + 2as\cos sx}{(a^2+s^2)^2} = \frac{1}{a^2+s^2}\sin\left(sx + \tan^{-1}\frac{2as}{a^2-s^2}\right).$$

(c) $R(x) = P(x)$,

$$y = \frac{1}{a^2}\left[P(x) + \frac{2P'(x)}{a} + \frac{3P''(x)}{a^2} + \dots + \frac{(n+1)P^{(n)}(x)}{a^n}\right]$$

(d) $R(x) = e^{rx}\sin sx$,

Replace a by $(a-r)$ in formula (2a) and multiply solution by e^{rx}

(e) $R(x) = P(x)e^{rx}$,

Replace a by $(a-r)$ in formula (2b) and multiply solution by e^{rx}.

(f) If $R(x) = P(x)\sin sx$,

$$y = \sin sx\Big[\frac{a^2-s^2}{(a^2+s^2)^2}P(x) + 2\frac{a^2-3as^2}{(a^2+s^2)^3}P'(x).$$
$$+ \dots + (k-1)\frac{a^k - \binom{k}{2}a^{k-2}s^2 + \binom{k}{4}a^{k-4}s^4 - \dots}{(a^2+s^2)^k}P^{(k-2)}(x) + \dots\Big]$$
$$+ \cos sx\Big[\frac{2as}{(a^2+s^2)^2}P(x) + 2\frac{3a^2s - s^3}{(a^2+s^2)^3}P'(x)$$
$$+ \dots + (k-1)\frac{\binom{k}{1}a^{k-1}s - \binom{k}{3}a^{k-3}s^3 + \dots}{(a^2+s^2)^k}P^{(k-2)}(x) + \dots\Big]$$

(g) $R(x) = P(x)e^{rx}\sin sx$,

Replace a by $(a-r)$ in formula (2f) and multiply solution by e^{rx}.

(h) If $R(x) = e^{ax}$, $\quad y = x^2e^{ax}/2$

(i) If $R(x) = e^{ax}\sin sx$, $\quad y = -e^{ax}\sin sx/s^2$

(j) If $R(x) = P(x)e^{ax}$, $\quad y = e^{ax}\int^x\int^y P(z)\,dz\,dy$

(k) $R(x) = P(x)e^{ax}\sin sx$,

$$y = -\frac{e^{ax}\sin sx}{s^2}\left[P(x) - \frac{3P''(x)}{s^2} + \frac{5P^{(4)}(x)}{s^4} + \dots\right] - \frac{e^{ax}\cos sx}{s^2}\left[\frac{2P(x)}{s} - \frac{4P'''(x)}{s^3} + \frac{6P^{(5)}(x)}{s^5} + \dots\right]$$

3. Particular solutions to $y'' + qy = R(x)$
   (a) If $R(x) = e^{rx}$, $\quad y = e^{rx}/(r^2+q)$
   (b) If $R(x) = \sin sx$, $\quad y = \sin sx/(q - s^2)$
   (c) $R(x) = P(x)$,
   $$y = \frac{1}{q}\left[P(x) - \frac{P''(x)}{q} + \frac{P^{(4)}(x)}{q^2} + \cdots + (-1)^k \frac{P^{(2k)}(x)}{q^k} + \cdots\right]$$
   (d) If $R(x) = e^{rx}\sin sx$, $\quad y = \frac{(r^2-s^2+q)e^{rx}\sin sx - 2rse^{rx}\cos sx}{(r^2-s^2+q)^2+(2rs)^2} =$
   $$\frac{e^{rx}}{\sqrt{(r^2-s^2+q)^2+(2rs)^2}}\sin\left[sx - \tan^{-1}\frac{2rs}{r^2-s^2+q}\right]$$
   (e) $R(x) = P(x)e^{rx}$,
   $$y = \frac{e^{rx}}{q+r^2}\left[P(x) - \frac{2r}{q+r^2}P'(x) + \frac{3r^2-q}{(q+r^2)^2}P''(x) + \cdots + (-1)^{k-1}\frac{\binom{k}{1}r^{k-1} - \binom{k}{3}r^{k-3}q + \dots}{(q+r^2)^{k-1}}P^{(k-1)}(x) + \dots\right]$$
   (f) $R(x) = P(x)\sin sx$,
   $$y = \frac{\sin sx}{q-s^2}\left[P(x) - \frac{3s^2+q}{(q-s^2)^2}P''(x) + \cdots + (-1)^k\frac{\binom{2k+1}{1}s^{2k} + \binom{2k+1}{3}s^{2k-2}q + \dots}{(q-s^2)^{2k}}P^{(2k)}(x) + \dots\right]$$
   $$- \frac{s\cos sx}{q-s^2}\left[\frac{2P'(x)}{(q-s^2)} - \frac{4s^2+4q}{(q-s^2)^3}P'''(x) + \cdots + (-1)^{k+1}\frac{\binom{2k}{1}s^{2k-2} + \binom{2k}{3}s^{2k-4}q + \dots}{(q-s^2)^{2k-1}}P^{(2k-1)}(x) + \dots\right]$$
4. Particular solutions to $y'' + b^2y = R(x)$
   (a) If $R(x) = \sin bx$, $\quad y = -x\cos bx/2b$
   (b) If $R(x) = P(x)\sin bx$, $\quad y = \frac{\sin bx}{(2b)^2}\left[P(x) - \frac{P''(x)}{(2b)^2} + \frac{P^{(4)}(x)}{(2b)^4} + \cdots\right] - \frac{\cos bx}{2b}\int\left[P(x) - \frac{P''(x)}{(2b)^2} + \cdots\right]dx$
5. Particular solutions to $y'' + py' + qy = R(x)$
   (a) If $R(x) = e^{rx}$, $\quad y = e^{rx}/(r^2+pr+q)$
   (b) If $R(x) = \sin sx$, $\quad y = \frac{(q-s^2)\sin sx - ps\cos sx}{(q-s^2)^2+(ps)^2} = \frac{\sin\left(sx - \tan^{-1}\frac{ps}{q-s^2}\right)}{\sqrt{(q-s^2)^2+(ps)^2}}$
   (c) If $R(x) = P(x)$,
   $$y = \frac{1}{q}\left[P(x) - \frac{p}{q}P'(x) + \frac{p^2-q}{q^2}P''(x) - \frac{p^2-2pq}{q^2}P'''(x) + \cdots + (-1)^n\frac{p^n - \binom{n-1}{1}p^{n-2}q + \binom{n-2}{2}p^{n-4}q^2 - \dots}{q^n}P^{(n)}(x)\right]$$
   (d) $R(x) = e^{rx}\sin sx$,
   In (5b): replace p by $(p+2r)$, q by $(q+pr+r^2)$, multiply by e^{rx}.

(e) $R(x) = P(x)e^{rx}$,
In (5c): replace p by $(p + 2r)$, q by $(q + pr + r^2)$, multiply by e^{rx}.

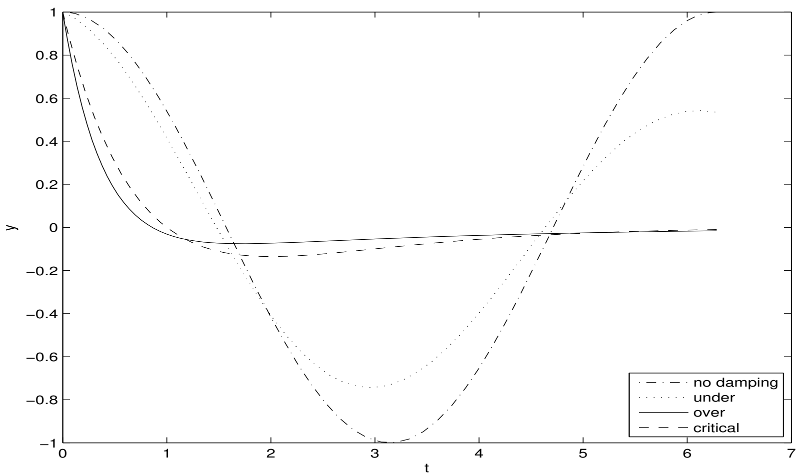
6. Particular solutions to $(D - a)^n y = R(x)$
   (a) If $R(x) = e^{rx}$, $y = e^{rx}/(r-a)^n$
   (b) If $R(x) = \sin sx$, $y = \frac{(-1)^n}{(a^2+s^2)^n}\Big[\Big(a^n - \binom{n}{2}a^{n-2}s^2 + \binom{n}{4}a^{n-4}s^4 - \dots\Big)\sin sx + \Big(\binom{n}{1}a^{n-1}s + \binom{n}{3}a^{n-3}s^3 + \dots\Big)\cos sx\Big]$
   (c) If $R(x) = P(x)$, $y = \frac{(-1)^n}{a^n}\left[P(x) + \binom{n}{1}\frac{P'(x)}{a} + \binom{n+1}{2}\frac{P''(x)}{a^2} + \binom{n+2}{3}\frac{P'''(x)}{a^3} + \dots\right]$
   (d) $R(x) = e^{rx}\sin sx$,
   Replace a by $(a - r)$ in formula (6b) and multiply solution by e^{rx}.
   (e) $R(x) = P(x)e^{rx}$,
   Replace a by $(a - r)$ in formula (6c) and multiply solution by e^{rx}.

5.6.1.5 Damping: none, under, over, and critical

Consider the linear ordinary differential equation $x'' + \mu x' + x = 0$. If the damping coefficient μ is positive, then all solutions decay to $x = 0$. If $\mu = 0$, the system is undamped and the solution oscillates without decaying. The value of μ such that the roots of the characteristic equation $\lambda^2 + \mu\lambda + 1 = 0$ are real and equal is the critical damping coefficient. If μ is less than (greater than) the critical damping coefficient, then the system is under (over) damped.

Consider four cases with the same initial values: $y(0) = 1$ and $y'(0) = 0$.

1. $y'' + y = 0$ Undamped
2. $y'' + 0.2y' + y = 0$ Underdamped
3. $y'' + 3y' + y = 0$ Overdamped
4. $y'' + 2y' + y = 0$ Critically damped



5.6.2 SOLUTION TECHNIQUES

Differential equation	Solution or solution technique
Autonomous equation $f(y^{(n)}, y^{(n-1)}, \dots, y'', y', y) = 0$	Change dependent variable to $u(y) = y'(x)$
Bernoulli's equation $y' + f(x)y = g(x)y^n$	Change dependent variable to $v(x) = (y(x))^{1-n}$
Clairaut's equation $f(xy' - y) = g(y')$	One solution is $f(xC - y) = g(C)$
Constant coefficient equation $a_0 y^{(n)} + a_1 y^{(n-1)} + \dots + a_{n-1}y' + a_n y = 0$	There are solutions of the form $y = x^k e^{\lambda x}$. See Section 5.6.1.3.
Dependent variable missing $f(y^{(n)}, y^{(n-1)}, \dots, y'', y', x) = 0$	Change dependent variable to $u(x) = y'(x)$
Euler's equation $a_0 x^n y^{(n)} + a_1 x^{n-1} y^{(n-1)} + \dots + a_{n-1}xy + a_n y = 0$	Change independent variable to $x = e^t$
Exact equation $M(x,y)\,dx + N(x,y)\,dy = 0$ with $\frac{\partial M}{\partial y} = \frac{\partial N}{\partial x}$	Integrate $M(x,y)$ with respect to x holding y constant, call this $m(x,y)$. Then $m(x,y) + \int \left(N - \frac{\partial m}{\partial y}\right) dy = C$
Homogeneous equation $y' = f\left(\frac{y}{x}\right)$	$\ln x = \int \frac{dv}{f(v) - v} + C$ unless $f(v) = v$, in which case $y = Cx$.
Linear first-order equation $y' + f(x)y = g(x)$	$y(x) = e^{-\int^x f(t)\,dt} \left[\int^x e^{\int^z f(t)\,dt} g(z)\,dz + C\right]$
Reducible to homogeneous $(a_1x + b_1y + c_1)\,dx + (a_2x + b_2y + c_2)\,dy = 0$ with $a_1/a_2 \neq b_1/b_2$	Change variables to $u = a_1x + b_1y + c$ and $v = a_2x + b_2y + c$
Reducible to separable $(a_1x + b_1y + c_1)\,dx + (a_2x + b_2y + c_2)\,dy = 0$ with $a_1/a_2 = b_1/b_2$	Change dependent variable to $u(x) = a_1x + b_1y$
Separation of variables $y' = f(x)g(y)$	$\int \frac{dy}{g(y)} = \int f(x)\,dx + C$

5.6.3 TRANSFORM TECHNIQUES

Transforms can sometimes be used to solve linear differential equations. Laplace transforms (page 508) are appropriate for initial value problems, while Fourier transforms (page 494) are appropriate for boundary value problems.

EXAMPLE Consider the linear second order equation $y'' + y = p(x)$, with the initial conditions $y(0) = 0$ and $y'(0) = 0$. Multiplying this equation by e^{-sx} and integrating with respect to x from 0 to ∞ results in

$$\int_0^\infty e^{-sx} y''(x)\,dx + \int_0^\infty e^{-sx} y(x)\,dx = \int_0^\infty e^{-sx} p(x)\,dx.$$

Integrating by parts, and recognizing that $Y(s) = \mathcal{L}[y(x)] = \int_0^\infty e^{-sx} y(x)\,dx$ is the Laplace transform of y, we simplify to

$$(s^2 + 1)Y(s) = \int_0^\infty e^{-sx} p(x)\,dx = \mathcal{L}[p(x)].$$

If $p(x) \equiv 1$, then $\mathcal{L}[p(x)] = s^{-1}$. The table of Laplace transforms (entry 20 in the table on page 525) shows that the $y(x)$ corresponding to $Y(s) = 1/[s(1 + s^2)]$ is $y(x) = \mathcal{L}^{-1}[Y(s)] = 1 - \cos x$.

5.6.4 INTEGRATING FACTORS

An integrating factor is a multiplicative term that makes a differential equation become exact; that is, form an exact differential. If the differential equation $M(x, y)\,dx + N(x, y)\,dy = 0$ is not exact (i.e., $M_y \neq N_x$), then it may be made exact by multiplying by the integrating factor.

1. If $\dfrac{1}{N}\left(\dfrac{\partial M}{\partial y} - \dfrac{\partial N}{\partial x}\right) = f(x)$, a function of x alone,

   then $u(x) = \exp\left(\int^x f(z)\,dz\right)$ is an integrating factor.

2. If $\dfrac{1}{M}\left(\dfrac{\partial N}{\partial y} - \dfrac{\partial M}{\partial x}\right) = g(y)$, a function of y alone,

   then $v(y) = \exp\left(\int^y g(z)\,dz\right)$ is an integrating factor.

EXAMPLE The equation $\dfrac{y}{x}\,dx + dy = 0$ has $\{M = y/x,\ N = 1\}$ and $f(x) = 1/x$. Hence $u(x) = \exp\left(\int^x \frac{1}{z}\,dz\right) = \exp(\log x) = x$ is an integrating factor. Multiplying the original equation by $u(x)$ results in $y\,dx + x\,dy = 0$ or $d(xy) = 0$.

5.6.5 VARIATION OF PARAMETERS

If the linear second order equation $L[y] = y'' + P(x)y' + Q(x)y = R(x)$ has the independent homogeneous solutions $u(x)$ and $v(x)$ (i.e., $L[u] = 0 = L[v]$), then the solution to the original equation is given by

$$y(x) = -u(x) \int \frac{v(x)R(x)}{W(u,v)}\, dx + v(x) \int \frac{u(x)R(x)}{W(u,v)}\, dx, \tag{5.6.6}$$

where $W(u,v) = uv' - u'v = \left| \begin{smallmatrix} u & v \\ u' & v' \end{smallmatrix} \right|$ is the Wronskian.

EXAMPLE The homogeneous solutions to $y'' + y = \csc x$ are $u(x) = \sin x$ and $v(x) = \cos x$. Here, $W(u,v) = -1$. Hence, $y(x) = \sin x \log(\sin x) - x \cos x$.

If the linear third order equation $L[y] = y''' + P_2(x)y'' + P(x)y' + Q(x)y = R(x)$ has the homogeneous solutions $y_1(x)$, $y_2(x)$, and $y_3(x)$ (i.e., $L[y_i] = 0$), then the solution to the original equation is given by

$$\begin{aligned} y(x) = y_1(x) \int \frac{\begin{vmatrix} 0 & y_2 & y_3 \\ 0 & y_2' & y_3' \\ R & y_2'' & y_3'' \end{vmatrix}}{W(y_1,y_2,y_3)}\, dx + y_2(x) \int \frac{\begin{vmatrix} y_1 & 0 & y_3 \\ y_1' & 0 & y_3' \\ y_1'' & R & y_3'' \end{vmatrix}}{W(y_1,y_2,y_3)}\, dx \\ + y_3(x) \int \frac{\begin{vmatrix} y_1 & y_2 & 0 \\ y_1' & y_2' & 0 \\ y_1'' & y_2'' & R \end{vmatrix}}{W(y_1,y_2,y_3)}\, dx \end{aligned} \tag{5.6.7}$$

where $W(y_1,y_2,y_3) = \begin{vmatrix} y_1 & y_2 & y_3 \\ y_1' & y_2' & y_3' \\ y_1'' & y_2'' & y_3'' \end{vmatrix}$ is the Wronskian.

5.6.6 GREEN'S FUNCTIONS

Let $L[y] = f(x)$ be an n^{th} order linear differential equation with linear and homogeneous boundary conditions $\{B_i[y] = \sum_{j=0}^{n-1} (a_{ij} y^{(j)}(x_0) + b_{ij} y^{(j)}(x_1)) = 0\}$, for $i = 1, 2, \ldots, n$. If there is a Green's function $G(x;z)$ that satisfies

$$\begin{aligned} L[G(x;z)] &= \delta(x-z), \\ B_i[G(x;z)] &= 0, \end{aligned} \tag{5.6.8}$$

where δ is Dirac's delta function, then the solution of the original system can be written as $y(x) = \int G(x;z) f(z)\, dz$, integrated over an appropriate region.

EXAMPLE To solve $y'' = f(x)$ with $y(0) = 0$ and $y(L) = 0$, the appropriate Green's function is

$$G(x;z) = \begin{cases} \dfrac{x(z-L)}{L} & \text{for } 0 \le x \le z, \\ \dfrac{z(x-L)}{L} & \text{for } z \le x \le L. \end{cases} \tag{5.6.9}$$

Hence, the solution is

$$y(x) = \int_0^L G(x;z)\, f(z)\, dz = \int_0^x \frac{z(x-L)}{L}\, f(z)\, dz + \int_x^L \frac{x(z-L)}{L}\, f(z)\, dz. \tag{5.6.10}$$

5.6.7 TABLE OF GREEN'S FUNCTIONS

The Green's function is $G(x,\xi)$ when $x \le \xi$ and $G(\xi,x)$ when $x \ge \xi$.

1. For the equation $\dfrac{d^2y}{dx^2} = f(x)$ with
   - (a) $y(0) = y(1) = 0$, $G(x,\xi) = -(1-\xi)x$,
   - (b) $y(0) = 0, y'(1) = 0$, $G(x,\xi) = -x$,
   - (c) $y(0) = -y(1), y'(0) = -y'(1)$, $G(x,\xi) = -\frac{1}{2}(x-\xi) - \frac{1}{4}$, and
   - (d) $y(-1) = y(1) = 0$, $G(x,\xi) = -\frac{1}{2}(x - \xi - x\xi + 1)$.
2. For the equation $\dfrac{d^2y}{dx^2} - y = f(x)$ with y finite $G(x,\xi) = -\frac{1}{2}e^{x-\xi}$.
3. For the equation $\dfrac{d^2y}{dx^2} + k^2y = f(x)$ with
   - (a) $y(0) = y(1) = 0$, $G(x,\xi) = -\dfrac{\sin kx \sin k(1-\xi)}{k \sin k}$,
   - (b) $y(-1) = y(1)$ and $y'(-1) = y'(1)$ $G(x,\xi) = \dfrac{\cos k(x-\xi+1)}{2k \sin k}$.
4. For the equation $\dfrac{d^2y}{dx^2} - k^2y = f(x)$ with
   - (a) $y(0) = y(1) = 0$, $G(x,\xi) = -\dfrac{\sinh kx \sinh k(1-\xi)}{k \sinh k}$,
   - (b) $y(-1) = y(1)$ and $y'(-1) = y'(1)$ $G(x,\xi) = -\dfrac{\cosh k(x-\xi+1)}{2k \sinh k}$.
5. For the equation $\dfrac{d}{dx}\left(x\dfrac{dy}{dx}\right) = f(x)$, with $y(0)$ finite and $y(1) = 0$

   $G(x,\xi) = \ln \xi$
6. For the equation $\dfrac{d^4y}{dx^4} = f(x)$, with $y(0) = y'(0) = y(1) = y'(1) = 0$

   $G(x,\xi) = -\dfrac{x^2(\xi-1)^2}{6}(2x\xi + x - 3\xi)$

5.6.8 STOCHASTIC DIFFERENTIAL EQUATIONS

A stochastic differential equation for the unknown $X(t)$ has the form (here, a and b are given):

$$dX(t) = a(X(t))\,dt + b(X(t))\,dB(t) \tag{5.6.11}$$

where $B(t)$ is a Brownian motion. Brownian motion has a Gaussian probability distribution and independent increments. The probability density function $f_{X(t)}$ for $X(t)$ satisfies the forward Kolmogorov equation

$$\frac{\partial}{\partial t} f_{X(t)}(x) = \frac{1}{2}\frac{\partial^2}{\partial x^2}\left[b^2(x) f_{X(t)}(x)\right] - \frac{\partial}{\partial x}\left[a(x) f_{X(t)}(x)\right]. \tag{5.6.12}$$

The conditional expectation of the function $\phi(X(t))$, which is $u(t,x) = \mathrm{E}\left[\phi(X(t)) \mid X(0) = x\right]$, satisfies

$$\frac{\partial}{\partial t}u(t,x) = \frac{1}{2}b^2(x)\frac{\partial^2}{\partial x^2}u(t,x) + a(x)\frac{\partial}{\partial x}u(t,x) \quad \text{with} \quad u(0,x) = \phi(x). \tag{5.6.13}$$

5.6.9 LIE GROUPS

An algorithm for integrating second order ordinary differential equations is:

1. Determine the admitted Lie algebra L_r, where r is the dimension.
2. If $r < 2$, then Lie groups are not useful for the given equation.
3. If $r > 2$, determine a subalgebra $L_2 \subset L_r$.
4. From the commutator and pseudoscalar product, change the basis to obtain one of the four cases in the table below.
5. Introduce canonical variables specified by the change of basis.
6. Integrate the new equation. Rewrite solution in terms of the original variables.

The invertible transformation $\{\overline{x} = \phi(x, y, a), \overline{y} = \psi(x, y, a)\}$ forms a 1 parameter group if $\phi(\overline{x}, \overline{y}, b) = \phi(x, y, a + b)$ and $\psi(\overline{x}, \overline{y}, b) = \psi(x, y, a + b)$. For small a,

$$\overline{x} = x + a\xi(x, y) + O(a^2) \qquad \text{and} \qquad \overline{y} = y + a\eta(x, y) + O(a^2) \tag{5.6.14}$$

If $D = \frac{\partial}{\partial x} + y'\frac{\partial}{\partial y} + y''\frac{\partial}{\partial y'} + \ldots$, then the derivatives of the new variables are

$$\begin{aligned} \overline{y}' &= \frac{d\overline{y}}{d\overline{x}} = \frac{D\psi}{D\phi} = \frac{\psi_x + y'\psi_y}{\phi_x + y'\phi_y} = P(x, y, y', a) = y' + a\zeta_1 + O(a^2), \quad \text{and} \\ \overline{y}'' &= \frac{d\overline{y}'}{d\overline{x}} = \frac{DP}{D\phi} = \frac{P_x + y'P_y + y''P_{y'}}{\phi_x + y'\phi_y} = y'' + a\zeta_2 + O(a^2), \end{aligned} \tag{5.6.15}$$

where

$$\begin{aligned} \zeta_1 &= D(\eta) - y'D(\xi) = \eta_x + (\eta_y - \xi_x)y' - y'^2\xi_y, \qquad \text{and} \\ \zeta_2 &= D(\zeta_1) - y''D(\xi) = \eta_{xx} + (2\eta_{xy} - \xi_{xx})y' + (\eta_{yy} - 2\xi_{xy})y'^2 \\ &\qquad - y'^3\xi_{yy} + (\eta_y - 2\xi_x - 3y'\xi_y)y''. \end{aligned} \tag{5.6.16}$$

Define the **infinitesimal generator** $X = \xi(x, y)\frac{\partial}{\partial x} + \eta(x, y)\frac{\partial}{\partial y}$, and its **prolongations** $X^{(1)} = X + \zeta_1\frac{\partial}{\partial y'}$ and $X^{(2)} = X^{(1)} + \zeta_2\frac{\partial}{\partial y''}$. For a given differential equation, these infinitesimal generators will generate an r-dimensional Lie group (L_r).

For the equation $F(x, y, y', y'') = 0$ to be invariant under the action of the above group, $X^{(2)}F\,|_{F=0} = 0$. When $F = y'' - f(x, y, y')$ this determining equation is

$$\begin{aligned} &\eta_{xx} + (2\eta_{xy} - \xi_{xx})y' + (\eta_{yy} - 2\xi_{xy})y'^2 - y'^3\xi_y y \\ &\quad + (\eta_y - 2\xi_x - 3y'\xi_y)f - \left[\eta_x + (\eta_y - \xi_x)y' - y'^2\right] f_{y'} - \xi f_x - \eta f_y = 0. \end{aligned}$$

Given two generators $X_1 = \xi_1\frac{\partial}{\partial x} + \eta_1\frac{\partial}{\partial y}$ and $X_2 = \xi_2\frac{\partial}{\partial x} + \eta_2\frac{\partial}{\partial y}$, the **pseudoscalar product** is $X_1 \vee X_2 = \xi_1\eta_2 - \xi_2\eta_1$ and the **commutator** is $[X_1, X_2] = X_1X_2 - X_2X_1$. By changing basis any Lie algebra L_2 can be changed to one of four types:

No.	Commutator	Pseudoscalar	Typified by
I	$[X_1, X_2] = 0$	$X_1 \vee X_2 \neq 0$	$\{X_1 = \frac{\partial}{\partial x},\quad X_2 = \frac{\partial}{\partial y}\}$
II	$[X_1, X_2] = 0$	$X_1 \vee X_2 = 0$	$\{X_1 = \frac{\partial}{\partial y},\quad X_2 = x\frac{\partial}{\partial y}\}$
III	$[X_1, X_2] = X_1$	$X_1 \vee X_2 \neq 0$	$\{X_1 = \frac{\partial}{\partial y},\quad X_2 = x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\}$
IV	$[X_1, X_2] = X_1$	$X_1 \vee X_2 = 0$	$\{X_1 = \frac{\partial}{\partial y},\quad X_2 = y\frac{\partial}{\partial y}\}$

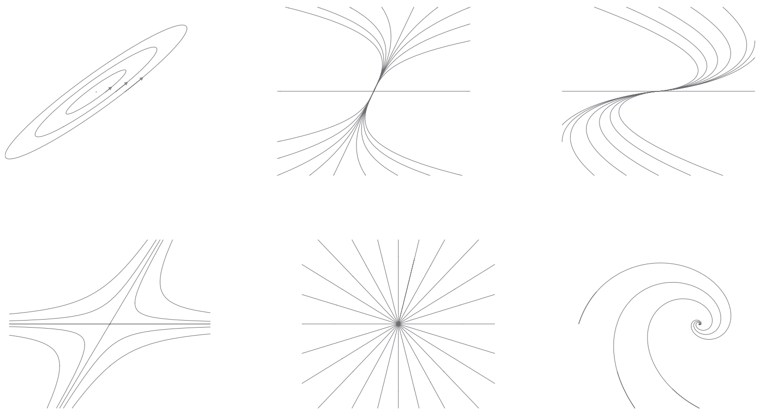
5.6.10 NAMED ORDINARY DIFFERENTIAL EQUATIONS

1. Airy equation $y'' = xy$
   Solution: $y = c_1 \operatorname{Ai}(x) + c_2 \operatorname{Bi}(x)$
2. Bernoulli equation $y' = a(x)y^n + b(x)y$
3. Bessel equation $x^2y'' + xy' + (\lambda^2x^2 - n^2)y = 0$
   Solution: $y = c_1 J_n(\lambda x) + c_2 Y_n(\lambda x)$
4. Bessel equation (transformed): $x^2y'' + (2p+1)xy' + (\lambda^2x^{2r} + \beta^2)y = 0$
   Solution: $y = x^{-p}\left[c_1 J_{q/r}\left(\frac{\lambda}{r}x^r\right) + c_2 Y_{q/r}\left(\frac{\lambda}{r}x^r\right)\right] \quad q \equiv \sqrt{p^2 - \beta^2}$
5. Bôcher equation: $y'' + \frac{1}{2}\left[\frac{m_1}{x-a_1} + \cdots + \frac{m_{n-1}}{x-a_{n-1}}\right]y'$
   $+\frac{1}{4}\left[\frac{A_0 + A_1x + \cdots + A_lx^l}{(x-a_1)^{m_1}(x-a_2)^{m_2}\cdots(x-a_{n-1})^{m_{n-1}}}\right]y = 0$
6. Duffing equation $y'' + y + \epsilon y^3 = 0$
7. Emden–Fowler equation $(x^py')' \pm x^\sigma y^n = 0$
8. Hypergeometric equation: $y'' + \left(\frac{1-\alpha-\alpha'}{x-a} + \frac{1-\beta-\beta'}{x-b} + \frac{1-\gamma-\gamma'}{x-c}\right)y'$
   $-\left(\frac{\alpha\alpha'}{(x-a)(b-c)} + \frac{\beta\beta'}{(x-b)(c-a)} + \frac{\gamma\gamma'}{(x-c)(a-b)}\right)\frac{(a-b)(b-c)(c-a)}{(x-a)(x-b)(x-c)}u = 0$
   Solution: $y = P\left\{\begin{matrix} a & b & c & \\ \alpha & \beta & \gamma & x \\ \alpha' & \beta' & \gamma' & \end{matrix}\right\}$ (Riemann's P function)
9. Legendre equation $(1-x^2)y'' - 2xy' + n(n+1)y = 0$
   Solution: $y = c_1P_n(x) + c_2Q_n(x)$
10. Mathieu equation $y'' + (a - 2q\cos 2x)y = 0$
11. Painlevé transcendents
    (a) first $y'' = 6y^2 + tx$
    (b) second $y'' = 2y^3 + xy + a$
    (c) third $y'' = \frac{1}{y}(y')^2 - \frac{1}{x}y' + \frac{1}{x}(\alpha y^2 + \beta) + \gamma y^3 + \frac{\delta}{y}$
    (d) fourth $y'' = \frac{1}{2y}(y')^2 + \frac{3y^3}{2} + 4xy^2 + 2(x^2 - \alpha)y + \frac{\beta}{y}$
    (e) fifth
    $y'' = \left(\frac{1}{2y} + \frac{1}{y-1}\right)(y')^2 - \frac{1}{x}y' + \frac{(y-1)^2}{x^2}\left(\alpha y + \frac{\beta}{y}\right) + \frac{\gamma y}{x} + \frac{\delta y(y+1)}{y-1}$
    (f) sixth $y'' = \frac{1}{2}\left(\frac{1}{y} + \frac{1}{y-1} + \frac{1}{y-x}\right)(y')^2 - \left(\frac{1}{x} + \frac{1}{x-1} + \frac{1}{y-x}\right)y' +$
    $\frac{y(y-1)(y-x)}{x^2(x-1)^2}\left[\alpha + \frac{\beta x}{y^2} + \frac{\gamma(x-1)}{(y-1)^2} + \frac{\delta x(x-1)}{(y-x)^2}\right]$
12. Parabolic cylinder equation $y'' + (ax^2 + bx + c)y = 0$
13. Riccati equation $y' = a(x)y^2 + b(x)y + c(x)$

5.6.11 TYPES OF CRITICAL POINTS

An ODE may have several types of critical points; these include improper node, deficient improper node, proper node, saddle, center, and focus. See Figure 5.1.

FIGURE 5.1
Types of critical points. Clockwise from upper left: center, improper node, deficient improper node, spiral, star, and saddle.



5.7 PARTIAL DIFFERENTIAL EQUATIONS

5.7.1 CLASSIFICATIONS OF PDES

Consider second order partial differential equations, with two independent variables, of the form

$$A(x,y)\frac{\partial^2 u}{\partial x^2} + B(x,y)\frac{\partial^2 u}{\partial x \partial y} + C(x,y)\frac{\partial^2 u}{\partial y^2} = \Psi\left(u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, x, y\right). \tag{5.7.1}$$

If $\begin{bmatrix} B^2 - 4AC > 0 \\ B^2 - 4AC = 0 \\ B^2 - 4AC < 0 \end{bmatrix}$ at some point (x,y), then Equation (5.7.1) is $\begin{bmatrix} \text{hyperbolic} \\ \text{parabolic} \\ \text{elliptic} \end{bmatrix}$ at that point. If an equation is of the same type at all points, then the equation is simply of that type.

5.7.2 WELL-POSEDNESS OF PDES

Partial differential equations involving $u(\mathbf{x})$ usually have the following types of boundary conditions:

1. Dirichlet conditions: $u = 0$ on the boundary
2. Neumann conditions: $\frac{\partial u}{\partial n} = 0$ on the boundary
3. Cauchy conditions: u and $\frac{\partial u}{\partial n}$ specified on the boundary

A well-posed differential equation meets these conditions:

1. The solution exists.
2. The solution is unique.
3. The solution is stable (i.e., the solution depends continuously on the boundary conditions and initial conditions).

Type of boundary conditions	**Type of equation** Elliptic	Hyperbolic	Parabolic
Dirichlet			
Open surface	Undetermined	Undetermined	Unique, stable solution in one direction
Closed surface	Unique, stable solution	Undetermined	Undetermined
Neumann			
Open surface	Undetermined	Undetermined	Unique, stable solution in one direction
Closed surface	Overdetermined	Overdetermined	Overdetermined
Cauchy			
Open surface	Not physical	Unique, stable solution	Overdetermined
Closed surface	Overdetermined	Overdetermined	Overdetermined

5.7.3 TRANSFORMING PARTIAL DIFFERENTIAL EQUATIONS

To transform a partial differential equation, construct a new function which depends upon new variables, and then differentiate with respect to the old variables to see how the derivatives transform.

EXAMPLE Consider transforming

$$f_{xx} + f_{yy} + x f_y = 0, \tag{5.7.2}$$

from the $\{x, y\}$ variables to the $\{u, v\}$ variables, where $\{u = x, v = x/y\}$. Note that the inverse transformation is given by $\{x = u, y = u/v\}$.

First, define $g(u, v)$ to be the function $f(x, y)$ when written in the new variables, that is

$$f(x, y) = g(u, v) = g\left(x, \frac{x}{y}\right). \tag{5.7.3}$$

Now create the needed derivative terms, carefully applying the chain rule. For example, differentiating Equation (5.7.3) with respect to x results in

$$\begin{aligned} f_x(x, y) &= g_u \frac{\partial}{\partial x}(u) + g_v \frac{\partial}{\partial x}(v) = g_1 \frac{\partial}{\partial x}(x) + g_2 \frac{\partial}{\partial x}\left(\frac{x}{y}\right) \\ &= g_1 + g_2 \frac{1}{y} = g_1 + \frac{v}{u} g_2, \end{aligned}$$

where a subscript of "1" ("2") indicates a derivative with respect to the first (second) argument of the function $g(u, v)$, that is, $g_1(u, v) = g_u(u, v)$. Use of this "slot notation" tends to minimize errors. In like manner

$$\begin{aligned} f_y(x, y) &= g_u \frac{\partial}{\partial y}(u) + g_v \frac{\partial}{\partial y}(v) = g_1 \frac{\partial}{\partial y}(x) + g_2 \frac{\partial}{\partial y}\left(\frac{x}{y}\right) \\ &= -\frac{x}{y^2} g_2 = -\frac{v^2}{u} g_2. \end{aligned}$$

The second order derivatives can be calculated similarly:

$$\begin{aligned} f_{xx}(x, y) &= \frac{\partial}{\partial x}\left(f_x(x, y)\right) = \frac{\partial}{\partial x}\left(g_1 + \frac{1}{y} g_2\right) \\ &= g_{11} + \frac{2v}{u} g_{12} + \frac{v^2}{u^2} g_{22}, \\ f_{xy}(x, y) &= \frac{\partial}{\partial x}\left(-\frac{x}{y^2} g_2\right) = -\frac{u^2}{v^2} g_2 - \frac{u^3}{v^3} g_{12} - \frac{u^2}{v^2} g_{22}, \\ f_{yy}(x, y) &= \frac{\partial}{\partial y}\left(-\frac{x}{y^2} g_2\right) = \frac{2v^3}{u^2} g_2 + \frac{v^4}{u^2} g_{22}. \end{aligned}$$

Finally, Equation (5.7.2) in the new variables has the form,

$$\begin{aligned} 0 &= f_{xx} + f_{yy} + x f_y \\ &= \left(g_{11} + \frac{2v}{u} g_{12} + \frac{v^2}{u^2} g_{22}\right) + \left(\frac{2v^3}{u^2} g_2 + \frac{v^4}{u^2} g_{22}\right) + (u)\left(-\frac{v^2}{u} g_2\right) \\ &= \frac{v^2(2v - u^2)}{u^2} g_v + g_{uu} + \frac{2v}{u} g_{uv} + \frac{v^2(1 + v^2)}{u^2} g_{vv}. \end{aligned}$$

5.7.4 GREEN'S FUNCTIONS

The Green's function, $G(\mathbf{r}; \mathbf{r}_0)$, of a linear differential operator $L[\cdot]$ is a solution of $L[G(\mathbf{r}; \mathbf{r}_0)] = \delta(\mathbf{r} - \mathbf{r}_0)$ where $\delta(\cdot)$ is the Dirac delta function.

In the following, $\mathbf{r} = (x, y, z)$, $\mathbf{r}_0 = (x_0, y_0, z_0)$, $R^2 = |\mathbf{r} - \mathbf{r}_0|^2 = (x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2$, and $P^2 = (x - x_0)^2 + (y - y_0)^2$.

1. For the potential equation $\nabla^2 G + k^2 G = -4\pi\delta(\mathbf{r} - \mathbf{r}_0)$, with the radiation condition (outgoing waves only), the solution is

$$G = \begin{cases} \dfrac{2\pi i}{k} e^{ik|x-x_0|} & \text{in one dimension,} \\ i\pi H_0^{(1)}(kP) & \text{in two dimensions, and} \\ \dfrac{e^{ikR}}{R} & \text{in three dimensions,} \end{cases} \tag{5.7.4}$$

where $H_0^{(1)}(\cdot)$ is a Hankel function (see page 460).

2. For the n-dimensional diffusion equation

$$\nabla^2 G - a^2 \frac{\partial G}{\partial t} = -4\pi\delta(\mathbf{r} - \mathbf{r}_0)\delta(t - t_0), \tag{5.7.5}$$

with the initial condition $G = 0$ for $t < t_0$, and the boundary condition $G = 0$ at $r = \infty$, the solution is

$$G = \frac{4\pi}{a^2} \left(\frac{a}{2\sqrt{\pi(t - t_0)}} \right)^N \exp\left(-\frac{a^2 |\mathbf{r} - \mathbf{r}_0|^2}{4(t - t_0)} \right). \tag{5.7.6}$$

3. For the wave equation

$$\nabla^2 G - \frac{1}{c^2} \frac{\partial^2 G}{\partial t^2} = -4\pi\delta(\mathbf{r} - \mathbf{r}_0)\delta(t - t_0), \tag{5.7.7}$$

with the initial conditions $G = G_t = 0$ for $t < t_0$, and the boundary condition $G = 0$ at $r = \infty$, the solution is

$$G = \begin{cases} 2c\pi H\left[(t - t_0) - \dfrac{|x - x_0|}{c}\right] & \text{in one space dimension,} \\ \dfrac{2c}{\sqrt{c^2(t - t_0)^2 - P^2}} H\left[(t - t_0) - \dfrac{P}{c}\right] & \text{in two space dimensions, and} \\ \dfrac{1}{R}\delta\left[\dfrac{R}{c} - (t - t_0)\right] & \text{in three space dimensions.} \end{cases} \tag{5.7.8}$$

where $H(\cdot)$ is the Heaviside function.

5.7.5 SEPARATION OF VARIABLES

A solution of a linear PDE in n dimensions is attempted in the form $u(\mathbf{x}) = u(x_1, x_2, \cdots, x_n) = X_1(x_1)X_2(x_2)\ldots X_n(x_n)$. Logic may determine the $\{X_i\}$.

1. The diffusion or heat equation in a circle is

$$\frac{\partial u}{\partial t} = \nabla^2 u = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial u}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 u}{\partial \theta^2} \tag{5.7.9}$$

for $u(t, r, \theta)$, where (r, θ) are the polar coordinates. If $u(t, r, \theta) = T(t)R(r)\Theta(\theta)$, then

$$\frac{1}{rR}\frac{d}{dr}\left(r\frac{dR}{dr}\right) + \frac{1}{r^2\Theta}\frac{d^2\Theta}{d\theta^2} - \frac{1}{T}\frac{dT}{dt} = 0. \tag{5.7.10}$$

Logic about which terms depend on which variables leads to

$$\frac{1}{T}\frac{dT}{dt} = -\lambda, \qquad \frac{1}{\Theta}\frac{d^2\Theta}{d\theta^2} = -\rho, \qquad r\frac{d}{dr}\left(r\frac{dR}{dr}\right) + (-\rho + r^2\lambda)R = 0.$$

where λ and ρ are unknown constants. Solving these ordinary differential equations yields the general solution,

$$\begin{aligned} u(t, r, \theta) = \int_{-\infty}^{\infty} d\lambda \int_{-\infty}^{\infty} d\rho\, e^{-\lambda t}\Big[B(\lambda, \rho)\sin(\sqrt{\rho}\theta) + C(\lambda, \rho)\cos(\sqrt{\rho}\theta)\Big] \\ \times \Big[D(\lambda, \rho)J_{\sqrt{\rho}}(\sqrt{\lambda}r) + E(\lambda, \rho)Y_{\sqrt{\rho}}(\sqrt{\lambda}r)\Big]. \end{aligned} \tag{5.7.11}$$

Boundary conditions are required to determine the $\{B, C, D, E\}$.

2. A necessary and sufficient condition for a system with Hamiltonian $H = \frac{1}{2}(p_x^2 + p_y^2) + V(x, y)$, to be separable in elliptic, polar, parabolic, or Cartesian coordinates is that the expression,

$$\begin{aligned} (V_{yy} - V_{xx})(-2axy - b'y - bx + d) + 2V_{xy}(ay^2 - ax^2 + by - b'x + c - c') \\ + V_x(6ay + 3b) + V_y(-6ax - 3b'), \end{aligned} \tag{5.7.12}$$

vanishes for some constants $(a, b, b', c, c', d) \neq (0, 0, 0, c, c, 0)$.

3. Consider the orthogonal coordinate system $\{u^1, u^2, u^3\}$, with the metric $\{g_{ii}\}$, and $g = g_{11}g_{22}g_{33}$. The Stäckel matrix is defined as

$$S = \begin{bmatrix} \Phi_{11}(u^1) & \Phi_{12}(u^1) & \Phi_{13}(u^1) \\ \Phi_{21}(u^2) & \Phi_{22}(u^2) & \Phi_{23}(u^2) \\ \Phi_{31}(u^3) & \Phi_{32}(u^3) & \Phi_{33}(u^3) \end{bmatrix}, \tag{5.7.13}$$

where the $\{\Phi_{ij}\}$, which are analogous to the $\{X_i\}$ above, have been tabulated for many different coordinate systems. The determinant of S can be written as $s = \Phi_{11}M_{11} + \Phi_{21}M_{21} + \Phi_{31}M_{33}$, where

$$M_{11} = \begin{vmatrix} \Phi_{22} & \Phi_{23} \\ \Phi_{32} & \Phi_{33} \end{vmatrix}, \quad M_{21} = -\begin{vmatrix} \Phi_{12} & \Phi_{13} \\ \Phi_{32} & \Phi_{33} \end{vmatrix} \quad M_{31} = \begin{vmatrix} \Phi_{12} & \Phi_{13} \\ \Phi_{22} & \Phi_{23} \end{vmatrix}. \tag{5.7.14}$$

If the separability conditions, $g_{ii} = s/M_{i1}$ and $\sqrt{g}/s = f_1(u^1)f_2(u^2)f_3(u^3)$, are met then the Helmholtz equation $\nabla^2 W + \lambda^2 W = 0$ separates with $W = X_1(u^1)X_2(u^2)X_3(u^3)$. Here the $\{X_i\}$ are defined by

$$\frac{1}{f_i}\frac{d}{du^i}\left(f_i\frac{dX_i}{du^i}\right) + X_i\sum_{j=1}^{3}\alpha_j\Phi_{ij} = 0, \tag{5.7.15}$$

with $\alpha_1 = \lambda^2$, and α_2 and α_3 arbitrary.

(a) Necessary and sufficient conditions for the separation of the Laplace equation ($\nabla^2 W = 0$) are

$$\frac{g_{ii}}{g_{jj}} = \frac{M_{j1}}{M_{i1}} \qquad \text{and} \qquad \frac{\sqrt{g}}{g_{ii}} = f_1(u^1)f_2(u^2)f_3(u^3)M_{i1}. \tag{5.7.16}$$

(b) Necessary and sufficient conditions for the separation of the scalar Helmholtz equation are

$$g_{ii} = \frac{S}{M_{i1}} \qquad \text{and} \qquad \frac{\sqrt{g}}{S} = f_1(u^1)f_2(u^2)f_3(u^3) \tag{5.7.17}$$

4. In parabolic coordinates $\{\mu, \nu, \psi\}$ the metric coefficients are $g_{11} = g_{22} = \mu^2 + \nu^2$ and $g_{33} = \mu^2\nu^2$. Hence, $\sqrt{g} = \mu\nu(\mu^2 + \nu^2)$. For the Stäckel matrix

$$S = \begin{bmatrix} \mu^2 & -1 & -1/\mu^2 \\ \nu^2 & 1 & -1/\nu^2 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.7.18}$$

(for which $s = \mu^2 + \nu^2$, $M_{11} = M_{21} = 1$, and $M_{31} = \mu^{-2} + \nu^{-2}$), the separability condition holds with $f_1 = \mu$, $f_2 = \nu$, and $f_3 = 1$. Hence, the Helmholtz equation separates in parabolic coordinates. The separated equations are

$$\begin{aligned} &\frac{1}{\mu}\frac{d}{d\mu}\left(\mu\frac{dX_1}{d\mu}\right) + X_1\left(\alpha_1\mu^2 - \alpha_2 - \frac{\alpha_3}{\mu^2}\right) = 0, \\ &\frac{1}{\nu}\frac{d}{d\nu}\left(\nu\frac{dX_2}{d\nu}\right) + X_2\left(\alpha_1\nu^2 + \alpha_2 - \frac{\alpha_3}{\nu^2}\right) = 0, \text{and} \\ &\frac{d^2X_3}{d\psi^2} + \alpha_3X_3 = 0, \end{aligned} \tag{5.7.19}$$

where $W = X_1(\mu)X_2(\nu)X_3(\psi)$.

5.7.6 SEPARATION OF VARIABLES FOR THE KLEIN–GORDON EQUATION IN 2D

The Klein–Gordon (KG) equation for w is given by $\Box w + \mu^2 w = 0$, in coordinates u^i on Minkowski space where $\mu = mc/\hbar$ is a constant, $\Box$ denotes the d'Alembertian operator

$$\Box w = \frac{1}{\sqrt{-g}}\partial_i \left(\sqrt{-g}\, g^{ij}\, \partial_j w\right),$$

g is the determinant of the matrix representing the metric tensor g_{ij} in coordinates u^i, and g^{ij} are the components of its inverse.

Now set $u^1 = u$, and $u^2 = v$. Listed below are the orthogonal coordinate systems for which the 2-dimensional KG equation admits product separable solutions of the form $w(u, v) = X(u)Y(v)$. For coordinate systems 1–10 below, except for number 2, the metric tensor is in Liouville form: $g_{11} = -(A(u) + B(v)) = -g_{22}$, and $g_{12} = 0$.

The Klein–Gordon equation is these cases can be written as

$$\left(-\frac{\partial^2 w}{\partial u^2} + \frac{\partial^2 w}{\partial v^2}\right) + \mu^2(A + B)w = 0.$$

Substituting $w = XY$ in the KG equation results in

$$YX''(u) - XY''(v) - \mu^2(A + B)XY = 0,$$

which, upon division by XY, is clearly separable. The KG equation is also clearly separable for coordinate system number 2 which is not, however, in Liouville form.

1. Cartesian coordinates
   Relations: $t = u, \quad x = v$
   Ranges: $-\infty < u < \infty, \quad -\infty < v < \infty$
   $g_{uu} = -g_{vv} = -1, \quad \sqrt{-g} = 1$

2. Rindler coordinates
   (a) for $-t^2 + x^2 < 0$:
      Relations: $t = u \cosh v, \quad x = u \sinh v$
      Ranges: $0 < u < \infty, \quad -\infty < v < \infty$
      $g_{uu} = -1, \quad g_{vv} = u^2, \quad \sqrt{-g} = u$
   (b) for $-t^2 + x^2 > 0$:
      Relations: $t = u \sinh v, \quad x = u \cosh v$
      Ranges: $0 < u < \infty, \quad -\infty < v < \infty$
      $g_{uu} = 1, \quad g_{vv} = -u^2, \quad \sqrt{-g} = u$

3. Real elliptic coordinates of type I
   Relations: $t = a \cosh u \sinh v, \quad x = a \sinh u \cosh v, \quad 0 < a$

Ranges: $0 < u < \infty, \quad 0 < v < \infty$
$g_{uu} = -g_{vv} = a^2(\cosh^2 u + \sinh^2 v), \quad \sqrt{-g} = a^2(\cosh^2 u + \sinh^2 v)$

4. Real elliptic coordinates of type II
   (a) for $|t| - |x| > a$:
      Relations: $t = a\cosh u \cosh v, \quad x = a \sinh v \sinh u, \quad 0 < a$
      Ranges: $0 < u < v < \infty$
      $g_{uu} = -g_{vv} = a^2(\cosh^2 v - \cosh^2 u),$
      $\sqrt{-g} = a^2(\cosh^2 v - \cosh^2 u)$
   (b) for $|t| - |x| < -a$:
      Relations: $t = a\sinh u \sinh v, \quad x = a \cosh v \cosh u, \quad 0 < a$
      Ranges: $0 < v < u < \infty$
      $g_{uu} = -g_{vv} = a^2(\cosh^2 u - \cosh^2 v),$
      $\sqrt{-g} = a^2(\cosh^2 u - \cosh^2 v)$
   (c) for $|t| + |x| < a$:
      Relations: $t = a\cos u \cos v, \quad x = a \sin v \sin u, \quad 0 < a$
      Ranges: $0 < v < u < \frac{\pi}{2}$
      $g_{uu} = -g_{vv} = a^2(\cos^2 u - \cos^2 v), \quad \sqrt{-g} = a^2(\cos^2 u - \cos^2 v)$

5. Complex elliptic coordinates
   Relations: $t - x = a\cosh(u+v), \quad t + x = a\sinh(u-v), \quad 0 < a$
   Ranges: $0 < |v| < u < \infty$
   $g_{uu} = -g_{vv} = a^2(\sinh 2u + \sinh 2v), \quad \sqrt{-g} = a^2(\sinh 2u + \sinh 2v)$

6. Null elliptic coordinates of type I
   Relations: $t - x = e^{u+v}, \quad t + x = 2\sinh(u-v)$
   Ranges: $-\infty < u < \infty, \quad -\infty < v < \infty$
   $g_{uu} = -g_{vv} = e^{2u} + e^{2v}, \quad \sqrt{-g} = e^{2u} + e^{2v}$

7. Null elliptic coordinates of type II
   (a) for $-t^2 + x^2 > |t - x|$:
      Relations: $t - x = -e^{u+v}, \quad t + x = 2\cosh(u-v)$
      Ranges: $-\infty < v < u < \infty$
      $g_{uu} = -g_{vv} = e^{2u} - e^{2v}, \quad \sqrt{-g} = e^{2u} - e^{2v}$
   (b) for $-t^2 + x^2 < -|t - x|$:
      Relations: $t - x = e^{u+v}, \quad t + x = 2\cosh(u-v)$
      Ranges: $-\infty < u < v < \infty$
      $g_{uu} = -g_{vv} = e^{2v} - e^{2u}, \quad \sqrt{-g} = e^{2v} - e^{2u}$

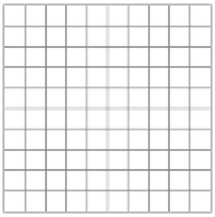
8. Timelike parabolic coordinates
   Relations: $t = \frac{1}{2}(u^2 + v^2), \quad x = uv$
   Ranges: $0 < v < u < \infty$
   $g_{uu} = -g_{vv} = -(u^2 - v^2), \quad \sqrt{-g} = u^2 - v^2$

9. Spacelike parabolic coordinates
   Relations: $t = uv, \quad x = \frac{1}{2}(u^2 + v^2)$
   Ranges: $0 < v < u < \infty$
   $g_{uu} = -g_{vv} = u^2 - v^2, \quad \sqrt{-g} = u^2 - v^2$

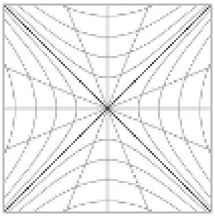
10. Null parabolic coordinates
    Relations: $t + x = -(u + v), \quad t - x = \frac{1}{2}(u - v)^2$
    Ranges: $0 < |v| < u < \infty$
    $g_{uu} = -g_{vv} = -(u - v), \quad \sqrt{-g} = u - v$

The following graphics show the above coordinate systems.[1] The empty white spaces are open regions which are not covered by the coordinate system.

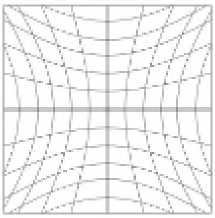
1. Cartesian



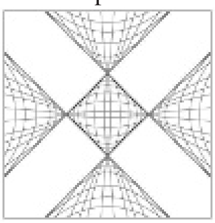
2. Rindler



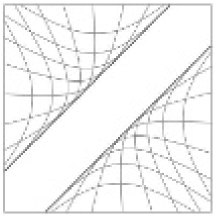
3. Real Elliptic I



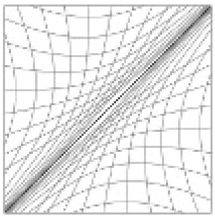
4. Real Elliptic II



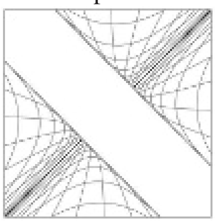
5. Complex Elliptic



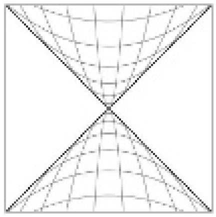
6. Null Elliptic I



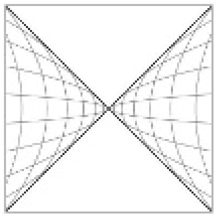
7. Null Elliptic II



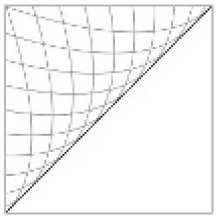
8. Timelike Parabolic



9. Spacelike Parabolic



10. Null Parabolic



[1] Reprinted with permission from: K. Rajaratnam, R. G. McLenaghan, and C. Valero, *Orthogonal Separation of the Hamilton-Jacobi Equation on Spaces of Constant Curvature*, SIGMA 12 (2016), 117, 30 pages, https://arxiv.org/abs/1607.00712.

5.7.7 SOLUTIONS TO THE WAVE EQUATION

1. Consider the wave equation $\frac{\partial^2 u}{\partial t^2} = \nabla^2 u = \frac{\partial^2 u}{\partial {x_1}^2} + \cdots + \frac{\partial^2 u}{\partial {x_n}^2}$, with $\mathbf{x} = (x_1, \ldots, x_n)$ and the initial data $u(0, \mathbf{x}) = f(\mathbf{x})$ and $u_t(0, \mathbf{x}) = g(\mathbf{x})$. When n is odd (and $n \geq 3$), the solution is

$$u(t, \mathbf{x}) = \frac{1}{1 \cdot 3 \cdots (n-2)} \left\{ \frac{\partial}{\partial t} \left(\frac{\partial}{t\, \partial t} \right)^{(n-3)/2} t^{n-2} \omega[f; \mathbf{x}, t] + \left(\frac{\partial}{t\, \partial t} \right)^{(n-3)/2} t^{n-2} \omega[g; \mathbf{x}, t] \right\}, \tag{5.7.20}$$

where $\omega[h; \mathbf{x}, t]$ is the average of the function $h(\mathbf{x})$ over the surface of an n-dimensional sphere of radius t centered at $\mathbf{x}$; that is, $\omega[h; \mathbf{x}, t] = \frac{1}{\sigma_{n-1}(t)} \int h(\zeta)\, d\Omega$, where $|\zeta - \mathbf{x}|^2 = t^2$, $\sigma_{n-1}(t)$ is the surface area of the n-dimensional sphere of radius t, and $d\Omega$ is an element of area.

When n is even, the solution is given by

$$u(t, \mathbf{x}) = \frac{1}{2 \cdot 4 \cdots (n-2)} \left\{ \frac{\partial}{\partial t} \left(\frac{\partial}{t\, \partial t} \right)^{(n-2)/2} \int_0^t \omega[f; \mathbf{x}, \rho] \frac{\rho^{n-1}\, d\rho}{\sqrt{t^2 - \rho^2}} + \left(\frac{\partial}{t\, \partial t} \right)^{(n-2)/2} \int_0^t \omega[g; \mathbf{x}, \rho] \frac{\rho^{n-1}\, d\rho}{\sqrt{t^2 - \rho^2}} \right\}, \tag{5.7.21}$$

where $\omega[h; \mathbf{x}, t]$ is defined as above. Since this expression is integrated over ρ, the values of f and g must be known everywhere in the *interior* of the n-dimensional sphere.

Using u_n for the solution in n dimensions, the above simplify to

$$u_1(x, t) = \frac{1}{2} \left[f(x - t) + f(x + t) \right] + \frac{1}{2} \int_{x-t}^{x+t} g(\zeta)\, d\zeta, \tag{5.7.22}$$

$$u_2(\mathbf{x}, t) = \frac{1}{2\pi} \frac{\partial}{\partial t} \iint_{R(t)} \frac{f(x_1 + \zeta_1, x_2 + \zeta_2)}{\sqrt{t^2 - \zeta_1^2 - \zeta_2^2}}\, d\zeta_1\, d\zeta_2 + \frac{1}{2\pi} \iint_{R(t)} \frac{g(x_1 + \zeta_1, x_2 + \zeta_2)}{\sqrt{t^2 - \zeta_1^2 - \zeta_2^2}}\, d\zeta_1\, d\zeta_2, \quad \text{and} \tag{5.7.23}$$

$$u_3(\mathbf{x}, t) = \frac{\partial}{\partial t} \Big(t\omega[f; \mathbf{x}, t] \Big) + t\omega[g; \mathbf{x}, t], \tag{5.7.24}$$

where $R(t)$ is the region $\{(\zeta_1, \zeta_2) \mid \zeta_1^2 + \zeta_2^2 \leq t^2\}$ and

$$\omega[h; \mathbf{x}, t] = \frac{1}{4\pi} \int_0^{2\pi} \int_0^{\pi} h(x_1 + t \sin\theta \cos\phi, x_2 + t \sin\theta \sin\phi, x_3 + t\cos\theta) \times \sin\theta\, d\theta\, d\phi. \tag{5.7.25}$$

2. The solution of the one-dimensional wave equation

$$\begin{aligned} v_{tt} &= c^2 v_{xx} \\ v(0,t) &= 0, && \text{for } 0 < t < \infty, \\ v(x,0) &= f(x), && \text{for } 0 \le x < \infty, \\ v_t(x,0) &= g(x), && \text{for } 0 \le x < \infty, \end{aligned} \tag{5.7.26}$$

is

$$v(x,t) = \begin{cases} \frac{1}{2}\left[f(x+ct)+f(x-ct)\right] + \frac{1}{2c}\int_{x-ct}^{x+ct} g(\zeta)\,d\zeta, & \text{for } x \ge ct, \\ \frac{1}{2}\left[f(x+ct)-f(ct-x)\right] + \frac{1}{2c}\int_{ct-x}^{x+ct} g(\zeta)\,d\zeta, & \text{for } x < ct. \end{cases} \tag{5.7.27}$$

3. The solution of the inhomogeneous wave equation

$$\frac{\partial^2 u}{\partial t^2} - \frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} - \frac{\partial^2 u}{\partial z^2} = F(t,x,y,z), \tag{5.7.28}$$

with the initial conditions $u(0,x,y,z) = 0$ and $u_t(0,x,y,z) = 0$, is

$$u(t,x,y,z) = \frac{1}{4\pi} \iiint\limits_{\rho \le t} \frac{F(t-\rho,\zeta,\eta,\xi)}{\rho}\,d\zeta\,d\eta\,d\xi, \tag{5.7.29}$$

with $\rho = \sqrt{(x-\zeta)^2 + (y-\eta)^2 + (z-\xi)^2}$.

5.7.8 SOLUTIONS OF LAPLACE'S EQUATION

1. If $\nabla^2 u = 0$ in a circle of radius R and $u(R,\theta) = f(\theta)$, for $0 \le \theta < 2\pi$, then $u(r,\theta)$ is

$$u(r,\theta) = \frac{1}{2\pi} \int_0^{2\pi} \frac{R^2 - r^2}{R^2 - 2Rr\cos(\theta-\phi) + r^2}\, f(\phi)\,d\phi.$$

2. If $\nabla^2 u = 0$ in a sphere of radius one and $u(1,\theta,\phi) = f(\theta,\phi)$, then

$$u(r,\theta,\phi) = \frac{1}{4\pi} \int_0^{\pi} \int_0^{2\pi} f(\Theta,\Phi) \frac{1-r^2}{(1-2r\cos\gamma + r^2)^{3/2}} \sin\Theta\,d\Theta\,d\Phi,$$

where $\cos\gamma = \cos\theta\cos\Theta + \sin\theta\sin\Theta\cos(\phi-\Phi)$.

3. If $\nabla^2 u = 0$ in the half plane $y \ge 0$, and $u(x,0) = f(x)$, then

$$u(x,y) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(t)y}{(x-t)^2 + y^2}\,dt.$$

4. If $\nabla^2 u = 0$ in the half space $z \ge 0$, and $u(x,y,0) = f(x,y)$, then

$$u(x,y,z) = \frac{z}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{f(\zeta,\eta)}{\left[(x-\zeta)^2 + (y-\eta)^2 + z^2\right]^{3/2}}\,d\zeta\,d\eta.$$

5.7.9 QUASI-LINEAR EQUATIONS

Consider the first-order quasi-linear differential equation for $u(\mathbf{x}) = u(x_1, \ldots, x_N)$,

$$a_1(\mathbf{x}, u)u_{x_1} + a_2(\mathbf{x}, u)u_{x_2} + \cdots + a_N(\mathbf{x}, u)u_{x_N} = b(\mathbf{x}, u). \tag{5.7.30}$$

Defining $\frac{\partial x_k}{\partial s} = a_k(\mathbf{x}, u)$, for $k = 1, 2, \ldots, N$, this equation becomes $\frac{du}{ds} = b(\mathbf{x}, u)$. The initial conditions can often be parameterized as (with $\mathbf{t} = (t_1, \ldots, t_{N-1})$)

$$\begin{aligned} u(s = 0, \mathbf{t}) &= v(\mathbf{t}), \\ x_1(s = 0, \mathbf{t}) &= h_1(\mathbf{t}), \\ x_2(s = 0, \mathbf{t}) &= h_2(\mathbf{t}), \\ &\vdots \\ x_N(s = 0, \mathbf{t}) &= h_N(\mathbf{t}), \end{aligned} \tag{5.7.31}$$

To solve the original equation, the differential equations for $u(s, \mathbf{t})$ and the $\{x_k(s, \mathbf{t})\}$ must be solved; these are the *characteristic equations*. This often results in an implicit solution.

EXAMPLES

- Consider the wave equation $cu_x + u_y = 0$, where c is a constant, with $x = x_0$ and $u = f(x_0)$ when $y = 0$. The characteristic equations are

  $$\frac{\partial x}{\partial s} = c, \qquad \frac{\partial y}{\partial s} = 1, \qquad \frac{du}{ds} = 0.$$

  with the conditions

  $$y(s = 0) = 0, \qquad x(s = 0) = x_0, \qquad u(s = 0) = f(x_0).$$

  The solutions of these equations are: $y = s, \quad x = x_0 + cs, \quad u = f(x_0)$.
  These can be combined to obtain $u = f(x_0) = f(x - cs) = f(x - cy)$,
  which represents a traveling wave.

- Consider the equation $u_x + x^2 u_y = -yu$ with $u = f(y)$ when $x = 0$. The characteristic equations are

  $$\frac{\partial x}{\partial s} = 1, \qquad \frac{\partial y}{\partial s} = x^2, \qquad \frac{du}{ds} = -yu.$$

  The original initial data can be written parametrically as

  $$x(s = 0, t_1) = 0, \qquad y(s = 0, t_1) = t_1, \qquad u(s = 0, t_1) = f(t_1).$$

  Solving for x results in $x(s, t_1) = s$. The equation for y is then integrated to obtain $y(s, t_1) = \frac{s^3}{3} + t_1$. Then the equation for u is integrated to obtain $u(s, t_1) = f(t_1)\exp\left(-\frac{s^4}{12} - st_1\right)$. These solutions constitute an implicit solution of the original system.
  In this case, it is possible to eliminate the s and t_1 variables analytically to obtain the explicit solution: $u(x, y) = f\left(y - \dfrac{x^3}{3}\right)\exp\left(\dfrac{x^4}{4} - xy\right)$.

5.7.10 PARTICULAR SOLUTIONS TO TWO PDES

In these tables, $P(x)$ is a polynomial of degree n.

	If $R(x)$ is	A solution to $z_x + mz_y = R(x,y)$ is
(1)	e^{ax+by}	$e^{ax+by}/(a+mb)$.
(2)	$f(ax+by)$	$\int^{ax+by} f(u)\,du/(a+mb)$.
(3)	$f(y-mx)$	$xf(y-mx)$.
(4)	$\phi(x,y)f(y-mx)$	evaluate $f(y-mx)\int \phi(x, a+mx)\,dx$; then substitute $a = y - mx$.

	If $R(x)$ is	A solution to $z_x + mz_y - kz = R(x,y)$ is
(5)	e^{ax+by}	$e^{ax+by}/(a+mb-k)$.
(6)	$\sin(ax+by)$	$-\frac{(a+bm)\cos(ax+by)+k\sin(ax+by)}{(a+bm)^2+k^2}$.
(7)	$e^{\alpha x+\beta y}\sin(ax+by)$	Replace k by $k-\alpha-m\beta$ in formula (6) and multiply by $e^{\alpha x+\beta y}$.
(8)	$e^{xk}f(ax+by)$	$e^{kx}\int f(y)\,du/(a+mb), \quad u = ax+by$.
(9)	$f(y-mx)$	$-f(y-mx)/k$.
(10)	$P(x)f(y-mx)$	$-\frac{1}{k}f(y-mx)\left[P(x)+\frac{P'(x)}{k}+\frac{P''(x)}{k^2}+\cdots+\frac{P^{(n)}(x)}{k^n}\right]$.
(11)	$e^{kx}f(y-mx)$	$xe^{kx}f(y-mx)$.

5.7.11 NAMED PARTIAL DIFFERENTIAL EQUATIONS

1. Biharmonic equation: $\nabla^4 u = 0$
2. Burgers' equation: $u_t + uu_x = \nu u_{xx}$
3. Diffusion (or heat) equation: $\nabla(c(\mathbf{x},t)\nabla u) = u_t$
4. Hamilton–Jacobi equation: $V_t + H(t, \mathbf{x}, V_{x_1}, \ldots, V_{x_n}) = 0$
5. Helmholtz equation: $\nabla^2 u + k^2 u = 0$
6. Korteweg de Vries equation: $u_t + u_{xxx} - 6uu_x = 0$
7. Laplace's equation: $\nabla^2 u = 0$
8. Navier–Stokes equations: $\mathbf{u}_t + (\mathbf{u}\cdot\nabla)\,\mathbf{u} = -\frac{\nabla P}{\rho} + \nu\nabla^2\mathbf{u}$
9. Poisson equation: $\nabla^2 u = -4\pi\rho(\mathbf{x})$
10. Schrödinger equation: $-\frac{\hbar^2}{2m}\nabla^2 u + V(\mathbf{x})u = i\hbar u_t$
11. Sine–Gordon equation: $u_{xx} - u_{yy} \pm \sin u = 0$
12. Telegraph equation: $u_{xx} = au_{tt} + bu_t + cu$
13. Tricomi equation: $u_{yy} = yu_{xx}$
14. Wave equation: $c^2\nabla^2 u = u_{tt}$

5.8 INTEGRAL EQUATIONS

$$h(x)u(x) = f(x) + \lambda \int_a^{b(x)} k(x,t)G[u(t);t]\,dt.$$

5.8.1 DEFINITIONS

1. Terms in an integral equation

• $k(x,t)$	kernel
• $u(x)$	function to be determined
• $h(x)$, $f(x)$, $b(x)$, $G[z,t]$	given functions
• λ	eigenvalue

2. Classification of integral equations

(a)	First kind	$h(x) = 0$
(b)	Second kind	$h(x) = 1$
(c)	Third kind	$h(x) \neq$ constant
(d)	Fredholm	$b(x) = b$
(e)	Homogeneous	$f(x) = 0$
(f)	Linear	$G[u(x);x] = u(x)$
(g)	Singular	$a = -\infty,\ b(x) = \infty$
(h)	Volterra	$b(x) = x$

3. Classification of kernels

(a)	Symmetric	$k(x,t) = k(t,x)$
(b)	Hermitian	$k(x,t) = \overline{k(t,x)}$
(c)	Separable/degenerate	$k(x,t) = \sum_{i=1}^{n} a_i(x)b_i(t),\ n < \infty$
(d)	Difference	$k(x,t) = k(x-t)$
(e)	Cauchy	$k(x,t) = \frac{1}{x-t}$
(f)	Singular	$k(x,t) \to \infty$ as $t \to x$
(g)	Hilbert–Schmidt	$\int_a^b \int_a^b \lvert k(x,t)\rvert^2 \,dx\,dt < \infty$

5.8.2 CONNECTION TO DIFFERENTIAL EQUATIONS

The initial value problem

$$\begin{aligned} &u''(x) + A(x)u'(x) + B(x)u(x) = g(x), \quad x > a, \\ &u(a) = c_1, \quad u'(a) = c_2, \end{aligned} \tag{5.8.1}$$

is equivalent to the Volterra integral equation,

$$\begin{aligned} u(x) &= f(x) + \int_a^x k(x,t)u(t)\,dt, \quad x \geq a, \\ f(x) &= \int_a^x (x-t)g(t)\,dt + (x-a)[A(a)c_1 + c_2] + c_1, \\ k(x,t) &= (t-x)[B(t) - A'(t)] - A(t). \end{aligned} \tag{5.8.2}$$

The boundary value problem

$$\begin{aligned} &u''(x) + A(x)u'(x) + B(x)u(x) = g(x), \quad a < x < b, \\ &u(a) = c_1, \quad u(b) = c_2, \end{aligned} \tag{5.8.3}$$

is equivalent to the Fredholm integral equation

$$\begin{aligned} u(x) &= f(x) + \int_a^b k(x,t)u(t)\,dt, \qquad a \leq x \leq b, \\ f(x) &= c_1 + \int_a^x (x-t)g(t)\,dt + \frac{x-a}{b-a}\left[c_2 - c_1 - \int_a^b (b-t)g(t)\,dt\right], \\ k(x,t) &= \begin{cases} \frac{x-b}{b-a}\{A(t) - (a-t)[A'(t) - B(t)]\}, & x > t, \\ \frac{x-a}{b-a}\{A(t) - (b-t)[A'(t) - B(t)]\}, & x < t. \end{cases} \end{aligned} \tag{5.8.4}$$

5.8.3 SPECIAL EQUATIONS WITH SOLUTIONS

1. Generalized Abel equation: $$\int_0^x \frac{u(t)\,dt}{[h(x)-h(t)]^\alpha} = f(x).$$

   The solution is $$u(x) = \frac{\sin(\alpha\pi)}{\pi}\frac{d}{dx}\int_0^x \frac{h'(t)f(t)\,dt}{[h(x)-h(t)]^{1-\alpha}}$$
   where $0 \leq x \leq 1$, $0 \leq \alpha < 1$, $0 \leq h(x) \leq 1$, $h'(x) > 0$, and $h'(x)$ is continuous.

2. Cauchy equation: $$\mu u(x) = f(x) + ⨍_0^1 \frac{u(t)}{t-x}\,dt.$$

   The solution is

   $$u(x) = \begin{cases} \dfrac{x^\gamma \sin^2(\pi\gamma)}{\pi^2}\dfrac{d}{dx}\displaystyle\int_x^1 \frac{ds}{(s-x)^\gamma}\int_0^s \frac{t^{-\gamma}f(t)}{(s-t)^{1-\gamma}}\,dt, & \mu < 0, \\ \dfrac{(1-x)^\gamma \sin^2(\pi\gamma)}{\pi^2}\dfrac{d}{dx}\displaystyle\int_0^x \frac{ds}{(x-s)^\gamma}\int_s^1 \frac{(1-t)^{-\gamma}f(t)}{(t-s)^{1-\gamma}}\,dt, & \mu > 0, \end{cases} \tag{5.8.6}$$

   where $0 < x < 1$, μ is real, $\mu \neq 0$, $|\mu| = \pi\cot(\pi\gamma)$, $0 < \gamma < 1/2$, and the integral is a Cauchy principal value integral.

3. Volterra equation with difference kernel:
$$u(x) = f(x) + \lambda \int_0^x k(x-t)u(t)\,dt.$$
The solution is $$u(x) = \mathcal{L}^{-1}\left[\frac{F(s)}{1-\lambda K(s)}\right],$$
for $x \geq 0$, $\mathcal{L}[f(x)] = F(s)$ and $\mathcal{L}[k(x)] = K(s)$.

4. Singular equation with difference kernel:
$$u(x) = f(x) + \lambda \int_{-\infty}^{\infty} k(x-t)u(t)\,dt.$$
The solution is $$u(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\alpha x} \frac{F(\alpha)}{1-\lambda K(\alpha)}\,d\alpha$$
where $-\infty < x < \infty$, $F(\alpha) = \mathcal{F}[f(x)]$ and $K(\alpha) = \mathcal{F}[k(x)]$

5. Fredholm equation with separable kernel:
$$u(x) = f(x) + \lambda \int_a^b \sum_{k=1}^n a_k(x)b_k(t)u(t)\,dt.$$
The solution is $$u(x) = f(x) + \lambda \sum_{k=1}^n c_k a_k(x),$$
with $c_m = \displaystyle\int_a^b b_m(t)f(t)\,dt + \lambda \sum_{k=1}^n c_k \int_a^b b_m(t)a_k(t)\,dt$ where $a \leq x \leq b$, $n < \infty$, and $m = 1, 2, \ldots, n$ (see page 494).

6. Fredholm equation with symmetric kernel:
$$u(x) = f(x) + \lambda \int_a^b k(x,t)u(t)\,dt.$$

   Solve $u_n(x) = \lambda_n \int_a^b k(x,t)u_n(t)\,dt$ for $\{u_n, \lambda_n\}_{n=1,2,\ldots}$. Then

   (a) For $\lambda \neq \lambda_n$, solution is $u(x) = f(x) + \lambda \displaystyle\sum_{n=1}^{\infty} \frac{u_n(x)\int_a^b f(t)u_n(t)\,dt}{(\lambda_n - \lambda)\int_a^b u_n^2(t)\,dt}$.

   (b) For $\lambda = \lambda_n$ and $\int_a^b f(t)u_m(t)\,dt = 0$ for all m, solutions are
   $$u(x) = f(x) + cu_m(x) + \lambda_m \sum_{\substack{n=1\\n\neq m}}^{\infty} \frac{u_n(x)\int_a^b f(t)u_n(t)\,dt}{(\lambda_n - \lambda_m)\int_a^b u_n^2(t)\,dt},$$
   for $m = 1, 2, \ldots$.

   where $a \leq x \leq b$, and $k(x,t) = k(t,x)$ (see Section 5.8.4).

7. Volterra equation of second kind:
$$u(x) = f(x) + \lambda \int_a^x k(x,t)u(t)\,dt.$$
The solution is $$u(x) = f(x) + \lambda \int_a^x \sum_{n=0}^{\infty} \lambda^n k_{n+1}(x,t)f(t)\,dt,$$
where $k_1(x,t) = k(x,t)$, $k_{n+1}(x,t) = \int_t^x k(x,s)k_n(s,t)\,ds$, when $k(x,t)$ and $f(x)$ are continuous, $\lambda \neq 0$, and $x \geq a$.

5.8.4 FREDHOLM ALTERNATIVE

For $u(x) = f(x) + \lambda \int_a^b k(x,t)u(t)\,dt$ with $\lambda \neq 0$, consider the solutions to $u_H(x) = \lambda \int_a^b k(x,t)u_H(t)\,dt$.

1. If the only solution is $u_H(x) = 0$, then there is a unique solution $u(x)$.
2. If $u_H(x) \neq 0$, then there is no solution unless $\int_a^b u_H^*(t)f(t)\,dt = 0$ for all $u_H^*(x)$ such that $u_H^*(x) = \lambda \int_a^b k(t,x)u_H^*(t)\,dt$. In this case, there are infinitely many solutions.

5.9 TENSOR ANALYSIS

5.9.1 DEFINITIONS

1. An *n-dimensional coordinate manifold* of class C^k, $k \geq 1$, is a point set M together with the totality of allowable coordinate systems on M. An *allowable coordinate system* (ϕ, U) on M is a one-to-one mapping $\phi : U \to M$, where U is an open subset of $\mathbb{R}^n$. The n-tuple $(x^1, \ldots, x^n) \in U$ give the *coordinates* of the corresponding point $\phi(x^1, \ldots, x^n) \in M$.

   If $(\tilde{\phi}, \tilde{U})$ is a second coordinate system on M, then the one-to-one correspondence $\tilde{\phi}^{-1} \circ \phi : U \to \tilde{U}$, called a *coordinate transformation* on M, is assumed to be of class C^k. It may be written as $\tilde{x}^i = \tilde{f}^i(x^1, \ldots, x^n), \quad i = 1, \ldots, n$,where the $\tilde{f}$ are defined by $(\tilde{\phi}^{-1} \circ \phi)(x^1, \ldots, x^n) = (\tilde{f}^1(x^1, \ldots, x^n), \ldots, \tilde{f}^n(x^1, \ldots, x^n))$.

   The coordinate transformation $\tilde{\phi}^{-1} \circ \phi$ has inverse $\phi^{-1} \circ \tilde{\phi}$, expressible in terms of the coordinates as $x^i = f^i(\tilde{x}^1, \ldots, \tilde{x}^n), \quad i = 1, \ldots, n$.
2. In the *Einstein summation convention*, a repeated upper and lower index signifies summation over the range $k = 1, \ldots, n$.
3. The Jacobian matrix of the transformation, $\frac{\partial \tilde{x}^i}{\partial x^j}$, satisfies $\frac{\partial \tilde{x}^i}{\partial x^k}\frac{\partial x^k}{\partial \tilde{x}^j} = \delta^i_j$ and $\dfrac{\partial x^i}{\partial \tilde{x}^k}\dfrac{\partial \tilde{x}^k}{\partial x^j} = \delta^i_j$, where $\delta^i_j = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$ denotes the Kronecker delta. Note also that $\det\left(\frac{\partial \tilde{x}^i}{\partial x^j}\right) \neq 0$.
4. A function $F : M \to \mathbb{R}$ is called a *scalar invariant* on M. The *coordinate representation of* F in any coordinate system (ϕ, U) is defined by $f := F \circ \phi$. The coordinate representations $\tilde{f}$ of F with respect to a second coordinate system $(\tilde{\phi}, \tilde{U})$ is related to f by $\tilde{f}(\tilde{x}^1, \ldots, \tilde{x}^n) = f(f^1(\tilde{x}^1, \ldots, \tilde{x}^n), \ldots, f^n(\tilde{x}^1, \ldots, \tilde{x}^n))$.

5. A *mixed tensor T of contravariant valence r, covariant valence s, and weight* w at $p \in M$, called a tensor of type (r, s, w), is an object which, with respect to each coordinate system on M, is represented by n^{r+s} real numbers whose values in any two coordinate systems, ϕ and $\tilde{\phi}$, are related by

$$\widetilde{T}^{i_1 \cdots i_r}{}_{j_1 \cdots j_s} = \left[\det\left(\frac{\partial x^i}{\partial \tilde{x}^j}\right)\right]^w T^{k_1 \cdots k_r}{}_{\ell_1 \cdots \ell_s} \underbrace{\frac{\partial \tilde{x}^{i_1}}{\partial x^{k_1}} \cdots \frac{\partial \tilde{x}^{i_r}}{\partial x^{k_r}}}_{r \text{ factors}} \underbrace{\frac{\partial x^{\ell_1}}{\partial \tilde{x}^{j_1}} \cdots \frac{\partial x^{\ell_s}}{\partial \tilde{x}^{j_s}}}_{s \text{ factors}}. \tag{5.9.1}$$

The superscripts are called *contravariant indices* and the subscripts *covariant indices*. If $w \neq 0$, then T is said to be a *relative tensor*. If $w = 0$, then T is said to be an *absolute tensor* or a tensor of type (r, s). In this section only absolute tensors, which will be called tensors, will be considered unless otherwise indicated. A *tensor field T of type* (r, s) is an assignment of a tensor of type (r, s) to each point of M. A tensor field T is C^k if its component functions are C^k for every coordinate system on M.

6. A *parameterized curve* on M is a mapping $\gamma : I \to M$, where $I \subset \mathbb{R}$ is some interval. The *coordinate representation of* γ in any coordinate system (ϕ, U) is a mapping $g : I \to \mathbb{R}^n$ defined by $g = \phi^{-1} \circ \gamma$. The mapping g defines a parameterized curve in $\mathbb{R}^n$. The component functions of g denoted by g^i (for $i = 1, \ldots, n$) are defined by $g(t) = (g^1(t), \cdots, g^n(t))$. The curve γ is C^k if and only if the functions g^i are C^k for every coordinate system on M. The coordinate representation $\tilde{g}$ of γ with respect to a second coordinate system $(\tilde{\phi}, \tilde{u})$ is related to g by $\tilde{g}^i(t) = \tilde{f}^i(g^1(t), \cdots, g^n(t))$.

5.9.2 ALGEBRAIC TENSOR OPERATIONS

1. *Addition:* The components of the *sum* of the tensors T_1 and T_2 of type (r, s) are given by

$$T_3{}^{i_1 \cdots i_r}{}_{j_1 \cdots j_s} = T_1{}^{i_1 \cdots i_r}{}_{j_1 \cdots j_s} + T_2{}^{i_1 \cdots i_r}{}_{j_1 \cdots j_s}. \tag{5.9.2}$$

2. *Multiplication:* The components of the *tensor* or *outer product* of a tensor T_1 of type (r, s) and a tensor T_2 of type (t, u) are given by

$$T_3{}^{i_1 \cdots i_r k_1 \cdots k_t}{}_{j_1 \cdots j_s \ell_1 \cdots \ell_u} = T_1{}^{i_1 \cdots i_r}{}_{j_1 \cdots j_s} T_2{}^{k_1 \cdots k_t}{}_{\ell_1 \cdots \ell_u}. \tag{5.9.3}$$

3. *Contraction:* The components of the *contraction* of the t^{th} contravariant index with the u^{th} covariant index of a tensor T of type (r, s), with $rs \neq 0$, are given by $T^{\,i_1 \cdots i_{t-1} k i_{t+1} \cdots i_r}{}_{j_1 \cdots j_{u-1} k j_{u+1} \cdots j_s}$.

4. *Permutation of indices:* Let T be any tensor of type $(0, r)$ and S_r the group of permutations of the set $\{1, \cdots, r\}$. The components of the tensor, obtained by permuting the indices of T with any $\sigma \in S_r$, are given by $(\sigma T)_{i_1 \cdots i_r} =$

$T_{i_{\sigma(1)}\cdots i_{\sigma(r)}}$. The *symmetric part* of T, denoted by $\mathcal{S}(T)$, is the tensor whose components are given by

$$\mathcal{S}(T)_{i_1\cdots i_r} = T_{(i_1\cdots i_r)} = \frac{1}{r!}\sum_{\sigma\epsilon S_r} T_{i_{\sigma(1)}\cdots i_{\sigma(r)}}. \tag{5.9.4}$$

The tensor T is said to be *symmetric* if and only if $T_{i_1\cdots i_r} = T_{(i_1\cdots i_r)}$. The *skew symmetric* part of T, denoted by $\mathcal{A}(T)$, is the tensor whose components are given by

$$\mathcal{A}(T)_{i_1\cdots i_r} = T_{[i_1\cdots i_r]} = \frac{1}{r!}\sum_{\sigma\in S_r} \text{sgn}(\sigma) T_{i_{\sigma(1)}\cdots i_{\sigma(r)}}, \tag{5.9.5}$$

where $\text{sgn}(\sigma) = \pm 1$ according to whether σ is an even or odd permutation. The tensor T is said to be *skew symmetric* if and only if $T_{i_1\cdots i_r} = T_{[i_1\cdots i_r]}$. If $r = 2$, $\mathcal{S}(T)_{i_1 i_2} = \frac{1}{2}(T_{i_1 i_2} + T_{i_2 i_1})$ and $\mathcal{A}(T)_{i_1 i_2} = \frac{1}{2}(T_{i_1 i_2} - T_{i_2 i_1})$.

5.9.3 DIFFERENTIATION OF TENSORS

1. In tensor analysis, a comma is used to denote partial differentiation and a semicolon to denote covariant differentiation.

2. A *linear connection* ∇ at $p \in M$ is an object which, with respect to each coordinate system on M, is represented by n^3 real numbers $\Gamma^i{}_{jk}$, called the *connection coefficients*, whose values in any two coordinate systems ϕ and $\tilde{\phi}$ are related by

$$\widetilde{\Gamma}^i{}_{jk} = \Gamma^\ell{}_{mn}\frac{\partial \tilde{x}^i}{\partial x^\ell}\frac{\partial x^m}{\partial \tilde{x}^j}\frac{\partial x^n}{\partial \tilde{x}^k} + \frac{\partial^2 x^\ell}{\partial \tilde{x}^j \partial \tilde{x}^k}\frac{\partial \tilde{x}^i}{\partial x^\ell} \tag{5.9.6}$$

The quantities $\Gamma^i{}_{jk}$ are *not* the components of a tensor of type $(1,2)$. A linear connection ∇ on M is an assignment of a linear connection to each point of M. A connection ∇ is C^k if its connection coefficients $\Gamma^i{}_{jk}$ are C^k in every coordinate system on M.

3. The components of the *covariant derivative* of a tensor field T of type (r,s), with respect to a connection ∇, are given by

$$\begin{aligned}\nabla_k T^{i_1\cdots i_r}{}_{j_1\cdots j_s} &= \partial_k T^{i_1\cdots i_r}{}_{j_1\cdots j_s} + \Gamma^{i_1}{}_{\ell k} T^{\ell i_2\cdots i_r}{}_{j_1\cdots j_s} + \cdots \\ &\cdots + \Gamma^{i_r}{}_{\ell k} T^{i_1\cdots i_{r-1}\ell}{}_{j_1\cdots j_s} - \Gamma^\ell{}_{j_1 k} T^{i_1\cdots i_r}{}_{\ell j_2\cdots j_s} \cdots - \Gamma^\ell{}_{j_s k} T^{i_1\cdots i_r}{}_{j_1\cdots j_{s-1}\ell},\end{aligned} \tag{5.9.7}$$

where $\partial_k T^{i_1\cdots i_r}{}_{j_i\cdots j_s} = T^{i_1\cdots i_r}{}_{j_1\cdots j_s,k} = \dfrac{\partial T^{i_1\cdots i_r}{}_{j_1\cdots j_s}}{\partial x^k}$.

This formula has this structure:

(a) A partial derivative term, a negative affine term for each covariant index and a positive affine term for each contravariant index.

(b) The second subscript in the Γ-symbols is always the differentiated index (k in this case).

4. Let $Y^i(t)$ be a contravariant vector field and $Z_i(t)$ a covariant vector field defined along a parameterized curve γ. The *absolute covariant derivatives* of Y^i and Z_i are defined as follows:

$$\begin{aligned} \frac{\delta Y^i}{\delta t} &= \frac{dY^i}{dt} + \Gamma^i{}_{jk}\, Y^j\, \frac{dx^k}{dt} \\ \frac{\delta Z_i}{\delta t} &= \frac{dZ_i}{dt} - \Gamma^j{}_{ik}\, Z_j\, \frac{dx^k}{dt} \end{aligned} \tag{5.9.8}$$

where x^i denotes the components of γ in the coordinate system ϕ. This derivative may also be defined for tensor fields of type (r, s) defined along γ.

5. A parameterized curve γ in M is said to be an *affinely parameterized geodesic* if the component functions of γ satisfy

$$\frac{\delta}{\delta t}\left(\frac{dx^i}{dt}\right) = \frac{d^2x^i}{dt^2} + \Gamma^i{}_{jk}\,\frac{dx^j}{dt}\,\frac{dx^k}{dt} = 0$$

which is equivalent to the statement that the tangent vector $\frac{dx^i}{dt}$ to γ is parallel along γ.

6. A vector field $Y^i(t)$ is *parallel along* a parameterized curve γ if $\frac{\delta Y^i}{\delta t} = 0$.

7. The components of the *torsion tensor* S of ∇ on M are defined by

$$S^i{}_{jk} = \Gamma^i{}_{jk} - \Gamma^i{}_{kj}. \tag{5.9.9}$$

8. The components of the *curvature tensor* R of ∇ on M are defined by

$$R^i{}_{jk\ell} = \partial_k \Gamma^i{}_{j\ell} - \partial_\ell \Gamma^i{}_{jk} + \Gamma^m{}_{j\ell}\Gamma^i{}_{mk} - \Gamma^m{}_{jk}\Gamma^i{}_{m\ell}. \tag{5.9.10}$$

(In some references R is defined with the opposite sign.)

9. The *Ricci tensor* of ∇ is defined by $R_{jk} = R^\ell{}_{jk\ell}$.

5.9.4 METRIC TENSOR

1. A *covariant metric tensor field* on M is a tensor field g_{ij} which satisfies $g_{ij} = g_{ji}$ and $g = \det(g_{ij}) \neq 0$ on M. The *contravariant metric* g^{ij} satisfies $g^{ik}g_{kj} = \delta^i_j$. The *line element* is expressible in terms of the metric tensor as $ds^2 = g_{ij}dx^i dx^j$.

2. *Signature of the metric:* For each $p \in M$, a coordinate system exists such that $g_{ij}(p) = \text{diag}\,\underbrace{(1, \cdots, 1}_{r}, \underbrace{-1, \cdots, -1)}_{n-r}$. The *signature* of g_{ij} is defined by $s = 2r - n$. It is independent of the coordinate system in which $g_{ij}(p)$ has the above diagonal form and is the same at every $p \in M$. A metric is said to be *positive definite* if $s = n$. A manifold, admitting a positive definite metric, is called a *Riemannian manifold*. A metric is said to be *indefinite* if $s \neq n$ and $s \neq -n$. A manifold, admitting an indefinite metric, is called a *pseudo-Riemannian manifold*. If $s = 2 - n$ or $n - 2$, the metric is said to be *Lorentzian* and the corresponding manifold is called a *Lorentzian manifold*.

3. The *inner product* of a pair of vectors X^i and Y^j is given by $g_{ij}X^iY^j$. If $X^i = Y^i$, then $g_{ij}X^iX^j$ defines the "square" of the length of X^i. If g_{ij} is positive definite, then $g_{ij}X^iX^j \geq 0$ for all X^i, and $g_{ij}X^iX^j = 0$ if and only if $X^i = 0$. In the positive definite case, the *angle* θ between two tangent vectors X^i and Y^j is defined by $\cos\theta = g_{ij}X^iY^j/(g_{k\ell}X^kX^\ell g_{mn}Y^mY^n)^{\frac{1}{2}}$. If g is indefinite, $g_{ij}X^iX^j$ may have a positive, negative, or zero value. A non-zero vector X^i, satisfying $g_{ij}X^iX^j = 0$, is called a *null vector*. If g_{ij} is indefinite, it is not possible in general to define the angle between two tangent vectors.

4. *Operation of lowering indices:* The components of the tensor resulting from *lowering the t^{th} contravariant index* of a tensor T of type (r, s), with $r \geq 1$, are given by

$$T^{i_1\cdots i_{t-1}\,\cdot\,i_{t+1}\cdots i_r}{}_{i_t\qquad j_1\cdots j_s} = g_{i_t k}T^{i_1\cdots i_{t-1}ki_{t+1}\cdots i_r}{}_{j_1\cdots j_s}. \tag{5.9.11}$$

5. *Operation of raising indices:* The components of the tensor resulting from *raising the t^{th} covariant index* of a tensor T of type (r, s), with $s \geq 1$, are given by

$$T^{\,i_1\cdots i_r\qquad\quad j_t}{}_{j_i\cdots j_{t-1}\,\cdot\,j_{t+1}\cdots j_s} = g^{j_t k}T^{i_1\cdots i_r}{}_{j_1\cdots j_{t-1}kj_{t+1}\cdots j_s}. \tag{5.9.12}$$

6. The *arc length* of a parameterized curve $\gamma : I \to M$, where $I = [a, b]$, and ϕ is any coordinate system, is defined by

$$L = \int_a^b \sqrt{\epsilon g_{ij}(x^1(t), \cdots, x^n(t))\dot{x}^i\dot{x}^j}\, dt, \tag{5.9.13}$$

where $\epsilon = \text{sgn}(g_{ij}\dot{x}^i\dot{x}^j) = \pm 1$ and $\dot{x}^i = \frac{dx^i}{dt}$.

5.9.5 RESULTS

The following results hold on any manifold M admitting any connection ∇:

1. The covariant derivative operator ∇_k is linear with respect to tensor addition, satisfies the product rule with respect to tensor multiplication, and commutes with contractions of tensors.

2. If T is any tensor of type $(0, r)$, then

$$\nabla_{[k}T_{i_1\cdots i_r]} = T_{[i_1\cdots i_r,k]} - \frac{1}{2}\left(S^{\ell}{}_{[i_1 k}T_{|\ell|i_2\cdots i_r]} + \cdots + S^{\ell}{}_{[i_{rk}}T_{i_1\cdots i_{r-1}]\ell}\right),$$

where $|\,|$ indicates that the enclosed indices are excluded from the symmetrization. Thus $T_{[i_1\cdots i_r,k]}$ defines a tensor of type $(0, r+1)$, and $\nabla_{[k}T_{i_1\cdots i_r]} = T_{[i_1\cdots i_r,k]}$ in the torsion free case. If $T_j = \nabla_j f = f_{,j}$, where f is any scalar invariant, then $\nabla_{[i}\nabla_{j]}f = \frac{1}{2}f_{,k}S^{k}{}_{ij}$. In the torsion free case, $\nabla_i\nabla_j f = \nabla_j\nabla_i f$.

3. If X^i is any contravariant vector field on M, then the identity $2\nabla_{[j}\nabla_{k]}X^i + \nabla_\ell X^i S^{\ell}{}_{jk} = X^\ell R^{i}{}_{\ell jk}$, called the *Ricci identity*, reduces to $2\nabla_{[j}\nabla_{k]}X^i = R^{i}{}_{\ell jk}X^\ell$, in the torsion free case. If Y_i is any covariant vector field, the Ricci identity has the form $2\nabla_{[i}\nabla_{j]}Y_k - \nabla_\ell Y_k S^{\ell}{}_{ij} = -Y_\ell R^{\ell}{}_{kij}$. The Ricci identity may be extended to tensor fields of type (r, s). For the tensor field $T^{i}{}_{jk}$, it has the form

$$2\nabla_{[i}\nabla_{j]}T^{k}{}_{\ell m} - \nabla_n T^{k}{}_{\ell m}S^{n}{}_{ij} = T^{n}{}_{\ell m}R^{k}{}_{nij} - T^{k}{}_{nm}R^{n}{}_{\ell ij} - T^{k}{}_{\ell n}R^{n}{}_{mij}.$$

If g is any metric tensor field, the above identity implies that

$$R_{(ij)k\ell} = \nabla_{[k}\nabla_{\ell]}g_{ij} - \tfrac{1}{2}\nabla_m g_{ij}S^{m}{}_{k\ell}.$$

4. The torsion tensor S and curvature tensor R satisfy the following identities:

$$S^{i}{}_{(jk)} = 0, \qquad 0 = R^{i}{}_{j[kl;m]} - R^{i}{}_{jn[k}S^{n}{}_{\ell m]}, \tag{5.9.14}$$

$$R^{i}{}_{j(k\ell)} = 0, \qquad R^{i}{}_{[jk\ell]} = -S^{i}{}_{[jk;\ell]} + S^{i}{}_{m[j}S^{m}{}_{k\ell]}. \tag{5.9.15}$$

In the torsion free case, these identities reduce to the *cyclical identity* $R^{i}{}_{[jk\ell]} = 0$ and *Bianchi's identity* $R^{i}{}_{j[k\ell;m]} = 0$.

The following results hold for any pseudo-Riemannian manifold M with a metric tensor field g_{ij}:

5. A unique connection ∇ called the *Levi–Civita* or *pseudo-Riemannian connection* with vanishing torsion ($S^{i}{}_{jk} = 0$) exists that satisfies $\nabla_i g_{jk} = 0$. It follows that $\nabla_i g^{jk} = 0$. The connection coefficients of ∇, called the *Christoffel symbols of the second kind*, are given by $\Gamma^{i}{}_{jk} = g^{i\ell}[jk,\ell]$, where $[jk,\ell] = \frac{1}{2}(g_{j\ell,k} + g_{k\ell,j} - g_{jk,\ell})$ are the *Christoffel symbols of the first kind.* $\Gamma^{k}{}_{jk} = \frac{1}{2}\partial_j(\log g) = |g|^{-\frac{1}{2}}\partial_j|g|^{\frac{1}{2}}$ and $g_{ij,k} = [ki,j] + [kj,i]$.

6. The operations of raising and lowering indices commute with the covariant derivative. For example if $X_i = g_{ij}X^j$, then $\nabla_k X_i = g_{ij}\nabla_k X^j$.
7. The *divergence* of a vector X^i is given by $\nabla_i X^i = |g|^{-\frac{1}{2}}\partial_i(|g|^{\frac{1}{2}}X^i)$. The *Laplacian* of a scalar invariant f is given by $\Delta f = g^{ij}\nabla_i\nabla_j f = \nabla_i(g^{ij}\nabla_j f) = |g|^{-\frac{1}{2}}\partial_i(|g|^{\frac{1}{2}}g^{ij}\partial_j f)$.
8. The equations of an affinely parameterized geodesic may be written as $\frac{d}{dt}(g_{ij}\dot{x}^j) - \frac{1}{2}g_{jk,i}\dot{x}^j\dot{x}^k = 0$.
9. Let X^i and Y^i be the components of any vector fields which are propagated in parallel along any parameterized curve γ. Then $\frac{d}{dt}(g_{ij}X^iY^j) = 0$, which implies that the inner product $g_{ij}X^iY^j$ is constant along γ. In particular, if $\dot{x}^i$ are the components of the tangent vector to γ, then $g_{ij}\dot{x}^i\dot{x}^j$ is constant along γ.
10. The *Riemann tensor*, defined by $R_{ijk\ell} = g_{im}R^m_{\ jk\ell}$, is given by

$$\begin{aligned} R_{ijk\ell} &= [j\ell, i]_{,k} - [jk, i]_{,\ell} + [i\ell, m]\Gamma^m_{\ jk} - [ik, m]\Gamma^m_{\ j\ell} \\ &= \frac{1}{2}(g_{i\ell,jk} + g_{jk,i\ell} - g_{j\ell,ik} - g_{ik,j\ell}) \\ &\qquad + g^{mn}([i\ell, m][jk, n] - [ik, m][j\ell, n]). \end{aligned} \tag{5.9.16}$$

It has the following symmetries:

$$R_{ij(k\ell)} = R_{(ij)k\ell} = 0, \quad R_{ijk\ell} = R_{k\ell ij}, \text{ and } R_{i[jk\ell]} = 0. \tag{5.9.17}$$

Consequently it has a maximum of $n^2(n^2-1)/12$ independent components.
11. The equations $R_{ijk\ell} = 0$ are necessary and sufficient conditions for M to be a *flat* pseudo-Riemannian manifold, that is, a manifold for which a coordinate system exists so that the components g_{ij} are *constant* on M.
12. The Ricci tensor is given by

$$\begin{aligned} R_{ij} &= \partial_j\Gamma^k_{\ ik} - \partial_k\Gamma^k_{\ ij} + \Gamma^k_{\ i\ell}\Gamma^\ell_{\ kj} - \Gamma^k_{\ ij}\Gamma^\ell_{\ k\ell} \\ &= \frac{1}{2}\partial_i\partial_j(\log|g|) - \frac{1}{2}\Gamma^k_{\ ij}\partial_k(\log|g|) - \partial_k\Gamma^k_{\ ij} + \Gamma^k_{\ im}\Gamma^m_{\ kj}. \end{aligned}$$

It possesses the symmetry $R_{ij} = R_{ji}$, and thus has a maximum of $n(n+1)/2$ independent components.
13. The *scalar curvature* or *curvature invariant* is defined by $R = g^{ij}R_{ij}$.
14. The *Einstein tensor* is defined by $G_{ij} = R_{ij} - \frac{1}{2}Rg_{ij}$. In view of the Bianchi identity, it satisfies: $g^{jk}\nabla_j G_{ki} = 0$.
15. A *normal coordinate system* with origin $x_0 \epsilon M$ is defined by $\overset{0}{g}_{ij}\, x^j = g_{ij}x^j$, where a "0" affixed over a quantity indicates that the quantity is evaluated at x_0. The connection coefficients satisfy $\overset{0}{\Gamma}{}^i_{\ (j_1j_2\cdots j_r)} = 0$ (for $r = 2, 3, 4, \dots$) in any normal coordinate system. The equations of the geodesics through x_0 are given by $x^i = sk^i$, where s is an affine parameter and k^i is any constant vector.

5.9.6 EXAMPLES OF TENSORS

1. The components of the *gradient* of a scalar invariant $\frac{\partial f}{\partial x^i}$ define a tensor of type (0,1), since they transform as $\frac{\partial \tilde{f}}{\partial \tilde{x}^i} = \frac{\partial f}{\partial x^j}\frac{\partial x^j}{\partial \tilde{x}^i}$.
2. The components of the *tangent vector* to a parameterized curve $\frac{dx^i}{dt}$ define a tensor of type (1,0), because they transform as $\frac{d\tilde{x}^i}{dt} = \frac{dx^j}{dt}\frac{\partial \tilde{x}^i}{\partial x^j}$.
3. The determinant of the metric tensor g defines a relative scalar invariant of weight of $w = 2$, because it transforms as $\tilde{g} = \left|\frac{\partial x^i}{\partial \tilde{x}^j}\right|^2 g$.
4. The Kronecker deltas δ_i^j are the components of a constant absolute tensor of type $(1,1)$, because $\delta_j^i = \delta_\ell^k \frac{\partial \tilde{x}^i}{\partial x^k}\frac{\partial x^\ell}{\partial \tilde{x}^j}$.
5. The permutation symbol defined by

$$e_{i_1\cdots i_n} = \begin{cases} 1, & \text{if } i_1\cdots i_n \text{ is an even permutation of } 1\cdots n, \\ -1, & \text{if } i_1\cdots i_n \text{ is an odd permutation of } 1\cdots n, \text{ and} \\ 0 & \text{otherwise,} \end{cases} \tag{5.9.18}$$

satisfies $\left|\frac{\partial x^i}{\partial \tilde{x}^i}\right| e_{j_1\cdots j_n} = e_{i_1\cdots i_n}\frac{\partial x^{i_1}}{\partial \tilde{x}^{j_1}}\cdots\frac{\partial x^{i_n}}{\partial \tilde{x}^{j_n}}$. Hence it defines a tensor of type $(0, n, -1)$, that is, it is a relative tensor of weight $w = -1$. The contravariant permutation symbol $e^{i_1\cdots i_n}$, defined in a similar way, is a relative tensor of weight $w = 1$.
6. The Levi–Civita symbol, $\epsilon_{i_1\cdots i_n} = |g|^{\frac{1}{2}} e_{i_1\cdots i_n}$, defines a covariant absolute tensor of valence n. The contravariant Levi–Civita tensor satisfies

$$\epsilon^{i_1\cdots i_n} = g^{i_1 j_1}\cdots g^{i_n j_n}\epsilon_{j_1\cdots j_n} = (-1)^{\frac{n-s}{2}}|g|^{-\frac{1}{2}} e^{i_1\cdots i_n}. \tag{5.9.19}$$

Using this symbol, the *dual* of a covariant skew-symmetric tensor of valence r is defined by ${}^*T_{i_1\cdots i_{n-r}} = \frac{1}{r!}\epsilon^{j_1\cdots j_r}{}_{i_1\cdots i_{n-r}} T_{j_1\cdots j_r}$.
7. *Cartesian tensors:* Let $M = E^3$ (i.e., Euclidean 3 space) with metric tensor $g_{ij} = \delta_{ij}$ with respect to Cartesian coordinates. The components of a *Cartesian tensor* of valence r transform as

$$\widetilde{T}_{i_1\cdots i_r} = T_{j_1\cdots j_r}\mathrm{O}_{i_1 j_1}\cdots\mathrm{O}_{i_r j_r}, \tag{5.9.20}$$

where O_{ij} are the components of a constant orthogonal matrix which satisfies $(\mathrm{O}^{-1})_{ij} = (\mathrm{O}^{\mathrm{T}})_{ij} = \mathrm{O}_{ji}$. For Cartesian tensors, all indices are written as covariant, because *no* distinction is required between covariant and contravariant indices.
8. Note the useful relations: $\epsilon_{ijk}\epsilon_{klm} = \delta_{il}\delta_{jm} - \delta_{im}\delta_{jl}$, $\epsilon_{ikl}\epsilon_{klm} = 2\delta_{im}$, and $\epsilon_{ijk}\epsilon_{lmn} = \delta_{il}\delta_{jm}\delta_{kn} + \delta_{im}\delta_{jn}\delta_{kl} + \delta_{in}\delta_{jl}\delta_{km} - \delta_{in}\delta_{jm}\delta_{kl} - \delta_{im}\delta_{jl}\delta_{kn} - \delta_{il}\delta_{jn}\delta_{km}$.
9. *Polar coordinates:* The line element is given by $ds^2 = dr^2 + r^2 d\theta^2$. Thus the metric tensor is $g_{ij} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & r^2 \end{smallmatrix}\right)$, and the non-zero Christoffel symbols are $[21, 2] = [12, 2] = -[22, 1] = r$.

10. *Orthogonal curvilinear coordinates*: Let M be a 3-dimensional Riemannian manifold admitting a coordinate system $[x^1, x^2, x^3]$ such that the metric tensor has the form $g_{ii} = h_i^2(x^1, x^2, x^3)$ for $i = 1, \ldots, 3$ with $g_{ij} = g^{ij} = 0$ for $i \neq j$. The metric tensor on E^3 has this form with respect to orthogonal curvilinear coordinates. The non-zero components of various quantities corresponding to this metric are:

    (a) Covariant metric tensor

    $$g_{11} = {h_1}^2, \qquad g_{22} = {h_2}^2, \qquad g_{33} = {h_3}^2 \tag{5.9.21}$$

    (b) Contravariant metric tensor

    $$g^{11} = {h_1}^{-2}, \qquad g^{22} = {h_2}^{-2}, \qquad g^{33} = {h_3}^{-2} \tag{5.9.22}$$

    (c) Christoffel symbols of the first kind (note that $[ij, k] = 0$ if i, j, and k are all different),

    $$\begin{array}{lll} [11,1] = h_1 h_{1,1} & [11,2] = -h_1 h_{1,2} & [11,3] = -h_1 h_{1,3} \\ [12,1] = h_1 h_{1,2} & [12,2] = h_2 h_{2,1} & [13,1] = h_1 h_{1,3} \\ [13,3] = h_3 h_{3,1} & [22,1] = -h_2 h_{2,1} & [22,2] = h_2 h_{2,2} \\ [22,3] = -h_2 h_{2,3} & [23,2] = h_2 h_{2,3} & [23,3] = h_3 h_{3,2} \\ [33,1] = -h_3 h_{3,1} & [33,2] = -h_3 h_{3,2} & [33,3] = h_3 h_{3,3}. \end{array}$$

    (d) Christoffel symbols of the second kind (note that $\Gamma^k{}_{ij} = 0$ if i, j, and k are all different),

    $$\begin{array}{lll} \Gamma^1{}_{11} = {h_1}^{-1} h_{1,1} & \Gamma^1{}_{12} = {h_1}^{-1} h_{1,2} & \Gamma^1{}_{13} = {h_1}^{-1} h_{1,3} \\ \Gamma^1{}_{22} = -{h_1}^{-2} h_2 h_{2,1} & \Gamma^1{}_{33} = -{h_1}^{-2} h_3 h_{3,1} & \Gamma^2{}_{11} = -h_1 {h_2}^{-2} h_{1,2} \\ \Gamma^2{}_{12} = {h_2}^{-1} h_{2,1} & \Gamma^2{}_{22} = {h_2}^{-1} h_{2,2} & \Gamma^2{}_{23} = {h_2}^{-1} h_{2,3} \\ \Gamma^2{}_{33} = -{h_2}^{-2} h_3 h_{3,2} & \Gamma^3{}_{11} = -h_1 {h_3}^{-2} h_{1,3} & \Gamma^3{}_{13} = {h_3}^{-1} h_{3,1} \\ \Gamma^3{}_{22} = -h_2 {h_3}^{-2} h_{2,3} & \Gamma^3{}_{23} = {h_3}^{-1} h_{3,2} & \Gamma^3{}_{33} = {h_3}^{-1} h_{3,3}. \end{array}$$

    (e) Vanishing Riemann tensor conditions (Lamé equations),

    $$\begin{aligned} & h_{1,2,3} - {h_2}^{-1} h_{1,2} h_{2,3} - {h_3}^{-1} h_{1,3} h_{3,2} = 0, \\ & h_{2,1,3} - {h_1}^{-1} h_{1,3} h_{2,1} - {h_3}^{-1} h_{3,1} h_{2,3} = 0, \\ & h_{3,1,2} - {h_1}^{-1} h_{1,2} h_{3,1} - {h_2}^{-1} h_{2,1} h_{3,2} = 0, \\ & h_2 h_{2,3,3} + h_3 h_{3,2,2} + {h_1}^{-2} h_2 h_3 h_{2,1} h_{3,1} - {h_2}^{-1} h_3 h_{2,2} h_{3,2} \\ & \qquad - h_2 {h_3}^{-1} h_{2,3} h_{3,3} = 0, \\ & h_1 h_{1,3,3} + h_3 h_{3,1,1} + h_1 {h_2}^{-2} h_3 h_{1,2} h_{3,2} - {h_1}^{-1} h_3 h_{1,1} h_{3,1} \\ & \qquad - h_1 {h_3}^{-1} h_{1,3} h_{3,3} = 0, \\ & h_1 h_{1,2,2} + h_2 h_{2,1,1} + h_1 h_2 {h_3}^{-2} h_{1,3} h_{2,3} - {h_1}^{-1} h_2 h_{1,1} h_{2,1} \\ & \qquad - h_1 {h_2}^{-1} h_{1,2} h_{2,2} = 0. \end{aligned}$$

11. *The 2-sphere*: A coordinate system $[\theta, \phi]$ for the 2-sphere $x^2 + y^2 + z^2 = r^2$ is given by $x = r \sin\theta \cos\phi$, $y = r \sin\theta \sin\phi$, $z = r\cos\theta$, where $[\theta, \phi] \in U = (0, \pi) \times (0, 2\pi)$. This is a non-Euclidean space. The non-zero independent components of various quantities defined on the sphere are:
    (a) Covariant metric tensor components: $g_{11} = r^2, \quad g_{22} = r^2 \sin^2\theta.$
    (b) Contravariant metric tensor components:
    $$g^{11} = r^{-2}, \quad g^{22} = r^{-2} \csc^2\theta.$$
    (c) Christoffel symbols of the first kind:
    $$[12, 2] = r^2 \sin\theta \cos\theta, \qquad [22, 1] = -r^2 \sin\theta \cos\theta.$$
    (d) Christoffel symbols of the second kind:
    $$\Gamma^1{}_{22} = -\sin\theta\cos\theta, \qquad \Gamma^2{}_{12} = -\cos\theta \csc\theta.$$
    (e) Covariant Riemann tensor components: $R_{1212} = r^2 \sin^2\theta.$
    (f) Covariant Ricci tensor components: $R_{11} = -1, \quad R_{22} = -\sin^2\theta.$
    (g) The Ricci scalar: $R = -2r^{-2}$.

12. *The 3-sphere*: A coordinate system $[\psi, \theta, \phi]$ for the 3-sphere $x^2 + y^2 + z^2 + w^2 = r^2$ is given by $x = r \sin\psi \sin\theta \cos\phi$, $y = r \sin\psi \sin\theta \sin\phi$, $z = r \sin\psi \cos\theta$, and $w = r\cos\psi$, where $[\psi, \theta, \phi] \in U = (0, \pi) \times (0, \pi) \times (0, 2\pi)$. The non-zero components of various quantities defined on the sphere are:
    (a) Covariant metric tensor components:
    $$g_{11} = r^2, \qquad g_{22} = r^2 \sin^2\psi, \qquad g_{33} = r^2 \sin^2\psi \sin^2\theta$$
    (b) Contravariant metric tensor components:
    $$g^{11} = r^{-2}, \qquad g^{22} = r^{-2} \csc^2\psi, \qquad g^{33} = r^{-2} \csc^2\psi \csc^2\theta$$
    (c) Christoffel symbols of the first kind:
    $$\begin{aligned} &[22, 1] = -r^2 \sin\psi \cos\psi && [33, 1] = -r^2 \sin\psi \cos\psi \sin^2\theta \\ &[12, 2] = r^2 \sin\psi \cos\psi && [33, 2] = -r^2 \sin^2\psi \sin\theta \cos\theta \\ &[13, 3] = r^2 \sin\psi \cos\psi \sin^2\theta && [23, 3] = r^2 \sin^2\psi \sin\theta \cos\theta. \end{aligned}$$
    (d) Christoffel symbols of the second kind:
    $$\begin{aligned} &\Gamma^1{}_{22} = -\sin\psi\cos\psi && \Gamma^1{}_{33} = -\sin\psi\cos\psi\sin^2\theta \\ &\Gamma^2{}_{12} = \cot\psi && \Gamma^2{}_{33} = -\sin\theta\cos\theta \\ &\Gamma^3{}_{13} = \cot\psi && \Gamma^3{}_{23} = \cot\theta. \end{aligned}$$
    (e) Covariant Riemann tensor components: $R_{1212} = r^2 \sin^2\psi,$
    $$R_{1313} = r^2 \sin^2\psi \sin^2\theta, \qquad R_{2323} = r^2 \sin^4\psi \sin^2\theta.$$
    (f) Covariant Ricci tensor components:
    $$R_{11} = -2, \quad R_{22} = -2\sin^2\psi, \quad R_{33} = -2\sin^2\psi \sin^2\theta.$$
    (g) The Ricci scalar: $R = -6r^{-2}$.
    (h) Covariant Einstein tensor components:
    $$G_{11} = 1, \qquad G_{22} = \sin^2\psi, \qquad G_{33} = \sin^2\psi \sin^2\theta$$

5.10 ORTHOGONAL COORDINATE SYSTEMS

In an orthogonal coordinate system, let $\{\mathbf{a}_i\}$ denote the unit vectors in each of the three coordinate directions, and let $\{u_i\}$ denote distance along each of these axes. The coordinate system may be designated by the *metric coefficients* $\{g_{11}, g_{22}, g_{33}\}$, defined by

$$g_{ii} = \left(\frac{\partial x_1}{\partial u_i}\right)^2 + \left(\frac{\partial x_2}{\partial u_i}\right)^2 + \left(\frac{\partial x_3}{\partial u_i}\right)^2, \tag{5.10.1}$$

where $\{x_1, x_2, x_3\}$ represent rectangular coordinates. With these, we define $g = g_{11}g_{22}g_{33}$.

Operations for orthogonal coordinate systems are sometimes written in terms of $\{h_i\}$ functions, instead of the $\{g_{ii}\}$ terms. Here, $h_i = \sqrt{g_{ii}}$, so that $\sqrt{g} = h_1h_2h_3$. For example, in cylindrical coordinates, $\{x_1 = r\cos\theta,\ x_2 = r\sin\theta,\ x_3 = z\}$, so that $\{h_1 = 1,\ h_2 = r,\ h_3 = 1\}$.

In the following, ϕ represents a scalar, and $\mathbf{E} = E_1\mathbf{a}_1 + E_2\mathbf{a}_2 + E_3\mathbf{a}_3$ and $\mathbf{F} = F_1\mathbf{a}_1 + F_2\mathbf{a}_2 + F_3\mathbf{a}_3$ represent vectors.

$$\begin{aligned}
\operatorname{grad}\phi &= \nabla\phi = \text{the } \textit{gradient} \text{ of } \phi \\
&= \frac{\mathbf{a}_1}{\sqrt{g_{11}}}\frac{\partial\phi}{\partial u_1} + \frac{\mathbf{a}_2}{\sqrt{g_{22}}}\frac{\partial\phi}{\partial u_2} + \frac{\mathbf{a}_3}{\sqrt{g_{33}}}\frac{\partial\phi}{\partial u_3},
\end{aligned} \tag{5.10.2}$$

$$\begin{aligned}
\operatorname{div}\mathbf{E} &= \nabla\cdot\mathbf{E} = \text{the } \textit{divergence} \text{ of } \mathbf{E} \\
&= \frac{1}{\sqrt{g}}\left\{\frac{\partial}{\partial u_1}\left(\frac{gE_1}{g_{11}}\right) + \frac{\partial}{\partial u_2}\left(\frac{gE_2}{g_{22}}\right) + \frac{\partial}{\partial u_3}\left(\frac{gE_3}{g_{33}}\right)\right\},
\end{aligned} \tag{5.10.3}$$

$$\begin{aligned}
\operatorname{curl}\mathbf{E} &= \nabla\times\mathbf{E} = \text{the } \textit{curl} \text{ of } \mathbf{E} \\
&= \mathbf{a}_1\frac{\Gamma_1}{\sqrt{g_{11}}} + \mathbf{a}_2\frac{\Gamma_2}{\sqrt{g_{22}}} + \mathbf{a}_3\frac{\Gamma_3}{\sqrt{g_{33}}}
\end{aligned} \tag{5.10.4}$$

$$= \begin{vmatrix} \frac{\mathbf{a}_1}{h_2h_3} & \frac{\mathbf{a}_2}{h_1h_3} & \frac{\mathbf{a}_3}{h_1h_2} \\ \frac{\partial}{\partial u_1} & \frac{\partial}{\partial u_2} & \frac{\partial}{\partial u_3} \\ h_1E_1 & h_2E_2 & h_3E_3 \end{vmatrix}, \tag{5.10.5}$$

$$\begin{aligned}
\left[(\mathbf{F}\cdot\nabla)\,\mathbf{E}\right]_j &= \text{the } \textit{convective operator} \\
&= \sum_{i=1}^{3}\left[\frac{F_i}{h_i}\frac{\partial E_j}{\partial u_i} + \frac{E_i}{h_ih_j}\left(F_j\frac{\partial h_j}{\partial u_i} - F_i\frac{\partial h_i}{\partial u_j}\right)\right],
\end{aligned} \tag{5.10.6}$$

$$\begin{aligned}
\nabla^2\phi &= \text{the } \textit{Laplacian} \text{ of } \phi \text{ (sometimes written as } \Delta\phi) \\
&= \frac{1}{h_1h_2h_3}\left\{\frac{\partial}{\partial u_1}\left[\frac{h_2h_3}{h_1}\frac{\partial\phi}{\partial u_1}\right] + \frac{\partial}{\partial u_2}\left[\frac{h_3h_1}{h_2}\frac{\partial\phi}{\partial u_2}\right] + \frac{\partial}{\partial u_3}\left[\frac{h_1h_2}{h_3}\frac{\partial\phi}{\partial u_3}\right]\right\} \\
&= \frac{1}{\sqrt{g}}\left\{\frac{\partial}{\partial u_1}\left[\frac{\sqrt{g}}{g_{11}}\frac{\partial\phi}{\partial u_1}\right] + \frac{\partial}{\partial u_2}\left[\frac{\sqrt{g}}{g_{22}}\frac{\partial\phi}{\partial u_2}\right] + \frac{\partial}{\partial u_3}\left[\frac{\sqrt{g}}{g_{33}}\frac{\partial\phi}{\partial u_3}\right]\right\},
\end{aligned} \tag{5.10.7}$$

$$\operatorname{grad}\operatorname{div}\mathbf{E} = \nabla(\nabla\cdot\mathbf{E}) = \frac{\mathbf{a}_1}{\sqrt{g_{11}}}\frac{\partial\Upsilon}{\partial x_1} + \frac{\mathbf{a}_2}{\sqrt{g_{22}}}\frac{\partial\Upsilon}{\partial x_2} + \frac{\mathbf{a}_3}{\sqrt{g_{33}}}\frac{\partial\Upsilon}{\partial x_3}, \tag{5.10.8}$$

$$\begin{aligned}\operatorname{curl}\operatorname{curl}\mathbf{E} &= \nabla\times(\nabla\times\mathbf{E})\\ &= \mathbf{a}_1\sqrt{\frac{g_{11}}{g}}\left[\frac{\partial\Gamma_3}{\partial x_2} - \frac{\partial\Gamma_2}{\partial x_3}\right] + \mathbf{a}_2\sqrt{\frac{g_{22}}{g}}\left[\frac{\partial\Gamma_1}{\partial x_3} - \frac{\partial\Gamma_3}{\partial x_1}\right]\\ &\quad + \mathbf{a}_3\sqrt{\frac{g_{33}}{g}}\left[\frac{\partial\Gamma_2}{\partial x_1} - \frac{\partial\Gamma_1}{\partial x_2}\right],\end{aligned} \tag{5.10.9}$$

$$\begin{aligned}\text{☼}\mathbf{E} &= \operatorname{grad}\operatorname{div}\mathbf{E} - \operatorname{curl}\operatorname{curl}\mathbf{E} = \text{the } \textit{vector Laplacian} \text{ of } \mathbf{E}\\ &= \nabla(\nabla\cdot\mathbf{E}) - \nabla\times(\nabla\times\mathbf{E})\\ &= \mathbf{a}_1\left\{\frac{1}{\sqrt{g_{11}}}\frac{\partial\Upsilon}{\partial x_1} + \sqrt{\frac{g_{11}}{g}}\left[\frac{\partial\Gamma_2}{\partial x_3} - \frac{\partial\Gamma_3}{\partial x_2}\right]\right\}\\ &\quad + \mathbf{a}_2\left\{\frac{1}{\sqrt{g_{22}}}\frac{\partial\Upsilon}{\partial x_2} + \sqrt{\frac{g_{22}}{g}}\left[\frac{\partial\Gamma_3}{\partial x_1} - \frac{\partial\Gamma_1}{\partial x_3}\right]\right\}\\ &\quad + \mathbf{a}_3\left\{\frac{1}{\sqrt{g_{33}}}\frac{\partial\Upsilon}{\partial x_3} + \sqrt{\frac{g_{33}}{g}}\left[\frac{\partial\Gamma_1}{\partial x_2} - \frac{\partial\Gamma_2}{\partial x_1}\right]\right\},\end{aligned} \tag{5.10.10}$$

where Υ and $\boldsymbol{\Gamma} = (\Gamma_1, \Gamma_2, \Gamma_3)$ are defined by

$$\begin{aligned}\Upsilon &= \frac{1}{\sqrt{g}}\left\{\frac{\partial}{\partial x_1}\left[E_1\sqrt{\frac{g}{g_{11}}}\right] + \frac{\partial}{\partial x_2}\left[E_2\sqrt{\frac{g}{g_{22}}}\right] + \frac{\partial}{\partial x_3}\left[E_3\sqrt{\frac{g}{g_{33}}}\right]\right\},\\ \Gamma_1 &= \frac{g_{11}}{\sqrt{g}}\left\{\frac{\partial}{\partial x_2}\left(\sqrt{g_{33}}E_3\right) - \frac{\partial}{\partial x_3}\left(\sqrt{g_{22}}E_2\right)\right\},\\ \Gamma_2 &= \frac{g_{22}}{\sqrt{g}}\left\{\frac{\partial}{\partial x_3}\left(\sqrt{g_{11}}E_1\right) - \frac{\partial}{\partial x_1}\left(\sqrt{g_{33}}E_3\right)\right\},\\ \Gamma_3 &= \frac{g_{22}}{\sqrt{g}}\left\{\frac{\partial}{\partial x_1}\left(\sqrt{g_{22}}E_2\right) - \frac{\partial}{\partial x_2}\left(\sqrt{g_{11}}E_1\right)\right\}.\end{aligned} \tag{5.10.11}$$

5.10.1 TABLE OF ORTHOGONAL COORDINATE SYSTEMS

The $\{f_i\}$ listed below are the separated components of the Laplace or Helmholtz equations (see (5.7.16) and (5.7.17)). See Moon and Spencer for details.

The eleven coordinate systems listed below are the only coordinate systems in Euclidean space (up to an isometry) for which the Laplace or Helmholtz equation may be solved by separation of variables.

The corresponding equations for the four separable coordinate systems which exist in two-dimensions may be obtained from the coordinate systems 1–4 listed below by suppressing the z coordinate and assuming that all functions depend only on the x and y coordinates.

1. **Rectangular coordinates** $\{x, y, z\}$
   Ranges: $-\infty < x < \infty$, $-\infty < y < \infty$, $-\infty < z < \infty$.
   $g_{11} = g_{22} = g_{33} = \sqrt{g} = 1$,
   $f_1 = f_2 = f_3 = 1$.
   In this coordinate system the following notation is sometimes used:
   $\mathbf{i} = \mathbf{a}_x$, $\mathbf{j} = \mathbf{a}_y$, $\mathbf{k} = \mathbf{a}_z$.

$$
\begin{aligned}
\operatorname{grad} f &= \mathbf{a}_x \frac{\partial f}{\partial x} + \mathbf{a}_y \frac{\partial f}{\partial y} + \mathbf{a}_z \frac{\partial f}{\partial z} \\
\operatorname{div} \mathbf{E} &= \frac{\partial}{\partial x}(E_x) + \frac{\partial}{\partial y}(E_y) + \frac{\partial}{\partial z}(E_z) \\
\operatorname{curl} \mathbf{E} &= \left(\frac{\partial E_z}{\partial y} - \frac{\partial E_y}{\partial z}\right)\mathbf{a}_x + \left(\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x}\right)\mathbf{a}_y + \left(\frac{\partial E_y}{\partial x} - \frac{\partial E_x}{\partial y}\right)\mathbf{a}_z \\
\nabla^2 f &= \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} \qquad (5.10.12) \\
\left[(\mathbf{F} \cdot \nabla)\, \mathbf{E}\right]_x &= F_x \frac{\partial E_x}{\partial x} + F_y \frac{\partial E_x}{\partial y} + F_z \frac{\partial E_x}{\partial z}
\end{aligned}
$$

2. **Circular cylinder coordinates** $\{r, \theta, z\}$
   Relations: $x = r\cos\theta$, $y = r\sin\theta$, $z = z$.
   Ranges: $0 \le r < \infty$, $0 \le \theta < 2\pi$, $-\infty < z < \infty$.
   $g_{11} = g_{33} = 1$, $g_{22} = r^2$, $\sqrt{g} = r$,
   $f_1 = r$, $f_2 = f_3 = 1$.

$$
\begin{aligned}
\operatorname{grad} f &= \mathbf{a}_r \frac{\partial f}{\partial r} + \frac{\mathbf{a}_\theta}{r}\frac{\partial f}{\partial \theta} + \mathbf{a}_z \frac{\partial f}{\partial z} \\
\operatorname{div} \mathbf{E} &= \frac{1}{r}\frac{\partial}{\partial r}(rE_r) + \frac{1}{r}\frac{\partial E_\theta}{\partial \theta} + \frac{\partial E_z}{\partial z} \\
(\operatorname{curl} \mathbf{E})_r &= \frac{1}{r}\frac{\partial E_z}{\partial \theta} - \frac{\partial E_\theta}{\partial z} \\
(\operatorname{curl} \mathbf{E})_\theta &= \frac{\partial E_r}{\partial z} - \frac{\partial E_z}{\partial r} \qquad (5.10.13) \\
(\operatorname{curl} \mathbf{E})_z &= \frac{1}{r}\frac{\partial (rE_\theta)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial \theta} \\
\nabla^2 f &= \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial f}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2 f}{\partial \theta^2} + \frac{\partial^2 f}{\partial z^2}
\end{aligned}
$$

3. **Elliptic cylinder coordinates** $\{\eta, \psi, z\}$
   Relations: $x = a\cosh\eta\cos\psi$, $y = a\sinh\eta\sin\psi$, $z = z$.
   Ranges: $0 \le \eta < \infty$, $0 \le \psi < 2\pi$, $-\infty < z < \infty$.
   $g_{11} = g_{22} = a^2(\cosh^2\eta - \cos^2\psi)$, $g_{33} = 1$,
   $\sqrt{g} = a^2(\cosh^2\eta - \cos^2\psi)$,
   $f_1 = f_2 = f_3 = 1$.

4. **Parabolic cylinder coordinates** $\{\mu, \nu, z\}$
   Relations: $x = \frac{1}{2}(\mu^2 - \nu^2)$, $\quad y = \mu\nu$, $\quad z = z$.
   Ranges: $0 \le \mu < \infty$, $\quad -\infty < \nu < 2\pi$, $\quad -\infty < z < \infty$.
   $g_{11} = g_{22} = \mu^2 + \nu^2$, $\quad g_{33} = 1$, $\quad \sqrt{g} = \mu^2 + \nu^2$,
   $f_1 = f_2 = f_3 = 1$.

5. **Spherical coordinates** $\{r, \theta, \psi\}$
   Relations: $x = r \sin\theta \cos\phi$, $\quad y = r \sin\theta \sin\phi$, $\quad z = r\cos\theta$.
   Ranges: $0 \le r < \infty$, $\quad 0 \le \theta \le \pi$, $\quad 0 \le \psi < 2\pi$.
   $g_{11} = 1$, $\quad g_{22} = r^2$, $\quad g_{33} = r^2 \sin^2\theta$, $\quad \sqrt{g} = r^2 \sin\theta$,
   $f_1 = r^2$, $\quad f_2 = \sin\theta$, $\quad f_3 = 1$.

$$
\begin{aligned}
\operatorname{grad} f &= \mathbf{e}_r \frac{\partial f}{\partial r} + \frac{\mathbf{e}_\theta}{r}\frac{\partial f}{\partial \theta} + \frac{\mathbf{e}_\phi}{r\sin\theta}\frac{\partial f}{\partial \phi} \\
\operatorname{div} \mathbf{E} &= \frac{1}{r^2}\frac{\partial}{\partial r}(r^2 E_r) + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(E_\theta \sin\theta) + \frac{1}{r\sin\theta}\frac{\partial E_\phi}{\partial\phi} \\
(\operatorname{curl}\mathbf{E})_r &= \frac{1}{r\sin\theta}\left[\frac{\partial}{\partial\theta}(E_\phi\sin\theta) - \frac{\partial A_\theta}{\partial\phi}\right] \qquad (5.10.14)\\
(\operatorname{curl}\mathbf{E})_\theta &= \frac{1}{r\sin\theta}\frac{\partial E_r}{\partial\phi} - \frac{1}{r}\frac{\partial(rE_\phi)}{\partial r} \\
(\operatorname{curl}\mathbf{E})_\phi &= \frac{1}{r}\frac{\partial(rE_\theta)}{\partial r} - \frac{1}{r}\frac{\partial E_r}{\partial\theta}
\end{aligned}
$$

$$
\nabla^2 f = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial f}{\partial r}\right) + \frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial f}{\partial\theta}\right) + \frac{1}{r^2\sin^2\theta}\frac{\partial^2 f}{\partial\phi^2}
$$

6. **Prolate spheroidal coordinates** $\{\eta, \theta, \psi\}$
   Relations: $x = a\sinh\eta\sin\theta\cos\psi$, $\quad y = a\sinh\eta\sin\theta\sin\psi$,
   $z = a\cosh\eta\cos\theta$.
   Ranges: $0 \le \eta < \infty$, $\quad 0 \le \theta \le \pi$, $\quad 0 \le \psi < 2\pi$.
   $g_{11} = g_{22} = a^2(\sinh^2\eta + \sin^2\theta)$, $\quad g_{33} = a^2\sinh^2\eta\sin^2\theta$,
   $\sqrt{g} = a^3(\sinh^2\eta + \sin^2\theta)\sinh\eta\sin\theta$.
   $f_1 = \sinh\eta$, $\quad f_2 = \sin\theta$, $\quad f_3 = a$.

7. **Oblate spheroidal coordinates** $\{\eta, \theta, \psi\}$
   Relations: $x = a\cosh\eta\sin\theta\cos\psi$, $\quad y = a\cosh\eta\sin\theta\sin\psi$,
   $z = a\sinh\eta\cos\theta$.
   Ranges: $0 \le \eta < \infty$, $\quad 0 \le \theta \le \pi$, $\quad 0 \le \psi < 2\pi$.
   $g_{11} = g_{22} = a^2(\cosh^2\eta - \sin^2\theta)$, $\quad g_{33} = a^2\cosh^2\eta\sin^2\theta$,
   $\sqrt{g} = a^3(\cosh^2\eta - \sin^2\theta)\cosh\eta\sin\theta$.
   $f_1 = \cosh\eta$, $\quad f_2 = \sin\theta$, $\quad f_3 = a$.

8. **Parabolic coordinates** $\{\mu, \nu, \psi\}$
   Relations: $x = \mu\nu\cos\psi, \qquad y = \mu\nu\sin\psi, \qquad z = \frac{1}{2}(\mu^2 - \nu^2).$
   Ranges: $0 \le \mu < \infty, \qquad 0 \le \nu \le \infty, \qquad 0 \le \psi < 2\pi.$
   $g_{11} = g_{22} = \mu^2 + \nu^2, \qquad g_{33} = \mu^2\nu^2, \qquad \sqrt{g} = \mu\nu(\mu^2 + \nu^2).$
   $f_1 = \mu, \qquad f_2 = \nu, \qquad f_3 = 1.$

9. **Conical coordinates** $\{r, \theta, \lambda\}$
   Relations: $x^2 = (r\theta\lambda/bc)^2, \qquad y^2 = r^2(\theta^2 - b^2)(b^2 - \lambda^2)/[b^2(c^2 - b^2)],$
   $z^2 = r^2(c^2 - \theta^2)(c^2 - \lambda^2)/[c^2(c^2 - b^2)].$
   Ranges: $0 \le r < \infty, \qquad b^2 < \theta^2 < c^2, \qquad 0 < \lambda^2 < b^2.$
   $g_{11} = 1, \qquad g_{22} = r^2(\theta^2 - \lambda^2)/((\theta^2 - b^2)(c^2 - \theta^2)),$
   $g_{33} = r^2(\theta^2 - \lambda^2)/((b^2 - \lambda^2)(c^2 - \lambda^2)),$
   $\sqrt{g} = r^2(\theta^2 - \lambda^2)/\sqrt{(\theta^2 - b^2)(c^2 - \theta^2)(b^2 - \lambda^2)(c^2 - \lambda^2)}.$
   $f_1 = r^2, \quad f_2 = \sqrt{(\theta^2 - b^2)(c^2 - \theta^2)}, \quad f_3 = \sqrt{(b^2 - \lambda^2)(c^2 - \lambda^2)}.$

10. **Ellipsoidal coordinates** $\{\eta, \theta, \lambda\}$
    Relations: $x^2 = (\eta\theta\lambda/bc)^2,$
    $y^2 = (\eta^2 - b^2)(\theta^2 - b^2)(b^2 - \lambda^2)/[b^2(c^2 - b^2)],$
    $z^2 = (\eta^2 - c^2)(c^2 - \theta^2)(c^2 - \lambda^2)/[c^2(c^2 - b^2)].$
    Ranges: $c^2 \le \eta^2 < \infty, \qquad b^2 < \theta^2 < c^2, \qquad 0 < \lambda^2 < b^2.$
    $g_{11} = (\eta^2 - \theta^2)(\eta^2 - \lambda^2)/((\eta^2 - b^2)(\eta^2 - c^2)),$
    $g_{22} = (\theta^2 - \lambda^2)(\eta^2 - \theta^2)/((\theta^2 - b^2)(c^2 - \theta^2)),$
    $g_{33} = (\eta^2 - \lambda^2)(\theta^2 - \lambda^2)/((b^2 - \lambda^2)(c^2 - \lambda^2)),$
    $\sqrt{g} = \frac{(\eta^2-\theta^2)(\eta^2-\lambda^2)(\theta^2-\lambda^2)}{\sqrt{(\eta^2-b^2)(\eta^2-c^2)(\theta^2-b^2)(c^2-\theta^2)(b^2-\lambda^2)(c^2-\lambda^2)}}.$
    $f_1 = \sqrt{(\eta^2 - b^2)(\eta^2 - c^2)}, \qquad f_2 = \sqrt{(\theta^2 - b^2)(c^2 - \theta^2)},$
    $f_3 = \sqrt{(b^2 - \lambda^2)(c^2 - \lambda^2)}.$

11. **Paraboloidal coordinates** $\{\mu, \nu, \lambda\}$
    Relations: $x^2 = 4(\mu - b)(b - \nu)(b - \lambda)/(b - c),$
    $y^2 = 4(\mu - c)(c - \nu)(\lambda - c)/(b - c), \qquad z = \mu + \nu + \lambda - b - c.$
    Ranges: $b < \mu < \infty, \qquad 0 < \nu < c, \qquad c < \lambda < b.$
    $g_{11} = (\mu - \nu)(\mu - \lambda)/((\mu - b)(\mu - c)),$
    $g_{22} = (\mu - \nu)(\lambda - \nu)/((b - \nu)(c - \nu)),$
    $g_{33} = (\lambda - \nu)(\mu - \lambda)/((b - \lambda)(\lambda - c)),$
    $\sqrt{g} = \frac{(\mu-\nu)(\mu-\lambda)(\lambda-\nu)}{\sqrt{(\mu-b)(\mu-c)(b-\nu)(c-\nu)(b-\lambda)(\lambda-c)}}.$
    $f_1 = \sqrt{(\mu - b)(\mu - c)}, \quad f_2 = \sqrt{(b - \nu)(c - \nu)}, \quad f_3 = \sqrt{(b - \lambda)(\lambda - c)}$

5.11 REAL ANALYSIS

5.11.1 RELATIONS

For two sets A and B, the *product* $A \times B$ is the set of all ordered pairs (a, b) where a is in A and b is in B. Any subset of the product $A \times B$ is called a *relation*. A relation R on a product $A \times A$ is called an *equivalence relation* if the following three properties hold:

1. *Reflexive*: (a, a) is in R for every a in A.
2. *Symmetric*: If (a, b) is in R, then (b, a) is in R.
3. *Transitive*: If (a, b) and (b, c) are in R, then (a, c) is in R.

When R is an equivalence relation then the *equivalence class* of an element a in A is the set of all b in A such that (a, b) is in R.

1. If $|A| = n$, there are 2^{n^2} relations on A.
2. If $|A| = n$, the number of equivalence relations on A is the Bell number B_n.

Set	Relation	Reflexive	Symmetric	Transitive
any nonempty set	$=$	yes	yes	yes
$\{1, 2, 3\}$	$\{(1, 1), (2, 2), (3, 3), (1, 2), (2, 1), (2, 3), (3, 2)\}$	yes	yes	no
real numbers	$\leq$	yes	no	yes
$\{1, 2, 3\}$	$\{(1, 1), (2, 2), (1, 2), (2, 1)\}$	no	yes	yes
any nonempty set	$\neq$	no	yes	no
real numbers	$<$	no	no	yes

EXAMPLE The set of rational numbers has an equivalence relation "=" defined by the requirement that an ordered pair $(\frac{a}{b}, \frac{c}{d})$ belongs in the relation if and only if $ad = bc$. The equivalence class of $\frac{1}{2}$ is the set $\{\frac{1}{2}, \frac{2}{4}, \frac{3}{6}, \dots, \frac{-1}{-2}, \frac{-2}{-4}, \dots\}$.

5.11.2 FUNCTIONS (MAPPINGS)

A relation f on a set $X \times Y$ is a *function* (or *mapping*) from X into Y if (x, y) and (x, z) in the relation implies that $y = z$, and each $x \in X$ has a $y \in Y$ such that (x, y) is in the relation. The last condition means that there is a unique pair in f whose first element is x. We write $f(x) = y$ to mean that (x, y) is in the relation f, and emphasize the idea of mapping by the notation $f\colon X \to Y$. The *domain* of f is the set X. The *range* of a function f is a set containing all the y for which there is a pair (x, y) in the relation. The *image* of a set A in the domain of a function f is the set of y in Y such that $y = f(x)$ for some x in A. The notation for the image of A

under f is $f[A]$. The *inverse image* of a set B in the range of a function f is the set of all x in X such that $f(x) = y$ for some y in B. The notation is $f^{-1}[B]$.

A function f is *one-to-one* (or *univalent*, or *injective*) if $f(x_1) = f(x_2)$ implies $x_1 = x_2$. A function $f\colon X \to Y$ is *onto* (or *surjective*) if for every y in Y there is some x in X such that $f(x) = y$. A function is *bijective* if it is both one-to-one and onto.

EXAMPLES

1. $f(x) = e^x$, as a mapping from $\mathbb{R}$ to $\mathbb{R}$, is one-to-one because $e^{x_1} = e^{x_2}$ implies $x_1 = x_2$ (by taking the natural logarithm). It is not onto because -1 is not the value of e^x for any x in $\mathbb{R}$.
2. $g(x) = x^3 - x$, as a mapping from $\mathbb{R}$ to $\mathbb{R}$, is onto because every real number is attained as a value of $g(x)$, for some x. It is not one-to-one because $g(-1) = g(0) = g(1)$.
3. $h(x) = x^3$, as a mapping from $\mathbb{R}$ to $\mathbb{R}$, is bijective.

For an injective function f mapping X into Y, there is an *inverse function* f^{-1} mapping the range of f into X which is defined by: $f^{-1}(y) = x$ if and only if $f(x) = y$.

EXAMPLE The function $f(x) = e^x$ mapping $\mathbb{R}$ into $\mathbb{R}^+$ (the set of positive reals) is bijective. Its inverse is $f^{-1}(x) = \ln(x)$ which maps $\mathbb{R}^+$ into $\mathbb{R}$.

For functions $f\colon X \to Y$ and $g\colon Y \to Z$, with the range of f contained in the domain of g, the *composition* $(g \circ f)\colon X \to Z$ is a function defined by $(g \circ f)(x) = g(f(x))$ for all x in the domain of f.

1. Note that $g \circ f$ may not be the same as $f \circ g$. For example, for $f(x) = x + 1$, and $g(x) = 2x$, we have $(g \circ f)(x) = g(f(x)) = 2f(x) = 2(x+1) = 2x+2$. However $(f \circ g)(x) = f(g(x)) = g(x) + 1 = 2x + 1$.
2. For every function f and its inverse f^{-1}, we have $(f \circ f^{-1})(x) = x$, for all x in the domain of f^{-1}, and $(f^{-1} \circ f)(x) = x$ for all x in the domain of f. (Note that the inverse function, f^{-1}, does not mean $\frac{1}{f}$).

5.11.3 SETS OF REAL NUMBERS

A *sequence* is the range of a function having the natural numbers as its domain. It can be denoted by $\{x_n \mid n \text{ is a natural number}\}$ or simply $\{x_n\}$. For a chosen natural number N, a *finite* sequence is the range of a function having natural numbers less than N as its domain. Sets A and B are in a *one-to-one correspondence* if there is a bijective function from A into B. Two sets A and B have the same cardinality if there is a one-to-one correspondence between them. A set which is equivalent to the set of natural numbers is *denumerable* (or *countably infinite)*. A set which is empty or is equivalent to a finite sequence is *finite* (or *finite countable*).

EXAMPLES The set of letters in the English alphabet is finite. The set of rational numbers is denumerable. The set of real numbers is uncountable.

5.11.3.1 Axioms of order

1. There is a subset P (positive numbers) of $\mathbb{R}$ for which $x + y$ and xy are in P for every x and y in P.
2. Exactly one of the following conditions can be satisfied by a number x in $\mathbb{R}$ (*trichotomy*): $x \in P$, $-x \in P$, or $x = 0$.

5.11.3.2 Definitions

A number b is an *upper* (or *lower*) *bound* of a subset S in $\mathbb{R}$ if $x \le b$ (or $x \ge b$) for every x in S. A number c is a *least upper bound* (*lub*, *supremum*, or *sup*) of a subset S in $\mathbb{R}$ if c is an upper bound of S and $b \ge c$ for every upper bound b of S. A number c is a *greatest lower bound* (*glb*, *infimum*, or *inf*) if c is a lower bound of S and $c \ge b$ for every lower bound b of S.

5.11.3.3 Completeness (or least upper bound) axiom

If a non-empty set of real numbers has an upper bound, then it has a least upper bound.

5.11.3.4 Characterization of the real numbers

The set of real numbers is the smallest complete ordered field that contains the rationals. Alternatively, the properties of a field, the order properties, and the least upper bound axiom characterize the set of real numbers. The least upper bound axiom distinguishes the set of real numbers from other ordered fields.

Archimedean property of $\mathbb{R}$: For every real number x, there is an integer N such that $x < N$. For every pair of real numbers x and y with $x < y$, there is a rational number r such that $x < r < y$. This is sometimes stated: The set of rational numbers is dense in $\mathbb{R}$.

5.11.3.5 Inequalities among real numbers

The expression $a > b$ means that $a - b$ is a positive real number.

1. If $a < b$ and $b < c$ then $a < c$.
2. If $a < b$ then $a \pm c < b \pm c$ for any real number c.
3. If $a < b$ and $\begin{cases} \text{if } c > 0 \text{ then } ac < bc \\ \text{if } c < 0 \text{ then } ac > bc \end{cases}$
4. If $a < b$ and $c < d$ then $a + c < b + d$.
5. If $0 < a < b$ and $0 < c < d$ then $ac < bd$.
6. If $a < b$ and $\left\{ \begin{matrix} ab > 0 \\ ab < 0 \end{matrix} \right\}$ then $\left\{ \begin{matrix} \dfrac{1}{a} > \dfrac{1}{b} \\ \dfrac{1}{a} < \dfrac{1}{b} \end{matrix} \right\}$

5.11.4 TOPOLOGICAL SPACE

A *topology* on a set X is a collection T of subsets of X (called *open sets*) having the following properties:

1. The empty set and X are in T.
2. The union of elements in an arbitrary subcollection of T is in T.
3. The intersection of elements in a finite subcollection of T is in T.

The complement of an open set is a *closed set*. A set is *compact* if every open cover has a finite subcover. The set X together with a topology T is a *topological space*.

5.11.4.1 Notes

1. A subset E of X is closed if and only if E contains all its limit points.
2. The union of finitely many closed sets is closed.
3. The intersection of an arbitrary collection of closed sets is closed.
4. The image of a compact set under a continuous function is compact.

5.11.5 METRIC SPACE

A *metric* (or *distance function*) on a set E is a function $\rho\colon E \times E \to \mathbb{R}$ that satisfies the following conditions:

1. *Positive definiteness*: $\rho(x, y) \geq 0$ for all x, y in E, and $\rho(x, y) = 0$ if and only if $x = y$.
2. *Symmetry*: $\rho(x, y) = \rho(y, x)$ for all x, y in E.
3. *Triangle inequality*: $\rho(x, y) \leq \rho(x, z) + \rho(z, y)$ for all x, y, z in E.

EXAMPLE

The set of real numbers with distance defined by $d(x, y) = |x - y|$ is a metric space.

A δ *neighborhood* of a point x in a metric space E is the set of all y in E such that $d(x, y) < \delta$. For example, a δ neighborhood of x in $\mathbb{R}$ is the interval centered at x with radius δ, $(x - \delta, x + \delta)$. In a metric space the topology is generated by the δ neighborhoods.

1. A subset G of $\mathbb{R}$ is *open* if, for every x in G, there is a δ neighborhood of x which is a subset of G. For example, intervals (a,b), (a,∞), $(-\infty,b)$ are open in $\mathbb{R}$.
2. A number x is a *limit point* (or a *point of closure,* or an *accumulation point)* of a set F if, for every $\delta > 0$, there is a point y in F, with $y \neq x$, such that $|x - y| < \delta$.
3. A subset F of $\mathbb{R}$ is *closed* if it contains all of its limit points. For example, intervals $[a, b]$, $(-\infty, b]$, and $[a, \infty)$ are closed in $\mathbb{R}$.
4. A subset F is *dense* in $\mathbb{R}$ if every element of $\mathbb{R}$ is a limit point of F.

5. A metric space is *separable* if it contains a denumerable dense set. For example, $\mathbb{R}$ is separable because the subset of rationals is a denumerable dense set.
6. Theorems:

   Bolzano–Weierstrass theorem Any bounded infinite set of real numbers has a limit point in $\mathbb{R}$.

   Heine–Borel theorem A subset of R is compact if and only if it is closed and bounded.

5.11.6 CONVERGENCE IN $\mathbb{R}$ WITH METRIC $|x - y|$

5.11.6.1 Limit of a sequence

A number L is a *limit point* of a sequence $\{x_n\}$ if, for every $\epsilon > 0$, there is a natural number N such that $|x_n - L| < \epsilon$ for all $n > N$. If it exists, a limit of a sequence is unique. A sequence is said to *converge* if it has a limit. A number L is a *cluster point* of a sequence $\{x_n\}$ if, for every $\epsilon > 0$ and every index N, there is an $n > N$ such that $|x_n - L| < \epsilon$.

EXAMPLE The limit of a sequence is a cluster point, as in $\{\frac{1}{n}\}$, which converges to 0. However, cluster points are not necessarily limits, as in $\{(-1)^n\}$, which has cluster points $+1$ and -1 but no limit.

Let $\{x_n\}$ be a sequence. A number L is the *limit superior (limsup)* if, for every $\epsilon > 0$, there is a natural number N such that $x_n > L - \epsilon$ for infinitely many $n \geq N$, and $x_n > L + \epsilon$ for only finitely many terms. An equivalent definition of the limit superior is given by

$$\limsup x_n = \inf_N \sup_{k \geq N} x_k. \tag{5.11.1}$$

The *limit inferior (liminf)* is defined in a similar way by

$$\liminf x_n = \sup_N \inf_{k \geq N} x_k. \tag{5.11.2}$$

For example, the sequence $\{x_n\}$ with $x_n = 1 + (-1)^n + \frac{1}{2^n}$ has $\limsup x_n = 2$, and $\liminf x_n = 0$.

Theorem Every bounded sequence $\{x_n\}$ in $\mathbb{R}$ has a $\limsup$ and a $\liminf$. In addition, if $\limsup x_n = \liminf x_n$, then the sequence converges to their common value.

A sequence $\{x_n\}$ is a *Cauchy* sequence if, for any $\epsilon > 0$, there exists a positive integer N such that $|x_n - x_m| < \epsilon$ for every $n > N$ and $m > N$.

Theorem A sequence $\{x_n\}$ in $\mathbb{R}$ converges if and only if it is a Cauchy sequence.

A metric space in which every Cauchy sequence converges to a point in the space is called *complete*. For example, $\mathbb{R}$ with the metric $d(x, y) = |x - y|$ is complete.

5.11.6.2 Limit of a function

A number L is a *limit* of a function f as x approaches a number a if, for every $\epsilon > 0$, there is a $\delta > 0$ such that $|f(x) - L| < \epsilon$ for all x with $0 < |x - a| < \delta$. This is represented by the notation $\lim_{x\to a} f(x) = L$. The symbol ∞ is the limit of a function f as x approaches a number a if, for every positive number M, there is a $\delta > 0$ such that $f(x) > M$ for all x with $0 < |x - a| < \delta$. The notation is $\lim_{x\to a} f(x) = \infty$. A number L is a limit of a function f as x approaches ∞ if, for every $\epsilon > 0$, there is a positive number M such that $|f(x) - L| < \epsilon$ for all $x > M$; this is written $\lim_{x\to\infty} f(x) = L$. The number L is said to be the *limit at infinity*.

EXAMPLES $\lim_{x\to 2} 3x - 1 = 5, \qquad \lim_{x\to 0} \frac{1}{x^2} = \infty, \qquad \lim_{x\to\infty} \frac{1}{x} = 0.$

5.11.6.3 Limit of a sequence of functions

A sequence of functions $\{f_n(x)\}$ is said to *converge pointwise* to the function $f(x)$ on a set E if for every $\epsilon > 0$ and $x \in E$ there is a positive integer N such that $|f(x) - f_n(x)| < \epsilon$ for every $n \geq N$. A sequence of functions $\{f_n(x)\}$ is said to *converge uniformly* to the function f on a set E if, for every $\epsilon > 0$, there exists a positive integer N such that $|f(x) - f_n(x)| < \epsilon$ for all x in E and $n \geq N$.

Note that these formulations of convergence are not equivalent. For example, the functions $f_n(x) = x^n$ on the interval $[0, 1]$ converge pointwise to the function $f(x) = 0$ for $0 \leq x < 1$, $f(1) = 1$. They do not converge uniformly because, for $\epsilon = 1/2$, there is no N such that $|f_n(x) - f(x)| < 1/2$ for all x in $[0, 1]$ and every $n \geq N$.

A function f is *Lipschitz* if there exists $k > 0$ in $\mathbb{R}$ such that $|f(x) - f(y)| \leq k|x - y|$ for all x and y in its domain. The function is a *contraction* if $0 < k < 1$.

> **Fixed point or contraction mapping theorem** Let E be a complete metric space. If the function $f: E \to E$ is a contraction, then there is a unique point x in E such that $f(x) = x$. The point x is called a *fixed point* of f.

EXAMPLE Newton's method for finding a zero of $f(x) = (x + 1)^2 - 2$ on the interval $[0, 1]$ produces $x_{n+1} = g(x_n)$ with the contraction $g(x) = \frac{x}{2} - \frac{1}{2} + \frac{1}{x+1}$. This has the unique fixed point $\sqrt{2} - 1$ in $[0, 1]$.

5.11.7 CONTINUITY IN $\mathbb{R}$ WITH METRIC $|x - y|$

A function $f: \mathbb{R} \to \mathbb{R}$ is *continuous at a point* a if f is defined at a and $\lim_{x\to a} f(x) = f(a)$. The function f is *continuous on a set* E if it is continuous at every point of E. A function f is *uniformly continuous* on a set E if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|f(x) - f(y)| < \epsilon$ for every x and y in E with $|x - y| < \delta$. A sequence $\{f_n(x)\}$ of continuous functions on the interval $[a, b]$ is *equicontinuous* if, for every $\epsilon > 0$, there exists a $\delta > 0$ such that $|f_n(x) - f_n(y)| < \epsilon$ for every n and for all x and y in $[a, b]$ with $|x - y| < \delta$.

1. A function can be continuous without being uniformly continuous. For example, the function $g(x) = \frac{1}{x}$ is continuous but not uniformly continuous on the open interval $(0, 1)$.
2. A collection of continuous functions can be bounded on a closed interval without having a uniformly convergent sub-sequence. For example, the continuous functions $f_n(x) = \frac{x^2}{x^2+(1-nx)^2}$ are each bounded by 1 in the closed interval $[0, 1]$ and for every x there is the limit: $\lim_{n\to\infty} f_n(x) = 0$. However, $f_n\left(\frac{1}{n}\right) = 1$ for every n, so that no sub-sequence can converge uniformly to 0 everywhere on $[0, 1]$. This sequence is not equicontinuous.
3. Theorems:

   Theorem Let $\{f_n(x)\}$ be a sequence of functions mapping $\mathbb{R}$ into $\mathbb{R}$ which converges uniformly to a function f. If each $f_n(x)$ is continuous at a point a, then $f(x)$ is also continuous at a.

   Theorem If a function f is continuous on a closed bounded set E, then it is uniformly continuous on E.

   Ascoli–Arzela theorem Let K be a compact set in $\mathbb{R}$. If $\{f_n(x)\}$ is uniformly bounded and equicontinuous on K, then $\{f_n(x)\}$ contains a uniformly convergent sub-sequence on K.

   Weierstrass polynomial approximation theorem Let K be a compact set in $\mathbb{R}$. If f is a continuous function on K, then there exists a sequence of polynomials that converges uniformly to f on K.

5.11.8 BANACH SPACE

A *norm* on a vector space E with scalar field $\mathbb{R}$ is a function $||\cdot||$ from E into $\mathbb{R}$ that satisfies the following conditions:

1. *Positive definiteness*: $||x|| \geq 0$ for all x in E, and $||x|| = 0$ if and only if $x = 0$.
2. *Scalar homogeneity*: For every x in E and a in $\mathbb{R}$, $||ax|| = |a|\,||x||$.
3. *Triangle inequality*: $||x + y|| \leq ||x|| + ||y||$ for all x, y in E.

Every norm $||\cdot||$ gives rise to a metric ρ by defining: $\rho(x, y) = ||x - y||$.

EXAMPLES

1. $\mathbb{R}$ with absolute value as the norm has the metric $\rho(x, y) = |x - y|$.
2. $\mathbb{R} \times \mathbb{R}$ (denoted $\mathbb{R}^2$) with the Euclidean norm $\|(x, y)\| = \sqrt{x^2 + y^2}$, has the metric $\rho((x_1, y_1), (x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$.

A *Banach space* is a complete normed space.

A widely studied example of a Banach space is the (vector) space of measurable functions f on $[a, b]$ for which $\int_a^b |f(x)|^p\,dx < \infty$ with $1 \leq p < \infty$. This is denoted by $L^p[a, b]$ or simply L^p. The space of essentially bounded measurable functions on $[a, b]$ is denoted by $L^\infty[a, b]$.

The L^p *norm* for $1 \leq p < \infty$ is defined by $\|f\|_p = \left(\int_a^b |f(x)|^p \, dx\right)^{1/p}$. The L^∞ norm is defined by

$$\|f\|_\infty = \text{ess} \sup_{a \leq x \leq b} |f(x)|, \tag{5.11.3}$$

where

$$\text{ess} \sup_{a \leq x \leq b} |f(x)| = \inf\{M | m\{t : f(t) > M\} = 0\}. \tag{5.11.4}$$

Let $\{f_n(x)\}$ be a sequence of functions in L^p (with $1 \leq p < \infty$) and f be some function in L^p. We say that $\{f_n\}$ *converges in the mean of order* p (or simply in L^p*-norm*) to f if $\lim_{n\to\infty} \|f_n - f\|_p = 0$.

Riesz–Fischer theorem The L^p spaces are complete.

5.11.8.1 Inequalities

1. *Arithmetic mean–geometric mean inequality* If A_n and G_n are the arithmetic and geometric means of the set of positive numbers $\{a_1, a_2, \ldots, a_n\}$ then $A_n \geq G_n$. That is

$$\frac{a_1 + a_2 + \ldots + a_n}{n} \geq (a_1 a_2 \cdots a_n)^{1/n}$$

2. *Carleman's inequality* If A_n and G_n are the arithmetic and geometric means of the set of positive numbers $\{a_1, a_2, \ldots, a_n\}$ then

$$\sum_{r=1}^{n} G_r \leq neA_n$$

3. *Hölder inequality* If p and q are non-negative extended real numbers such that $1/p + 1/q = 1$ and $f \in L^p$ and $g \in L^q$, then $\|fg\|_1 \leq \|f\|_p \|g\|_q$:

$$\int_a^b |fg| \leq \left(\int_a^b |f|^p\right)^{1/p} \left(\int_a^b |g|^q\right)^{1/q} \quad \text{for } 1 \leq p < \infty, \tag{5.11.5}$$

and (taking the limit as $q \to 1$ and $p \to 0$)

$$\int_a^b |fg| \leq (\text{ess}\sup |f|) \int_a^b |g|. \tag{5.11.6}$$

4. *Jensen inequality* A continuous function f on $[a, b]$ is called *convex* if the Jensen inequality holds

$$f\left(\frac{x_1 + x_2}{2}\right) \leq \frac{f(x_1) + f(x_2)}{2} \tag{5.11.7}$$

for all x_1 and x_2 in $[a, b]$.

5. *Minkowski inequality* If f and g are in L^p with $1 \leq p \leq \infty$, then $\|f+g\|_p \leq \|f\|_p + \|g\|_p$:

$$\left(\int_a^b |f+g|^p\right)^{1/p} \leq \left(\int_a^b |f|^p\right)^{1/p} + \left(\int_a^b |g|^p\right)^{1/p} \quad \text{for } 1 \leq p < \infty, \tag{5.11.8}$$

$$\operatorname{ess\,sup}\ |f+g| \leq \operatorname{ess\,sup}\ |f| + \operatorname{ess\,sup}\ |g|.$$

6. *Schwartz* (or *Cauchy–Schwartz*) *inequality* If f and g are in L^2, then $\|fg\|_1 \leq \|f\|_2 \|g\|_2$. This is the special case of Hölder's inequality with $p = q = 2$.

5.11.9 HILBERT SPACE

An *inner product* on a vector space E with scalar field $\mathbb{C}$ (complex numbers) is a function from $E \times E$ into $\mathbb{C}$ that satisfies the following conditions:

1. $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.
2. $\langle x + y, z \rangle = \langle x, z \rangle + \langle y, z \rangle$
3. $\langle cx, y \rangle = c \langle x, y \rangle$
4. $\langle x, y \rangle = \overline{\langle y, x \rangle}$

Every inner product $\langle x, y \rangle$ gives rise to a norm $\|x\|$ by defining $\|x\| = \langle x, x \rangle^{1/2}$.

A *Hilbert space* is a complete inner product space. A widely studied Hilbert space is $L^2[a,b]$ with the inner product $\langle f, g \rangle = \int_a^b f(t)\overline{g(t)}\, dt$.

Two functions f and g in $L^2[a,b]$ are *orthogonal* if $\int_a^b fg = 0$. A set of L^2 functions $\{\phi_n\}$ is *orthogonal* if $\int_a^b \phi_m \phi_n = 0$ for $m \neq n$. The set is *orthonormal* if, in addition, each member has norm 1. That is, $\|\phi_n\|_2 = 1$. For example, the functions $\{\sin nx\}$ are mutually orthogonal on $(-\pi, \pi)$. The functions $\{\frac{\sin nx}{\sqrt{\pi}}\}$ form an orthonormal set on $(-\pi, \pi)$.

Let $\{\phi_n\}$ be an orthonormal set in L^2 and f be in L^2. The numbers $c_n = \int_a^b f\phi_n\, dx$ are the *generalized Fourier coefficients* of f with respect to $\{\phi_n\}$, and the series $\sum_{n=1}^{\infty} c_n\phi_n(x)$ is called the *generalized Fourier series* of f with respect to $\{\phi_n\}$.

For a function f in L^2, the *mean square error* of approximating f by the sum $\sum_{n=1}^{N} a_n\phi_n$ is $\frac{1}{b-a}\int_a^b |f(x) - \sum_{n=1}^{N} a_n\phi_n(x)|^2\, dx$. An orthonormal set $\{\phi_n\}$ is *complete* if the only measurable function f that is orthogonal to every ϕ_n is zero. That is, $f = 0$ *a.e.* (In the context of elementary measure theory, two measurable functions f and g are *equivalent* if they are equal except on a set of measure zero. They are said to be equal *almost everywhere*. This is denoted by $f = g$ *a.e.*)

Bessel's inequality: For a function f in L^2 having generalized Fourier coefficients $\{c_n\}$, $\sum_{n=1}^{\infty} \left|c_n^2\right| \leq \int_a^b |f(x)|^2\, dx$.

1. **Riesz–Fischer theorem** Let $\{\phi_n\}$ be an orthonormal set in L^2 and let $\{c_n\}$ be constants such that $\sum_{n=1}^{\infty} \left|c_n^2\right|$ converges. Then a unique function f in L^2 exists such that the c_n are the Fourier coefficients of f with respect to $\{\phi_n\}$ and $\sum_{n=1}^{\infty} c_n \phi_n$ converges in the mean (or order 2) to f.
2. **Theorem** The generalized Fourier series of f in L^2 converges in the mean (of order 2) to f.
3. **Theorem** Parseval's identity holds: $\displaystyle\int_a^b |f(x)|^2\,dx = \sum_{n=1}^{\infty} \left|c_n^2\right|$.
4. **Theorem** The mean square error of approximating f by the series $\sum_{n=1}^{\infty} a_n \phi_n$ is minimum when all coefficients a_n are the Fourier coefficients of f with respect to $\{\phi_n\}$.

EXAMPLE Suppose that the series $\frac{a_o}{2} + \sum_{n=1}^{\infty}(|a_n|^2 + |b_n|^2)$ converges. Then the trigonometric series $\frac{a_o}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx)$ is the Fourier series of some function in L^2.

5.11.10 ASYMPTOTIC RELATIONSHIPS

Asymptotic relationships are indicated by the symbols O, Ω, Θ, o, and $\sim$.

1. The symbol O (pronounced "big-oh"): $f(x) \in O(g(x))$ as $x \to x_0$ if a positive constant C exists such that $|f(x)| \le C\,|g(x)|$ for all x sufficiently close to x_0. Note that $O(g(x))$ is a class of functions. Sometimes the statement $f(x) \in O(g(x))$ is written (imprecisely) as $f = O(g)$.
2. The symbol Ω : $f(x) \in \Omega(g(x))$ as $x \to x_0$ if a positive constant C exists such that $|g(x)| \le C\,|f(x)|$ for all x sufficiently close to x_0.
3. The symbol Θ : $f(x) \in \Theta(g(x))$ as $x \to x_0$ if positive constants c_1 and c_2 exist such that $c_1 g(x) \le f(x) \le c_2 g(x)$ for all x sufficiently close to x_0. This is equivalent to: $f(x) = O(g(x))$ and $g(x) = O(f(x))$. The symbol $\approx$ is often used for Θ (i.e., $f(x) \approx g(x)$).
4. The symbol o (pronounced "little-oh"): $f(x) \in o(g(x))$ as $x \to x_0$ if, given any $\mu > 0$, we have $|f(x)| < \mu|g(x)|$ for all x sufficiently close to x_0.
5. The symbol $\sim$ (pronounced "asymptotic to"): $f(x) \sim (g(x))$ as $x \to x_0$ if $f(x) = g(x)\,[1 + o(1)]$ as $x \to x_0$.
6. Two functions, $f(x)$ and $g(x)$, are *asymptotically equivalent* as $x \to x_0$ if $f(x)/g(x) \sim 1$ as $x \to x_0$.
7. A sequence of functions, $\{g_k(x)\}$, forms an *asymptotic series* at x_0 if $g_{k+1}(x) = o(g_k(x))$ as $x \to x_0$.
8. Given a function $f(x)$ and an asymptotic series $\{g_k(x)\}$ at x_0, the formal series $\sum_{k=0}^{\infty} a_k g_k(x)$, where the $\{a_k\}$ are given constants, is an *asymptotic expansion* of $f(x)$ if $f(x) - \sum_{k=0}^{n} a_k g_k(x) = o(g_n(x))$ as $x \to x_0$ for every n; this is expressed as $f(x) \sim \sum_{k=0}^{\infty} a_k g_k(x)$. Partial sums of this formal series are called *asymptotic approximations* to $f(x)$. This formal series need not converge.

Think of O being an upper bound on a function, Ω being a lower bound, and Θ being both an upper and lower bound. For example: $\sin x \in O(x)$ as $x \to 0$, $\log n \in o(n)$ as $n \to \infty$, and $n^9 \in \Omega(n^9 + n^2)$ as $n \to \infty$.

The statements: $n^2 = o(n^5)$, $n^5 = o(2^n)$, $2^n = o(n!)$, and $n! = o(n^n)$ as $n \to \infty$ can be illustrated as follows. If a computer can perform 10^9 operations per second, and a procedure takes $f(n)$ operations, then the following table indicates approximately how long it will take a computer to perform the procedure, for various functions and values of n.

complexity	$n = 10$	$n = 20$	$n = 50$	$n = 100$	$n = 300$
$f(n) = n^2$	10^{-7} sec	10^{-7} sec	10^{-6} sec	10^{-5} sec	10^{-4} sec
$f(n) = n^5$	10^{-4} sec	10^{-3} sec	0.3 sec	10 sec	41 minutes
$f(n) = 2^n$	10^{-6} sec	10^{-3} sec	2 weeks	10^{11} centuries	10^{72} centuries
$f(n) = n!$	10^{-3} sec	77 years	10^{46} centuries	10^{139} centuries	10^{596} centuries
$f(n) = n^n$	10 sec	10^{7} centuries	10^{66} centuries	10^{181} centuries	10^{724} centuries

5.12 GENERALIZED FUNCTIONS

5.12.1 DELTA FUNCTION

Dirac's delta function is a distribution defined by $\delta(x) = \begin{cases} 0 & x \neq 0 \\ \infty & x = 0 \end{cases}$, and is normalized so that $\int_{-\infty}^{\infty} \delta(x)\,dx = 1$. Properties include (assuming that $f(x)$ is continuous):

1. $\displaystyle\int_{-\infty}^{\infty} f(x)\delta(x-a)\,dx = f(a)$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)\frac{d^m\delta(x)}{dx^m}\,dx = (-1)^m \frac{d^m f(0)}{dx^m}$
3. $x\delta(x)$, as a distribution, equals zero.
4. $\delta(ax) = \dfrac{1}{|a|}\delta(x)$ when $a \neq 0$
5. $\delta(x^2 - a^2) = \dfrac{1}{2a}[\delta(x+a) + \delta(x-a)]$
6. $\delta(x) = \dfrac{1}{2L} + \dfrac{1}{L}\displaystyle\sum_{n=1}^{\infty} \cos\frac{n\pi x}{L}$ (Fourier series)
7. $\delta(x-\xi) = \dfrac{2}{L}\displaystyle\sum_{n=1}^{\infty} \sin\frac{n\pi\xi}{L}\sin\frac{n\pi x}{L}$ for $0 < \xi < L$ (Fourier sine series)

8. $\delta(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{ikx}\,dk$ (Fourier transform)

9. $\delta(\rho - \rho') = \rho \int_{0}^{\infty} kJ_m(k\rho)J_m(k\rho')\,dk$

Sequences of functions $\{\phi_n\}$ that approximate the delta function as $n \to \infty$ are known as delta sequences.

EXAMPLES

(a) $\phi_n(x) = \frac{n}{\pi}\frac{1}{1+n^2x^2}$

(b) $\phi_n(x) = \frac{n}{\sqrt{\pi}}e^{-n^2x^2}$

(c) $\phi_n(x) = \frac{1}{n\pi}\frac{\sin^2 nx}{x^2}$

(d) $\phi_n(x) = \begin{cases} 0 & |x| \geq 1/n \\ n/2 & |x| < 1/n \end{cases}$

The delta function $\delta(\mathbf{x} - \mathbf{x}') = \delta(x_1 - x_1')\delta(x_2 - x_2')\delta(x_3 - x_3')$ in terms of the coordinates (ξ_1, ξ_2, ξ_3), related to (x_1, x_2, x_3), via the Jacobian $J(x_i, \xi_j)$, is written

$$\delta(\mathbf{x} - \mathbf{x}') = \frac{1}{|J(x_i, \xi_j)|}\delta(\xi_1 - \xi_1')\delta(\xi_2 - \xi_2')\delta(\xi_3 - \xi_3'). \tag{5.12.1}$$

For example, in spherical polar coordinates

$$\delta(\mathbf{x} - \mathbf{x}') = \frac{1}{r^2}\delta(r - r')\delta(\phi - \phi')\delta(\cos\theta - \cos\theta'). \tag{5.12.2}$$

The solutions to differential equations involving delta functions are called Green's functions (see pages 358 and 365).

5.12.2 OTHER GENERALIZED FUNCTIONS

The Heaviside function, or step function, is defined as

$$H(x) = \int_{-\infty}^{x} \delta(t)\,dt = \begin{cases} 0 & x < 0 \\ 1 & x > 0. \end{cases} \tag{5.12.3}$$

Sometimes $H(0)$ is stated to be $1/2$. This function has the representations:

1. $H(x) = \frac{1}{2} + \frac{2}{\pi}\sum_{n=\text{odd}}^{\infty} \frac{1}{n}\sin\frac{n\pi x}{L}$ (when $|x| < L$)

2. $H(x) = \frac{1}{2\pi i}\int_{-\infty}^{\infty} \frac{e^{ikx}}{k}\,dk$ (this is a principal-value integral)

The related signum function gives the sign of its argument:

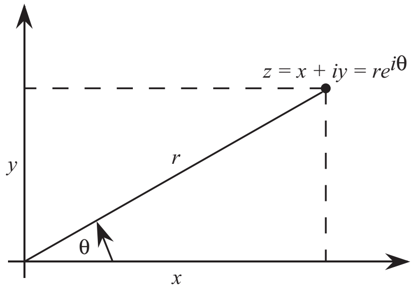
$$\operatorname{sgn}(x) = 2H(x) - 1 = \begin{cases} -1 & \text{if } x < 0 \\ 1 & \text{if } x > 0. \end{cases} \tag{5.12.4}$$

5.13 COMPLEX ANALYSIS

5.13.1 DEFINITIONS

A complex number z has the form $z = x + iy$ where x and y are real numbers, and $i = \sqrt{-1}$; the number i is sometimes called the imaginary unit. We write $x = \text{Re}\, z$ and $y = \text{Im}\, z$. The number x is called the *real* part of z and y is called the *imaginary* part of z. This form is also called the *Cartesian form* of the complex number.

Complex numbers can also be written in *polar form*, $z = re^{i\theta}$, where r, called the *modulus*, is given by $r = |z| = \sqrt{x^2 + y^2}$, and θ is called the *argument*: $\theta = \arg z = \tan^{-1} \frac{y}{x}$ (when $x > 0$). The geometric relationship between Cartesian and polar forms is shown below



The *complex conjugate* of z, denoted $\overline{z}$, is defined as $\overline{z} = x - iy = re^{-i\theta}$. Note that $|z| = |\overline{z}|$, $\arg \overline{z} = -\arg z$, and $|z| = \sqrt{z\overline{z}}$. In addition, $\overline{\overline{z}} = z$, $\overline{z_1 + z_2} = \overline{z}_1 + \overline{z}_2$, and $\overline{z_1 z_2} = \overline{z}_1 \overline{z}_2$.

5.13.2 OPERATIONS ON COMPLEX NUMBERS

1. Addition and subtraction:

$$z_1 \pm z_2 = (x_1 + iy_1) \pm (x_2 + iy_2) = (x_1 \pm x_2) + i(y_1 \pm y_2).$$

2. Multiplication:

$$z_1 z_2 = (x_1 + iy_1)(x_2 + iy_2) = (x_1 x_2 - y_1 y_2) + i(x_1 y_2 + x_2 y_1) = r_1 r_2 e^{i(\theta_1 + \theta_2)}.$$
$$|z_1 z_2| = |z_1||z_2|, \qquad \arg(z_1 z_2) = \arg z_1 + \arg z_2 = \theta_1 + \theta_2.$$

3. Division:

$$\frac{z_1}{z_2} = \frac{z_1 \overline{z}_2}{z_2 \overline{z}_2} = \frac{(x_1 x_2 + y_1 y_2) + i(x_2 y_1 - x_1 y_2)}{x_2^2 + y_2^2} = \frac{r_1}{r_2} e^{i(\theta_1 - \theta_2)}.$$
$$\left|\frac{z_1}{z_2}\right| = \frac{|z_1|}{|z_2|}, \qquad \arg\left(\frac{z_1}{z_2}\right) = \arg z_1 - \arg z_2 = \theta_1 - \theta_2.$$

4. Powers:

$$z^n = r^n e^{in\theta} = r^n(\cos n\theta + i \sin n\theta) \qquad \text{(DeMoivre's Theorem).}$$

5. Roots: $z^{1/n} = r^{1/n} e^{i(\theta+2k\pi)/n} = r^{1/n}\left(\cos\frac{\theta+2k\pi}{n} + i\sin\frac{\theta+2k\pi}{n}\right)$, for $k = 0, 1, \ldots, n-1$. The *principal root* has $-\pi < \theta \leq \pi$ and $k = 0$.

5.13.3 FUNCTIONS OF A COMPLEX VARIABLE

A complex function

$$w = f(z) = u(x, y) + iv(x, y) = |w|e^{i\phi},$$

where $z = x + iy$, associates one or more values of the complex dependent variable w with each value of the complex independent variable z for those values of z in a given domain.

A function $f(z)$ is said to be *analytic* (or *holomorphic*) at a point z_0 if $f(z)$ is defined in each point z of a disc with positive radius R around z_0, h is any complex number with $|h| < R$, and the limit of $[f(z_0+h)-f(z_0)]/h$ exists and is independent of the mode in which h tends to zero. This limiting value is the *derivative* of $f(z)$ at z_0 denoted by $f'(z_0)$. A function is called *analytic* in a connected domain if it is analytic at every point in that domain.

A function is called *entire* if it is analytic in $\mathbb{C}$.

Liouville's theorem: A bounded entire function is constant.

EXAMPLES

1. $f(z) = z^n$ is analytic everywhere when n is a non-negative integer. If n is a negative integer, then $f(z)$ is analytic except at the origin.
2. $f(z) = \overline{z}$ is nowhere analytic.
3. $f(z) = e^z$ is analytic everywhere.

5.13.4 CAUCHY–RIEMANN EQUATIONS

A necessary and sufficient condition for $f(z) = u(x, y) + iv(x, y)$ to be analytic is that it satisfies the *Cauchy–Riemann equations*:

$$\frac{\partial u}{\partial x} = \frac{\partial v}{\partial y}, \quad \text{and} \quad \frac{\partial u}{\partial y} = -\frac{\partial v}{\partial x}. \tag{5.13.1}$$

1. If these equations are satisfied, then both u and v are *harmonic functions*. That is, they satisfy $\nabla^2 u = u_{xx} + u_{yy} = 0$ and $\nabla^2 v = v_{xx} + v_{yy} = 0$.
2. The Cauchy–Riemann equations can be written as $\frac{\partial f}{\partial \overline{z}} = 0$ for $z = x + iy$.

5.13.5 DERIVATIVES IN COMPLEX COORDINATES

If $z = x + iy$ then $dz = dx + idy$, $d\overline{z} = dx - idy$, and the derivatives are

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right) \qquad \frac{\partial}{\partial \overline{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) \tag{5.13.2}$$

Using these, the differential of a complex valued function $f(z)$ is

$$df = \frac{\partial f(z)}{\partial z}\, dz + \frac{\partial f(z)}{\partial \overline{z}}\, d\overline{z} \tag{5.13.3}$$

The derivatives of the conjugate function $\overline{f}(z)$ satisfy

$$\frac{\partial \overline{f}}{\partial z} = \overline{\left(\frac{\partial f(z)}{\partial \overline{z}}\right)} \qquad \frac{\partial \overline{f}}{\partial \overline{z}} = \overline{\left(\frac{\partial f(z)}{\partial z}\right)} \tag{5.13.4}$$

In multiple dimensions

$$df = \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}^{\mathrm{T}}} d\mathbf{z} + \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}^{\mathrm{H}}} d\overline{\mathbf{z}} \tag{5.13.5}$$

where $\frac{\partial}{\partial \mathbf{z}} = \left[\frac{\partial}{\partial z_1}, \dots, \frac{\partial}{\partial z_n}\right]^{\mathrm{T}}$ and $\frac{\partial}{\partial \overline{\mathbf{z}}} = \left[\frac{\partial}{\partial \overline{z}_1}, \dots, \frac{\partial}{\partial \overline{z}_n}\right]^{\mathrm{T}}$.

EXAMPLES

1. Note the following

$$\begin{aligned} &\frac{\partial}{\partial z}\overline{z} = 0 \qquad \frac{\partial}{\partial \overline{z}} z = 0 \qquad \frac{\partial}{\partial z}|z|^2 = \frac{\partial}{\partial z} z\overline{z} = \overline{z} \\ &\frac{\partial}{\partial z}\exp\left(-|z|^2\right) = \frac{\partial}{\partial z}\exp\left(-z\overline{z}\right) = -\overline{z}\exp\left(-|z|^2\right) \end{aligned} \tag{5.13.6}$$

2. If $f(\mathbf{z}) = \mathbf{z}^{\mathrm{H}} A\mathbf{z}$ then

$$\begin{aligned} \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}} &= A^{\mathrm{T}}\overline{\mathbf{z}} \qquad & \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}^{\mathrm{T}}} &= \mathbf{z}^{\mathrm{H}} A \\ \frac{\partial f(\mathbf{z})}{\partial \overline{\mathbf{z}}} &= A\mathbf{z} \qquad & \frac{\partial f(\mathbf{z})}{\partial \mathbf{z}^{\mathrm{H}}} &= \mathbf{z}^{\mathrm{T}} A^{\mathrm{T}} \end{aligned} \tag{5.13.7}$$

5.13.6 TAYLOR SERIES EXPANSIONS

If $f(z)$ is analytic inside of and on a circle C of radius r centered at the point z_0, then a unique and uniformly convergent series expansion exists in powers of $(z - z_0)$

$$f(z) = \sum_{n=0}^{\infty} a_n (z - z_0)^n, \quad |z - z_0| < r, \quad z_0 \neq \infty, \tag{5.13.8}$$

where

$$a_n = \frac{1}{n!} f^{(n)}(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)}{(z - z_0)^{n+1}}\, dz. \tag{5.13.9}$$

If $M(r)$ is an upper bound of $|f(z)|$ on C, then

$$|a_n| = \frac{1}{n!}|f^{(n)}(z_0)| \leq \frac{M(r)}{r^n} \qquad \text{(Cauchy's inequality).} \tag{5.13.10}$$

If the series is truncated with the term $a_n(z - z_0)^n$, the remainder $R_n(z)$ is given by

$$R_n(z) = \frac{(z - z_0)^{n+1}}{2\pi i}\int_C \frac{f(s)}{(s - z)(s - z_0)^{n+1}}\, ds, \tag{5.13.11}$$

and

$$|R_n(z)| \leq \left(\frac{|z - z_0|}{r}\right)^n \frac{rM(r)}{r - |z - z_0|}. \tag{5.13.12}$$

5.13.7 LAURENT SERIES EXPANSIONS

If $f(z)$ is analytic inside the annulus between the concentric circles C_1 and C_2 centered at z_0 with radii r_1 and r_2 ($r_1 < r_2$), respectively, then a unique series expansion exists in terms of positive and negative powers of $z - z_0$ of the following form:

$$\begin{aligned} f(z) &= \sum_{n=1}^{\infty} b_n (z-z_0)^{-n} + \sum_{n=0}^{\infty} a_n (z-z_0)^n \\ &= \dots + \frac{b_2}{(z-z_0)^2} + \frac{b_1}{z-z_0} + a_0 + a_1(z-z_0) + a_2(z-z_0)^2 + \dots \end{aligned} \tag{5.13.13}$$

with (here C is a contour between C_1 and C_2)

$$\begin{aligned} a_n &= \frac{1}{2\pi i}\int_C \frac{f(s)}{(s-z_0)^{n+1}}\,ds, \quad n = 0, 1, 2, \dots, \\ b_n &= \frac{1}{2\pi i}\int_C f(s)(s-z_0)^{n-1}\,ds, \quad n = 1, 2, 3, \dots \end{aligned} \tag{5.13.14}$$

Equation (5.13.13) is often written in the form

$$\begin{aligned} f(z) &= \sum_{n=-\infty}^{\infty} c_n (z-z_0)^n \qquad \text{for } r_1 < |z-z_0| < r_2 \\ \text{with} \quad c_n &= \frac{1}{2\pi i}\int_C \frac{f(z)}{(z-z_0)^{n+1}}\,dz \qquad \text{for } n = 0, \pm 1, \pm 2, \dots. \end{aligned} \tag{5.13.15}$$

5.13.8 ZEROS AND SINGULARITIES

The points z for which $f(z) = 0$ are called *zeros* of $f(z)$. A function $f(z)$ which is analytic at z_0 has a zero of *order* m there, where m is a positive integer, if and only if the first m coefficients $a_0, a_1, \dots, a_{m-1}$ in the Taylor expansion at z_0 vanish.

A *singular point* or *singularity* of the function $f(z)$ is any point at which $f(z)$ is not analytic. An *isolated singularity* of $f(z)$ at z_0 may be classified as one of:

1. A *removable* singularity if and only if all coefficients b_n in the Laurent series expansion of $f(z)$ at z_0 vanish. This implies that $f(z)$ can be analytically extended to z_0.
2. A *pole* is of order m if and only if $(z-z_0)^m f(z)$, but not $(z-z_0)^{m-1} f(z)$, is analytic at z_0, (i.e., if and only if $b_m \neq 0$ and $0 = b_{m+1} = b_{m+2} = \dots$ in the Laurent series expansion of $f(z)$ at z_0). Equivalently, $f(z)$ has a pole of order m if $1/f(z)$ is analytic at z_0 and has a zero of order m there.
3. An isolated *essential singularity* if and only if the Laurent series expansion of $f(z)$ at z_0 has an infinite number of terms involving negative powers of $z - z_0$.

If not isolated, then a point may be a *branch point*. These are usually the result of a multi-valued function (e.g., $\sqrt{z}$ or $\log z$). A function can be made single valued within a domain by introducing a cut (a line or curve) which allows discontinuous values of the function on opposite sides of the cut.

Riemann removable singularity theorem: Suppose that a function f is analytic and bounded in some deleted neighborhood $0 < |z - z_0| < \epsilon$ of a point z_0. If f is not analytic at z_0, then it has a removable singularity there.

Casorati–Weierstrass theorem: Suppose that z_0 is an essential singularity of a function f, and let w be an arbitrary complex number. Then, for any $\epsilon > 0$, the inequality $|f(z) - w| < \epsilon$ is satisfied at some point z in each deleted neighborhood $0 < |z - z_0| < \delta$ of z_0.

5.13.9 RESIDUES

Given a point z_0 where $f(z)$ is either analytic or has an isolated singularity, the *residue* of $f(z)$ is the coefficient of $(z - z_0)^{-1}$ in the Laurent series expansion of $f(z)$ at z_0, or

$$\operatorname{Res}(z_0) = b_1 = \frac{1}{2\pi i} \int_C f(z)\, dz. \tag{5.13.16}$$

If $f(z)$ is either analytic or has a removable singularity at z_0, then $b_1 = 0$ there. If z_0 is a pole of order m, then

$$b_1 = \frac{1}{(m-1)!} \frac{d^{m-1}}{dz^{m-1}} [(z - z_0)^m f(z)] \bigg|_{z=z_0}. \tag{5.13.17}$$

For every simple closed contour C enclosing at most a finite number of singularities $z_1, z_2, \ldots, z_n$ of an function analytic in a neighborhood of C,

$$\int_C f(z)\, dz = 2\pi i \sum_{k=1}^{n} \operatorname{Res}(z_k), \tag{5.13.18}$$

where $\operatorname{Res}(z_k)$ is the residue of $f(z)$ at z_k.

5.13.10 THE ARGUMENT PRINCIPLE

Let $f(z)$ be analytic on a simple closed curve C with no zeros on C and analytic everywhere inside C except possibly at a finite number of poles. Let $\Delta_C \arg f(z)$ denote the change in the argument of $f(z)$ (that is, final value $-$ initial value) as z transverses the curve once in the positive sense. Then

$$\frac{1}{2\pi} \Delta_C \arg f(z) = \frac{1}{2\pi i} \int_C \frac{f'(z)}{f(z)}\, dz = N - P, \tag{5.13.19}$$

where N is number of zeros of $f(z)$ inside C, and P is the number of poles inside C. The zeros and poles are counted according to their multiplicities.

5.13.11 CAUCHY INTEGRAL THEOREM AND FORMULA

Cauchy integral theorem

If $f(z)$ is analytic at all points within and on a simple closed curve C, then

$$\int_C f(z)\,dz = 0. \tag{5.13.20}$$

Cauchy integral formula

If $f(z)$ is analytic inside and on a simple closed contour C and if z_0 is interior to C, then

$$f(z_0) = \frac{1}{2\pi i}\int_C \frac{f(z)}{z - z_0}\,dz. \tag{5.13.21}$$

Moreover, since the derivatives $f'(z_0)$, $f''(z_0)$, ... of all orders exist, then

$$f^{(n)}(z_0) = \frac{n!}{2\pi i}\int_C \frac{f(z)}{(z - z_0)^{n+1}}\,dz. \tag{5.13.22}$$

5.13.12 TRANSFORMATIONS AND MAPPINGS

A function $w = f(z) = u(z) + iv(z)$ maps points of the z-plane into corresponding points of the w-plane. At every point z such that $f(z)$ is analytic and $f'(z) \neq 0$, the mapping is *conformal*, i.e., the angle between two curves in the z-plane through such a point is equal in magnitude and sense to the angle between the corresponding curves in the w-plane. A table giving real and imaginary parts, zeros, and singularities for frequently used functions of a complex variable and a table illustrating a number of special transformations of interest are at the end of this section.

A function is said to be *simple* in a domain D if it is analytic in D and assumes no value more than once in D.

Riemann's mapping theorem: If D is a simply connected domain in the complex z plane, which is neither the z plane nor the extended z plane, then there is a simple function $f(z)$ such that $w = f(z)$ maps D onto the disc $|w| < 1$.

5.13.12.1 Bilinear transformations

The *bilinear* transformation (also called a *linear fractional transformation*) is defined by $w = \dfrac{az + b}{cz + d}$, where a, b, c, and d are complex numbers and $ad \neq bc$. The bilinear transformation is defined for all $z \neq -d/c$. The bilinear transformation is conformal and maps circles and lines onto circles and lines.

The inverse transformation is given by $z = \dfrac{-dw + b}{cw - a}$, which is also a bilinear transformation. Note that $w \neq a/c$.

1. The *cross-ratio* of four distinct complex numbers z_k (for $k = 1, 2, 3, 4$) is given by

$$(z_1, z_2, z_3, z_4) = \frac{(z_1 - z_2)(z_3 - z_4)}{(z_1 - z_4)(z_3 - z_2)}.$$

If any of the z_k's is complex infinity, the cross-ratio is redefined so that the quotient of the two terms on the right containing z_k is equal to 1. Under the bilinear transformation, the cross-ratio of four points is invariant: $(w_1, w_2, w_3, w_4) = (z_1, z_2, z_3, z_4)$.

2. Composition of bilinear transformation is equivalent to multiplying matrices. Consider the bilinear transformations $f_{(a,b,c,d)}(z) = \frac{az+b}{cz+d}$ and $f_{(A,B,C,D)}(z) = \frac{Az+B}{Cz+D}$. The composition of these is

$$\left(f_{(a,b,c,d)} \circ f_{(A,B,C,D)}\right) = \frac{A\frac{az+b}{cz+d} + B}{C\frac{az+b}{cz+d} + D} = \frac{(Aa+Bc)z + (Ab+Bd)}{(Ca+Dc)z + (Cb+Dd)}$$

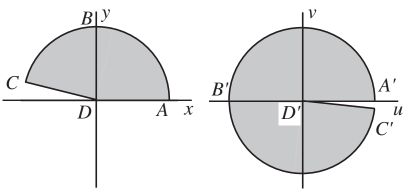
which is another bilinear transformation. Equivalently, note the matrix product

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \begin{bmatrix} Aa+Bc & Ab+Bd \\ Ca+Dc & Cb+Dd \end{bmatrix}$$

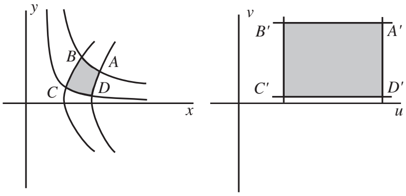
3. The *Möbius* transformation is a special case of the bilinear transformation; it is defined by $w(z) = \frac{z-a}{1-\overline{a}z}$ where a is a complex constant of modulus less than 1. It maps the unit disk onto itself.

5.13.12.2 Table of conformal mappings

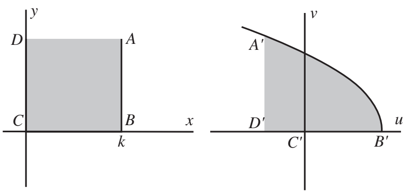
In the following functions $z = x + iy$ and $w = u + iv = \rho e^{i\phi}$.

1. 

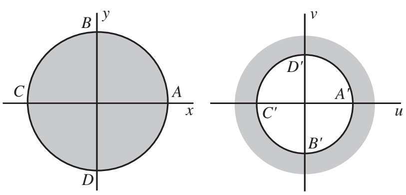
$w = z^2$.

2. 

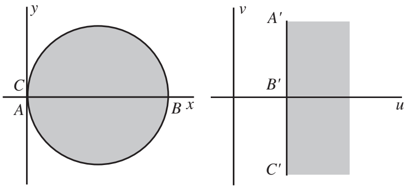
$w = z^2$.

3. 

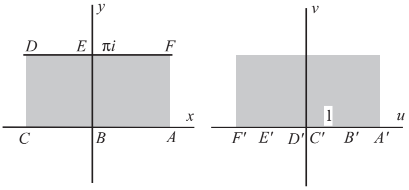
$w = z^2$; A', B' on the parabola $\rho = \dfrac{2k^2}{1+\cos\phi}$.

4. 

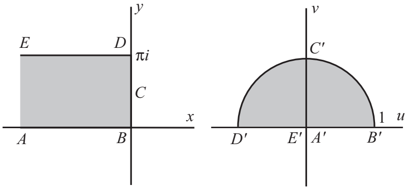
$w = \dfrac{1}{z}$.

5. 

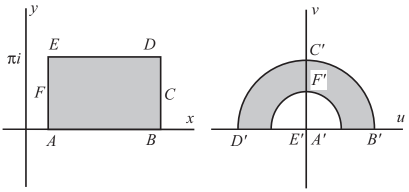
$w = \frac{1}{z}$.

6. 

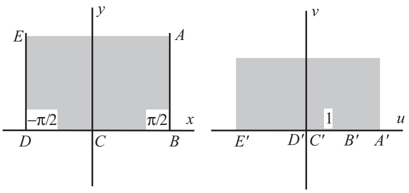
$w = e^z$.

7. 

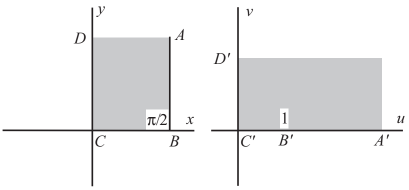
$w = e^z$.

8. 

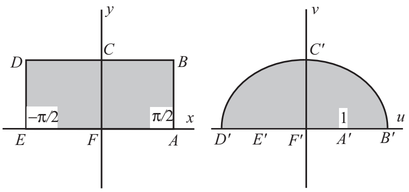
$w = e^z$.

9. 

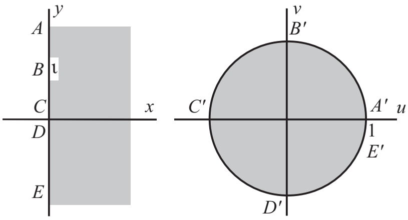
$w = \sin z$.

10. 

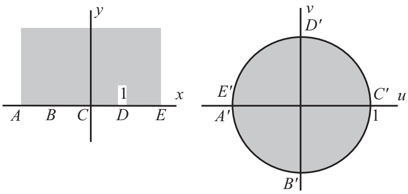
$w = \sin z$.

11. 

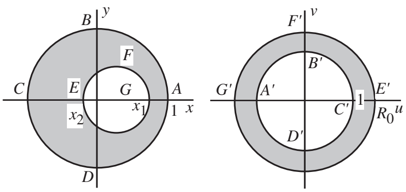
$w = \sin z$; BCD: $y = k$, $B'C'D'$ is on the ellipse
$$\left(\frac{u}{\cosh k}\right)^2 + \left(\frac{v}{\sinh k}\right)^2 = 1.$$

12. 

$w = \dfrac{z-1}{z+1}.$

13. 

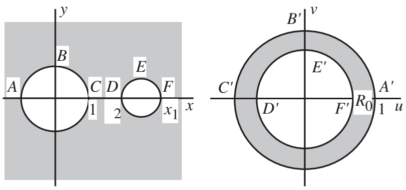
$w = \dfrac{i-z}{i+z}.$

14. 

$$w = \frac{z-a}{az-1};\; a = \frac{1 + x_1x_2 + \sqrt{(1-x_1^2)(1-x_2^2)}}{x_1 + x_2};$$

$$R_0 = \frac{1 - x_1x_2 + \sqrt{(1-x_1^2)(1-x_2^2)}}{x_1 - x_2}$$

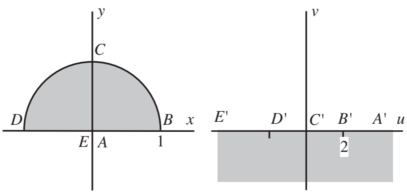
($a > 1$ and $R_0 > 1$ when $-1 < x_2 < x_1 < 1$).

15. 

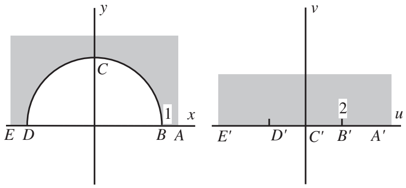
$$w = \frac{z-a}{az-1};\; a = \frac{1 + x_1x_2 + \sqrt{(x_1^2-1)(x_2^2-1)}}{x_1 + x_2};$$

$$R_0 = \frac{x_1x_2 - 1 - \sqrt{(x_1^2-1)(x_2^2-1)}}{x_1 - x_2}$$

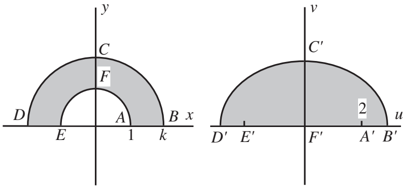
($x_2 < a < x_1$ and $0 < R_0 < 1$ when $1 < x_2 < x_1$).

16. 

$w = z + \dfrac{1}{z}.$

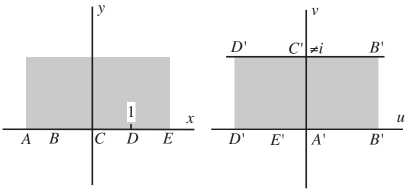
17.


$$w = z + \frac{1}{z}.$$

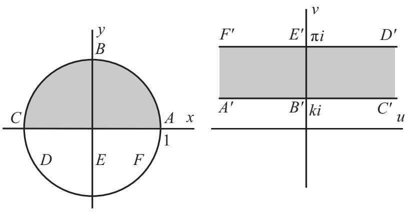
18.


$w = z + 1/z$; $B'C'D'$ is on the ellipse

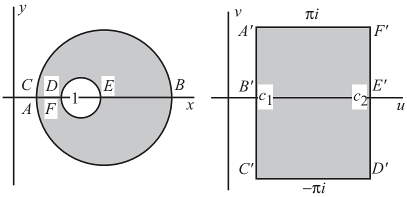
$$\left(\frac{ku}{k^2+1}\right)^2 + \left(\frac{kv}{k^2-1}\right)^2 = 1.$$

19.


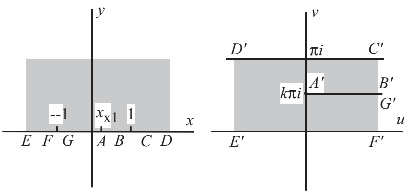
$$w = \log\frac{z-1}{z+1}.$$

20.


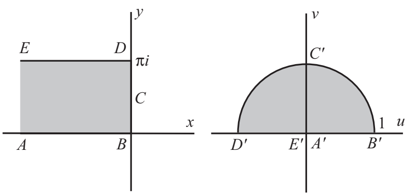
$w = \log\dfrac{z-1}{z+1}$; ABC is on the circle $x^2 + y^2 - 2y\cot k = 1$.

21.


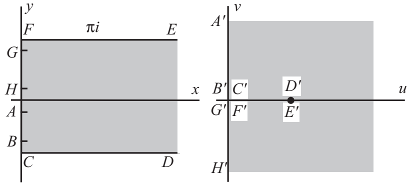
$w = \log\dfrac{z+1}{z-1}$; relationship between centers and radii: centers of circles at $z_n = \coth c_n$, radii are $\operatorname{csch} c_n$, $n = 1, 2$.

22.


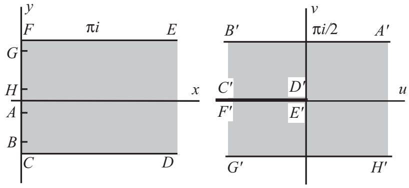
$w = k\log\dfrac{k}{1-k} + \log 2(1-k) + i\pi - k\log(z+1) - (1-k)\log(z-1)$; $x_1 = 2k - 1$.

23.


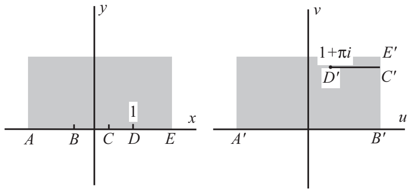
$$w = \tan^2\left(\frac{z}{2}\right).$$

24. 

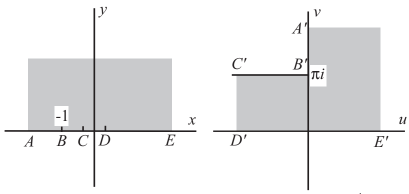
$w = \coth\left(\frac{z}{2}\right)$.

25. 

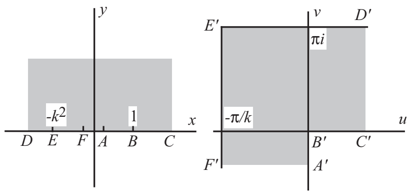
$w = \log\coth\left(\frac{z}{2}\right)$.

26. 

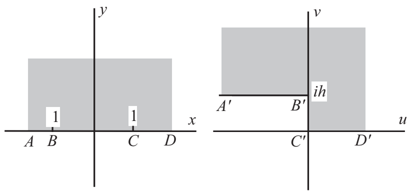
$w = \pi i + z - \log z$.

27. 

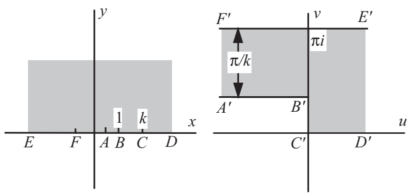
$$w = 2(z+1)^{1/2} + \log\frac{(z+1)^{1/2}-1}{(z+1)^{1/2}+1}.$$

28. 

$$w = \frac{i}{k}\log\frac{1+ikt}{1-ikt} + \log\frac{1+t}{1-t};$$
$$t = \left(\frac{z-1}{z+k^2}\right)^{1/2}.$$

29. 

$$w = \frac{h}{\pi}\left[(z^2-1)^{1/2} + \cosh^{-1} z\right].$$

30. 

$$w = \cosh^{-1}\left(\frac{2z-k-1}{k-1}\right) - \frac{1}{k}\cosh^{-1}\left[\frac{(k+1)z-2k}{(k-1)z}\right].$$

5.13.12.3 Table of transformations

$f(z) = w(x,y)$	$u(x,y) = \text{Re } w(x,y)$	$v(x,y) = \text{Im } w(x,y)$	Zeros (and order m)	Singularities (and order m)
z	x	y	$z=0$, $m=1$	Pole $(m=1)$ at $z=\infty$
z^2	x^2-y^2	$2xy$	$z=0$, $m=2$	Pole $(m=2)$ at $z=\infty$
$\dfrac{1}{z}$	$\dfrac{x}{x^2+y^2}$	$\dfrac{-y}{x^2+y^2}$	$z=\infty$, $m=1$	Pole $(m=1)$ at $z=0$
$\dfrac{1}{z^2}$	$\dfrac{x^2-y^2}{(x^2+y^2)^2}$	$\dfrac{-2xy}{(x^2+y^2)^2}$	$z=\infty$, $m=2$	Pole $(m=2)$ at $z=0$
$\dfrac{1}{z-(a+ib)}$ a, b real	$\dfrac{x-a}{(x-a)^2+(y-b)^2}$	$\dfrac{-(y-b)}{(x-a)^2+(y-b)^2}$	$z=\infty$, $m=1$	Pole $(m=1)$ at $z=a+ib$
$\sqrt{z}$	$\pm\left(\dfrac{x+\sqrt{x^2+y^2}}{2}\right)^{1/2}$	$\pm\left(\dfrac{-x+\sqrt{x^2+y^2}}{2}\right)^{1/2}$	$z=0$, $m=1$	Branch point $(m=1)$ at $z=0$ Branch point $(m=1)$ at $z=\infty$
e^z	$e^x\cos y$	$e^x\sin y$	None	Essential singularity at $z=\infty$
$\sin z$	$\sin x\cosh y$	$\cos x\sinh y$	$z=k\pi$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$
$\cos z$	$\cos x\cosh y$	$-\sin x\sinh y$	$z=(k+1/2)\pi$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$
$\sinh z$	$\sinh x\cos y$	$\cosh x\sin y$	$z=k\pi i$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$
$\cosh z$	$\cosh x\cos y$	$\sinh x\sin y$	$z=(k+1/2)\pi i$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$
$\tan z$	$\dfrac{\sin 2x}{\cos 2x+\cosh 2y}$	$\dfrac{\sinh 2y}{\cos 2x+\cosh 2y}$	$z=k\pi$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$ Poles $(m=1)$ at $z=(k+1/2)\pi$ $(k=0,\pm1,\pm2,\dots)$
$\tanh z$	$\dfrac{\sinh 2x}{\cosh 2x+\cos 2y}$	$\dfrac{\sin 2y}{\cosh 2x+\cos 2y}$	$z=k\pi i$, $m=1$ $(k=0,\pm1,\pm2,\dots)$	Essential singularity at $z=\infty$ Poles $(m=1)$ at $z=(k+1/2)\pi i$ $(k=0,\pm1,\pm2,\dots)$
$\log z$	$\frac{1}{2}\log(x^2+y^2)$	$\tan^{-1}\dfrac{x}{y}+2k\pi$ $(k=0,\pm1,\pm2,\dots)$	$z=1$, $m=1$	Branch points at $z=0$, $z=\infty$

5.14 SIGNIFICANT MATHEMATICAL EQUATIONS

Equations that contain profound information

1. Understanding of zero $\boxed{1-1=0}$
2. Cardinality of the continuum – understanding of infinities $\boxed{2^{\aleph_0} > \aleph_0}$
3. Euler's formula – connection between mathematical constants $\boxed{e^{i\pi}=-1}$
4. Quadratic Reciprocity Theorem – connection between primes
$$\boxed{\left(\frac{p}{q}\right) = (-1)^{\frac{p-1}{2}\frac{q-1}{2}}}$$

Equations that describe the world

1. Einstein equation – understanding time and space
$$\boxed{R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + g_{\mu\nu}\Lambda = \frac{8\pi G}{\pi^4}T_{\mu\nu}}$$
2. Euler's polyhedron formula – understanding polyhedra geometry
$$\boxed{\chi = V - E + F}$$
3. Gibbs equation – understanding thermodynamics $\boxed{\Delta G = \Delta H - T\Delta S}$
4. Maxwell's equations – understanding electromagnetism
$$\boxed{\begin{aligned} \nabla\cdot\mathbf{E} &= \frac{\rho}{\epsilon_0} \\ \nabla\cdot\mathbf{B} &= 0 \\ \nabla\times\mathbf{E} &= -\frac{\partial\mathbf{B}}{\partial t} \\ \nabla\times\mathbf{B} &= \mu_0\mathbf{J} + \mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t} \end{aligned}}$$
5. Navier–Stokes equations – understanding fluids
$$\boxed{\rho\left(\frac{\partial\mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla P + \nabla\cdot\mathbf{T} + \mathbf{f}}$$
6. Principle of least action – understanding physical optimization $\boxed{\delta S = 0}$
7. Schrödinger equation – understanding quantum mechanics
$$\boxed{i\hbar\frac{\partial}{\partial t}\Psi = H\Psi}$$

Chapter 6

Special Functions

6.1 CEILING AND FLOOR FUNCTIONS

The *ceiling function* of x, denoted $\lceil x \rceil$, is the least integer that is not smaller than x. For example, $\lceil \pi \rceil = 4$, $\lceil 5 \rceil = 5$, and $\lceil -1.5 \rceil = -1$.

The *floor function* of x, denoted $\lfloor x \rfloor$, is the largest integer that is not larger than x. For example, $\lfloor \pi \rfloor = 3$, $\lfloor 5 \rfloor = 5$, and $\lfloor -1.5 \rfloor = -2$.

6.2 EXPONENTIATION

For a any real number and m a positive integer, the *exponential* a^m is defined as

$$a^m = \underbrace{a \cdot a \cdot a \cdots a}_{m \text{ terms}}. \tag{6.2.1}$$

The following three laws of exponents follow for $a \neq 0$:

1. $a^n \cdot a^m = a^{m+n}$.
2. $(a^m)^n = a^{(mn)}$.
3. $\dfrac{a^m}{a^n} = \begin{cases} a^{m-n}, & \text{if } m > n, \\ 1, & \text{if } m = n, \\ \frac{1}{a^{n-m}}, & \text{if } m < n. \end{cases}$

The n^{th} *root function* is defined as the inverse of the n^{th} power function:

$$\text{If } b^n = a, \text{ then } b = \sqrt[n]{a} = a^{(1/n)}. \tag{6.2.2}$$

If n is odd, there will be a unique real number satisfying the definition of $\sqrt[n]{a}$, for any real value of a. If n is even, for positive values of a there will be two real values for $\sqrt[n]{a}$, one positive and one negative. By convention, the symbol $\sqrt[n]{a}$ means the positive value. If n is even and a is negative, then there are no real values for $\sqrt[n]{a}$.

To extend the definition to include a^t (for t not necessarily a positive integer), so as to maintain the laws of exponents, the following definitions are required (where we now restrict a to be positive, p and q are integers):

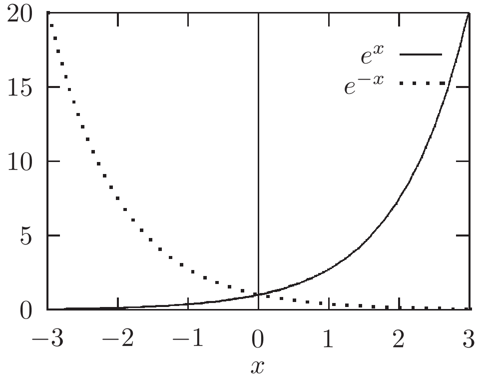
$$a^0 = 1 \qquad a^{p/q} = \sqrt[q]{a^p} \qquad a^{-t} = \frac{1}{a^t} \tag{6.2.3}$$

With these restrictions, the third law of exponents can be written as $\dfrac{a^m}{a^n} = a^{m-n}$.

If $a > 1$, then the function a^x is monotone increasing while, if $0 < a < 1$ then the function a^x is monotone decreasing.

6.3 EXPONENTIAL FUNCTION

$$\begin{aligned} \exp(z) = e^z &= \lim_{m\to\infty}\left(1+\frac{z}{m}\right)^m \\ &= 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{z^4}{4!} + \dots . \end{aligned} \tag{6.3.1}$$



- If $z = x + iy$, then $e^z = e^x e^{iy} = e^x(\cos y + i \sin y)$.
- $\int e^x \, dx = \frac{de^x}{dx} = e^x$.
- The value of e is on page 15.

6.4 LOGARITHMIC FUNCTIONS

6.4.1 DEFINITION OF THE NATURAL LOGARITHM

The *natural logarithm* (also known as the *Napierian logarithm*) of z is written as $\ln z$ or as $\log_e z$. It is sometimes written $\log z$ (this is also used to represent a "generic" logarithm, a logarithm to any base). One definition is

$$\ln z = \int_1^z \frac{dt}{t}, \tag{6.4.1}$$

where the integration path from 1 to z does not cross the origin or the negative real axis.

For complex values of z the natural logarithm, as defined above, can be represented in terms of its magnitude and phase. If $z = x + iy = re^{i\theta}$, then

$$\ln z = \ln r + i(\theta + 2k\pi), \tag{6.4.2}$$

for some $k = 0, \pm 1, \dots$, where $r = \sqrt{x^2 + y^2}$, $x = r \cos\theta$, and $y = r \sin\theta$. Usually, the value of k is chosen so that $0 \le (\theta + 2k\pi) < 2\pi$.

6.4.2 LOGARITHMS TO A DIFFERENT BASE

The logarithmic function to the base a, written $\log_a$, is defined as

$$\log_a z = \frac{\log_b z}{\log_b a} = \frac{\ln z}{\ln a}. \tag{6.4.3}$$

Note the properties:

1. $\log_a a^p = p$.
2. $\log_a b = \dfrac{1}{\log_b a}$.
3. $\log_{10} z = \dfrac{\ln z}{\ln 10} = (\log_{10} e) \ln z \approx (0.4342944819) \ln z$.
4. $\ln z = (\ln 10) \log_{10} z \approx (2.3025850929) \log_{10} z$.

6.4.3 LOGARITHM OF SPECIAL VALUES

- $\ln(-1) = i\pi + 2\pi i k$
- $\ln 0 = -\infty$
- $\ln 1 = 0$
- $\ln e = 1$
- $\ln(\pm i) = \pm \dfrac{i\pi}{2} + 2\pi i k$

6.4.4 RELATING THE LOGARITHM TO THE EXPONENTIAL

For positive values of z the logarithm is a monotonic function, as is the exponential for real z. Any monotonic function has a single-valued inverse function; the natural logarithm is the inverse of the exponential. If $x = e^y$, then $y = \ln x$, and $x = e^{\ln x}$. The same inverse relations hold for bases other than e. That is, if $u = a^w$, then $w = \log_a u$, and $u = a^{\log_a u}$.

6.4.5 IDENTITIES

$$\log_a z_1 z_2 = \log_a z_1 + \log_a z_2, \qquad \text{for } (-\pi < \arg z_1 + \arg z_2 < \pi).$$

$$\log_a \frac{z_1}{z_2} = \log_a z_1 - \log_a z_2, \qquad \text{for } (-\pi < \arg z_1 - \arg z_2 < \pi).$$

$$\log_a z^n = n \log_a z, \qquad \text{for } (-\pi < n \arg z < \pi), \text{ when } n \text{ is an integer.}$$

6.4.6 SERIES EXPANSIONS FOR THE NATURAL LOGARITHM

$$\ln(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \dots, \qquad \text{for } |z| < 1.$$

$$\ln z = \left(\frac{z-1}{z}\right) + \frac{1}{2}\left(\frac{z-1}{z}\right)^2 + \frac{1}{3}\left(\frac{z-1}{z}\right)^3 + \dots, \qquad \text{for Re } z \geq \frac{1}{2}.$$

6.4.7 DERIVATIVE AND INTEGRATION FORMULAS

$$\frac{d\ln z}{dz} = \frac{1}{z}, \qquad \int \frac{dz}{z} = \ln z, \qquad \int \ln z\, dz = z\ln z - z.$$

6.5 TRIGONOMETRIC FUNCTIONS

6.5.1 CIRCULAR FUNCTIONS

Consider the rectangular coordinate system in Figure 6.1. The coordinate x is positive to the right of the origin and the coordinate y is positive above the origin. The radius vector **r** shown terminating on the point $P(x, y)$ is shown rotated by the angle α up from the x axis. The radius vector **r** has component values x and y.

The trigonometric or circular functions of the angle α are defined in terms of the signed coordinates x and y and the length r which is always positive. Note that the coordinate x is negative in quadrants II and III and the coordinate y is negative in quadrants III and IV. The definitions of the trigonometric functions in terms of the Cartesian coordinates x and y of the point $P(x, y)$ are shown below. The angle α can be specified in radians, degrees, or any other unit.

$$\begin{aligned} \text{sine}\,\alpha &= \sin\alpha = y/r, & \text{cosine}\,\alpha &= \cos\alpha = x/r, \\ \text{tangent}\,\alpha &= \tan\alpha = y/x, & \text{cotangent}\,\alpha &= \cot\alpha = x/y, \\ \text{cosecant}\,\alpha &= \csc\alpha = r/y, & \text{secant}\,\alpha &= \sec\alpha = r/x. \end{aligned} \tag{6.5.1}$$

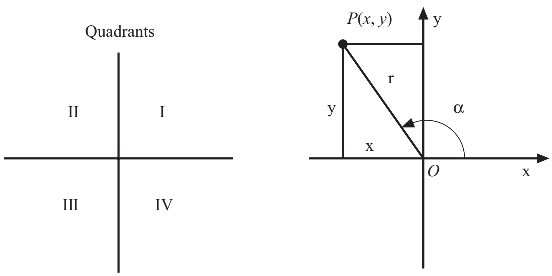
There are also the following seldom used functions:

versed sine of α = versine of α = vers $\alpha = 1 - \cos\alpha$,
coversed sine of α = versed cosine of α = covers $\alpha = 1 - \sin\alpha$,
exsecant of α = exsec $\alpha = \sec\alpha - 1$,
haversine of α = hav $\alpha = \frac{1}{2}$ vers $\alpha = \frac{1}{2}(1 - \cos\alpha)$.

6.5.1.1 Signs in the four quadrants

Quadrant	sin	cos	tan	csc	sec	cot
I	+	+	+	+	+	+
II	+	−	−	+	−	−
III	−	−	+	−	−	+
IV	−	+	−	−	+	−

FIGURE 6.1
The four quadrants (left) and notation for trigonometric functions (right).



6.5.2 DEFINITION OF ANGLES

If two lines intersect and one line is rotated about the point of intersection, the angle of rotation is designated positive if the rotation is counterclockwise. Angles are commonly measured in units of radians or degrees. Degrees are a historical unit related to the calendar defined by a complete revolution equaling 360 degrees (the approximate number of days in a year), written 360°. Radians are the angular unit usually used for mathematics and science. Radians are specified by the arc length traced by the tip of a rotating line divided by the length of that line. Thus a complete rotation of a line about the origin corresponds to 2π radians of rotation. It is a convenient convention that a full rotation of 2π radians is divided into four angular segments of $\pi/2$ each and that these are referred to as the four quadrants using Roman numerals I, II, III, and IV to designate them (see Figure 6.1).

A *right* angle is the angle between two perpendicular lines. It is equal to $\pi/2$ radians or 90 degrees. An *acute* angle is an angle less than $\pi/2$ radians. An *obtuse* angle is one between $\pi/2$ and π radians. A *convex* angle is one between 0 and π radians.

The angle π radians corresponds to 180 degrees. Therefore,

$$\begin{aligned}\text{one radian} &= \frac{180}{\pi}\,\text{degrees} = 57.30\,\text{degrees},\\ \text{one degree} &= \frac{\pi}{180}\,\text{radians} = 0.01745\,\text{radians}.\end{aligned} \tag{6.5.2}$$

FIGURE 6.2
Definitions of angles.

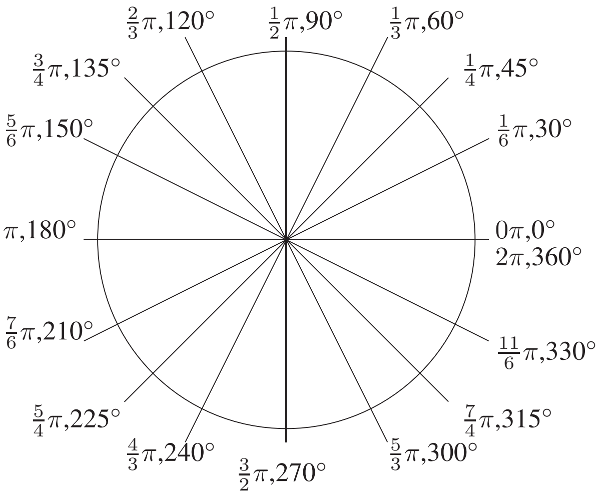
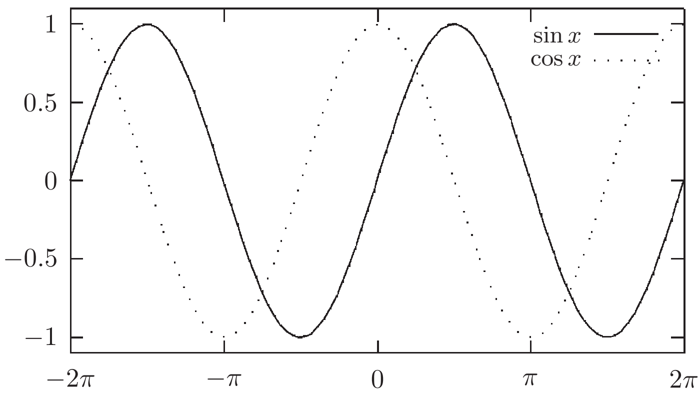


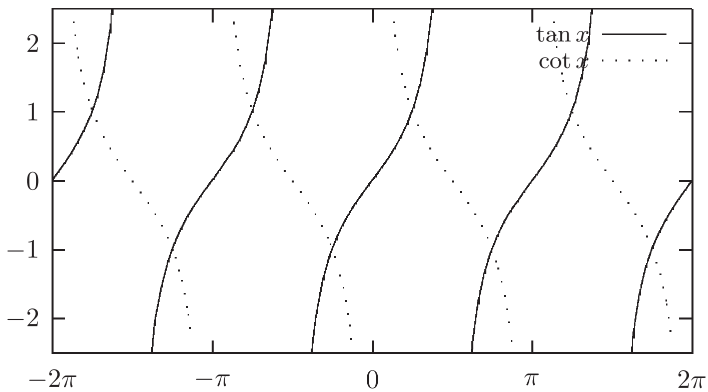
FIGURE 6.3
Sine and cosine, angles are in radians.



6.5.3 SYMMETRY AND PERIODICITY RELATIONSHIPS

$$\sin(-\alpha) = -\sin\alpha, \qquad \cos(-\alpha) = +\cos\alpha, \qquad \tan(-\alpha) = -\tan\alpha. \quad (6.5.3)$$

FIGURE 6.4
Tangent and cotangent, angles are in radians.



When n is any integer,

$$\begin{aligned}\sin(\alpha + n2\pi) &= \sin\alpha,\\ \cos(\alpha + n2\pi) &= \cos\alpha, \qquad (6.5.4)\\ \tan(\alpha + n\pi) &= \tan\alpha.\end{aligned}$$

6.5.4 ONE CIRCULAR FUNCTION IN TERMS OF ANOTHER

For $0 \le x \le \pi/2$,

	$\sin x$	$\cos x$	$\tan x$
$\sin x =$	$\sin x$	$\sqrt{1-\cos^2 x}$	$\dfrac{\tan x}{\sqrt{1+\tan^2 x}}$
$\cos x =$	$\sqrt{1-\sin^2 x}$	$\cos x$	$\dfrac{1}{\sqrt{1+\tan^2 x}}$
$\tan x =$	$\dfrac{\sin x}{\sqrt{1-\sin^2 x}}$	$\dfrac{\sqrt{1-\cos^2 x}}{\cos x}$	$\tan x$
$\csc x =$	$\dfrac{1}{\sin x}$	$\dfrac{1}{\sqrt{1-\cos^2 x}}$	$\dfrac{\sqrt{1+\tan^2 x}}{\tan x}$
$\sec x =$	$\dfrac{1}{\sqrt{1-\sin^2 x}}$	$\dfrac{1}{\cos x}$	$\sqrt{1+\tan^2 x}$
$\cot x =$	$\dfrac{\sqrt{1-\sin^2 x}}{\sin x}$	$\dfrac{\cos x}{\sqrt{1-\cos^2 x}}$	$\dfrac{1}{\tan x}$

	$\csc x$	$\sec x$	$\cot x$
$\sin x =$	$\dfrac{1}{\csc x}$	$\dfrac{\sqrt{\sec^2 x - 1}}{\sec x}$	$\dfrac{1}{\sqrt{1 + \cot^2 x}}$
$\cos x =$	$\dfrac{\sqrt{\csc^2 x - 1}}{\csc x}$	$\dfrac{1}{\sec x}$	$\dfrac{\cot x}{\sqrt{1 + \cot^2 x}}$
$\tan x =$	$\dfrac{1}{\sqrt{\csc^2 x - 1}}$	$\sqrt{\sec^2 x - 1}$	$\dfrac{1}{\cot x}$
$\csc x =$	$\csc x$	$\dfrac{\sec x}{\sqrt{\sec^2 x - 1}}$	$\sqrt{1 + \cot^2 x}$
$\sec x =$	$\dfrac{\csc x}{\sqrt{\csc^2 x - 1}}$	$\sec x$	$\dfrac{\sqrt{1 + \cot^2 x}}{\cot x}$
$\cot x =$	$\sqrt{\csc^2 x - 1}$	$\dfrac{1}{\sqrt{\sec^2 x - 1}}$	$\cot x$

6.5.5 CIRCULAR FUNCTIONS IN TERMS OF EXPONENTIALS

$$\cos z = \frac{e^{iz} + e^{-iz}}{2} \qquad e^{iz} = \cos z + i \sin z$$

$$\sin z = \frac{e^{iz} - e^{-iz}}{2i} \qquad e^{-iz} = \cos z - i \sin z$$

$$\tan z = \frac{\sin z}{\cos z} = \frac{e^{iz} - e^{-iz}}{i(e^{iz} + e^{-iz})}$$

6.5.6 FUNCTIONS IN TERMS OF ANGLES IN THE FIRST QUADRANT

When n is any integer:

	$-\alpha$	$\dfrac{\pi}{2} \pm \alpha$	$\pi \pm \alpha$	$\dfrac{3\pi}{2} \pm \alpha$	$2n\pi \pm \alpha$
sin	$-\sin \alpha$	$\cos \alpha$	$\mp \sin \alpha$	$-\cos \alpha$	$\pm \sin \alpha$
cos	$\cos \alpha$	$\mp \sin \alpha$	$-\cos \alpha$	$\pm \sin \alpha$	$+\cos \alpha$
tan	$-\tan \alpha$	$\mp \cot \alpha$	$\pm \tan \alpha$	$\mp \cot \alpha$	$\pm \tan \alpha$
csc	$-\csc \alpha$	$\sec \alpha$	$\mp \csc \alpha$	$-\sec \alpha$	$\pm \csc \alpha$
sec	$\sec \alpha$	$\mp \csc \alpha$	$-\sec \alpha$	$\pm \csc \alpha$	$\sec \alpha$
cot	$-\cot \alpha$	$\mp \tan \alpha$	$\pm \cot \alpha$	$\mp \tan \alpha$	$\pm \cot \alpha$

6.5.7 FUNDAMENTAL IDENTITIES

1. Reciprocal relations

$$\sin\alpha = \frac{1}{\csc\alpha}, \qquad \cos\alpha = \frac{1}{\sec\alpha}, \qquad \tan\alpha = \frac{\sin\alpha}{\cos\alpha} = \frac{1}{\cot\alpha},$$
$$\csc\alpha = \frac{1}{\sin\alpha}, \qquad \sec\alpha = \frac{1}{\cos\alpha}, \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha} = \frac{1}{\tan\alpha}.$$

2. Pythagorean theorem

$$\sin^2 z + \cos^2 z = 1 \qquad \sec^2 z - \tan^2 z = 1$$
$$\csc^2 z - \cot^2 z = 1$$

3. Product relations

$$\sin\alpha = \tan\alpha\cos\alpha \qquad \cos\alpha = \cot\alpha\sin\alpha$$
$$\tan\alpha = \sin\alpha\sec\alpha \qquad \cot\alpha = \cos\alpha\csc\alpha$$
$$\sec\alpha = \csc\alpha\tan\alpha \qquad \csc\alpha = \sec\alpha\cot\alpha$$

4. Quotient relations

$$\sin\alpha = \frac{\tan\alpha}{\sec\alpha} \qquad \cos\alpha = \frac{\cot\alpha}{\csc\alpha} \qquad \tan\alpha = \frac{\sin\alpha}{\cos\alpha}$$
$$\csc\alpha = \frac{\sec\alpha}{\tan\alpha} \qquad \sec\alpha = \frac{\csc\alpha}{\cot\alpha} \qquad \cot\alpha = \frac{\cos\alpha}{\sin\alpha}$$

6.5.8 PRODUCTS OF SINE AND COSINE

$$\cos\alpha\cos\beta = \frac{1}{2}\cos(\alpha-\beta) + \frac{1}{2}\cos(\alpha+\beta).$$
$$\sin\alpha\sin\beta = \frac{1}{2}\cos(\alpha-\beta) - \frac{1}{2}\cos(\alpha+\beta).$$
$$\sin\alpha\cos\beta = \frac{1}{2}\sin(\alpha-\beta) + \frac{1}{2}\sin(\alpha+\beta).$$

6.5.9 ANGLE SUM AND DIFFERENCE RELATIONSHIPS

$$\sin(\alpha\pm\beta) = \sin\alpha\cos\beta \pm \cos\alpha\sin\beta$$
$$\cos(\alpha\pm\beta) = \cos\alpha\cos\beta \mp \sin\alpha\sin\beta$$
$$\tan(\alpha\pm\beta) = \frac{\tan\alpha \pm \tan\beta}{1 \mp \tan\alpha\tan\beta}$$
$$\cot(\alpha\pm\beta) = \frac{\cot\alpha\cot\beta \mp 1}{\cot\beta \pm \cot\alpha}$$

6.5.10 MULTIPLE ANGLE FORMULAS

$$\sin 2\alpha = 2\sin\alpha\cos\alpha = \frac{2\tan\alpha}{1+\tan^2\alpha}$$

$$\sin 3\alpha = -4\sin^3\alpha + 3\sin\alpha.$$

$$\sin 4\alpha = -8\sin^3\alpha\cos\alpha + 4\sin\alpha\cos\alpha.$$

$$\sin n\alpha = 2\sin(n-1)\alpha\cos\alpha - \sin(n-2)\alpha.$$

$$\cos 2\alpha = 2\cos^2\alpha - 1 = 1 - 2\sin^2\alpha = \cos^2\alpha - \sin^2\alpha = \frac{1-\tan^2\alpha}{1+\tan^2\alpha}$$

$$\cos 3\alpha = 4\cos^3\alpha - 3\cos\alpha.$$

$$\cos 4\alpha = 8\cos^4\alpha - 8\cos^2\alpha + 1.$$

$$\cos n\alpha = 2\cos(n-1)\alpha\cos\alpha - \cos(n-2)\alpha.$$

$$\tan 2\alpha = \frac{2\tan\alpha}{1-\tan^2\alpha}$$

$$\tan 3\alpha = \frac{-\tan^3\alpha + 3\tan\alpha}{-3\tan^2\alpha + 1}.$$

$$\tan 4\alpha = \frac{-4\tan^3\alpha + 4\tan\alpha}{\tan^4\alpha - 6\tan^2\alpha + 1}.$$

$$\tan n\alpha = \frac{\tan(n-1)\alpha + \tan\alpha}{-\tan(n-1)\alpha\tan\alpha + 1}$$

$$\cot 2\alpha = \frac{\cot^2\alpha - 1}{2\cot\alpha}$$

6.5.11 HALF ANGLE FORMULAS

$$\cos\frac{\alpha}{2} = \pm\sqrt{\frac{1+\cos\alpha}{2}}$$

(positive if $\alpha/2$ is in quadrant I or IV, negative if in II or III).

$$\sin\frac{\alpha}{2} = \pm\sqrt{\frac{1-\cos\alpha}{2}}$$

(positive if $\alpha/2$ is in quadrant I or II, negative if in III or IV).

$$\tan\frac{\alpha}{2} = \frac{1-\cos\alpha}{\sin\alpha} = \frac{\sin\alpha}{1+\cos\alpha} = \pm\sqrt{\frac{1-\cos\alpha}{1+\cos\alpha}}$$

(positive if $\alpha/2$ is in quadrant I or III, negative if in II or IV).

$$\cot\frac{\alpha}{2} = \frac{1+\cos\alpha}{\sin\alpha} = \frac{\sin\alpha}{1-\cos\alpha} = \pm\sqrt{\frac{1+\cos\alpha}{1-\cos\alpha}}$$

(positive if $\alpha/2$ is in quadrant I or III, negative if in II or IV).

6.5.12 POWERS OF CIRCULAR FUNCTIONS

$$\sin^2\alpha = \frac{1}{2}(1-\cos 2\alpha) \qquad \cos^2\alpha = \frac{1}{2}(1+\cos 2\alpha)$$

$$\sin^3\alpha = \frac{1}{4}(-\sin 3\alpha + 3\sin\alpha) \qquad \cos^3\alpha = \frac{1}{4}(\cos 3\alpha + 3\cos\alpha)$$

$$\sin^4\alpha = \frac{1}{8}(3 - 4\cos 2\alpha + \cos 4\alpha) \qquad \cos^4\alpha = \frac{1}{8}(3 + 4\cos 2\alpha + \cos 4\alpha)$$

$$\tan^2\alpha = \frac{1-\cos 2\alpha}{1+\cos 2\alpha} \qquad \cot^2\alpha = \frac{1+\cos 2\alpha}{1-\cos 2\alpha}$$

6.5.13 SUMS OF CIRCULAR FUNCTIONS

$$\sin\alpha \pm \sin\beta = 2\sin\frac{\alpha\pm\beta}{2}\cos\frac{\alpha\mp\beta}{2}$$

$$\cos\alpha + \cos\beta = 2\cos\frac{\alpha+\beta}{2}\cos\frac{\alpha-\beta}{2}$$

$$\cos\alpha - \cos\beta = -2\sin\frac{\alpha+\beta}{2}\sin\frac{\alpha-\beta}{2}$$

$$\tan\alpha \pm \tan\beta = \frac{\sin\alpha\pm\beta}{\cos\alpha\cos\beta} \qquad \cot\alpha \pm \cot\beta = \frac{\sin\beta\pm\alpha}{\sin\alpha\sin\beta}$$

$$\frac{\sin\alpha+\sin\beta}{\sin\alpha-\sin\beta} = \frac{\tan\frac{\alpha+\beta}{2}}{\tan\frac{\alpha-\beta}{2}} \qquad \frac{\sin\alpha+\sin\beta}{\cos\alpha-\cos\beta} = \cot\frac{-\alpha+\beta}{2}$$

$$\frac{\sin\alpha+\sin\beta}{\cos\alpha+\cos\beta} = \tan\frac{\alpha+\beta}{2} \qquad \frac{\sin\alpha-\sin\beta}{\cos\alpha+\cos\beta} = \tan\frac{\alpha-\beta}{2}$$

6.5.14 CIRCULAR FUNCTIONS AS SERIES

See Section 1.9.6.4 on page 60.

6.5.15 RATIONAL TRIGONOMETRY

"Rational trigonometry" is a new approach to trigonometry. Whereas (usual) trigonometry considers *distance* and *angles* as fundamental quantities, Rational trigonometry uses *quadrance* and *spread*.

1. A *point* is an ordered pair of numbers $A = [x, y]$. Two points define a *line* ℓ.
2. The *quadrance* between points $A_1 = [x_1, y_1]$ and $A_2 = [x_2, y_2]$ is the number

$$Q(A_1, A_2) = (x_1 - x_2)^2 + (y_1 - y_2)^2 \equiv Q_{12} \tag{6.5.5}$$

3. For the points $\{A_1, A_2, A_3\}$ define: $Q_{23} = Q(A_2, A_3)$, $Q_{13} = Q(A_1, A_3)$, and $Q_{12} = Q(A_1, A_2)$. If the points $\{A_1, A_2, A_3\}$ are collinear (i.e., are on

the line $ax + by + c = 0$) then the *Triple quad formula* holds:

$$(Q_{12} + Q_{23} + Q_{13})^2 = 2\left(Q_{12}^2 + Q_{23}^2 + Q_{12}^2\right) \tag{6.5.6}$$

which can be written as

$$(Q_{23} + Q_{13} - Q_{12})^2 = 4Q_{23}Q_{13} \tag{6.5.7}$$

4. Consider two lines: ℓ_1 (defined by $a_1x + b_1y + c_1 = 0$) and ℓ_2 (defined by $a_2x + b_2y + c_2 = 0$). The *spread* of the lines ℓ_1 and ℓ_2 is the number

$$s(\ell_1, \ell_2) = \frac{(a_1b_2 - a_2b_1)^2}{(a_1^2 + b_1^2)(a_2^2 + b_2^2)} = s(\ell_2, \ell_1) \tag{6.5.8}$$

   (a) The spread s is in $[0, 1]$ and is the square of the sine of the angle.
   (b) If $s = 0$ then the lines are parallel and $a_1b_2 - a_2b_1 = 0$
   (c) If $s = 1$ then the lines are perpendicular and $a_1a_2 + b_1b_2 = 0$
   (d) The spread of lines meeting: at an angle of $30°$ or $150°$ is $s = {}^1\!/_4$; at an angle of $45°$ or $135°$ is $s = {}^1\!/_2$; at an angle of $60°$ or $120°$ is $s = {}^3\!/_4$.

5. Main laws of rational trigonometry:
   (define $s_{i,jk} = s(\ell_{ij}, \ell_{ik}) = s(A_iA_j, A_iA_k) = s_{i,kj}$)

   (a) *Pythagoras' theorem* The lines A_1A_3 and A_2A_3 are perpendicular precisely when $Q_{23} + Q_{13} = Q_{12}$.
   (b) *Spread law* (similar to the law of sines)
   For the triangle created by the points $\{A_1, A_2, A_3\}$

$$\frac{s_{1,23}}{Q_{23}} = \frac{s_{2,13}}{Q_{13}} = \frac{s_{3,12}}{Q_{12}} \tag{6.5.9}$$

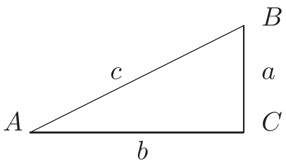
   (c) *Cross law* (similar to the law of cosines)
   For the triangle created by the points $\{A_1, A_2, A_3\}$

$$(Q_{23} + Q_{13} - Q_{12})^2 = 4Q_{23}Q_{13}\underbrace{(1 - s_{3,12})}_{\text{the "cross"}} \tag{6.5.10}$$

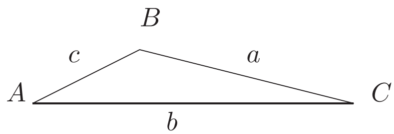
   (d) *Triple spread law* For the triangle created by the points $\{A_1, A_2, A_3\}$

$$(s_{1,23} + s_{2,13} + s_{3,12})^2 = 2\left(s_{1,23}^2 + s_{2,13}^2 + s_{3,12}^2\right) + 4s_{1,23}s_{2,13}s_{3,12} \tag{6.5.11}$$

6.6 CIRCULAR FUNCTIONS AND PLANAR TRIANGLES



Right triangle



General triangle

6.6.1 RIGHT TRIANGLES

Let A, B, and C be the vertices of a right triangle with C the right angle and a, b, and c the lengths of the sides opposite the corresponding vertices.

1. Trigonometric functions in terms of the sides of the triangle (a mnemonic is SOH-CAH-TOA)

$$\begin{aligned}
\sin A &= \frac{a}{c} = \frac{\text{opposite}}{\text{hypotenuse}} = \frac{1}{\csc A} \\
\cos A &= \frac{b}{c} = \frac{\text{adjacent}}{\text{hypotenuse}} = \frac{1}{\sec A} \\
\tan A &= \frac{a}{b} = \frac{\text{opposite}}{\text{adjacent}} = \frac{1}{\cot A}
\end{aligned} \tag{6.6.1}$$

2. The Pythagorean theorem states that $a^2 + b^2 = c^2$.
3. The sum of the interior angles equals π, that is $A + B + C = \pi$.

6.6.2 GENERAL PLANE TRIANGLES

Let A, B, and C be the interior angles of a general triangle, let a, b, and c be the length of the sides opposite those angles, and let $s = \frac{1}{2}(a + b + c)$ be the semi-perimeter.

1. Radius of the inscribed circle: $$r = \sqrt{\frac{(s-a)(s-b)(s-c)}{s}},$$

2. Radius of the circumscribed circle:
$$R = \frac{a}{2\sin A} = \frac{b}{2\sin B} = \frac{c}{2\sin C} = \frac{abc}{4(\text{Area})}.$$

3. Law of sines: $$\frac{a}{\sin A} = \frac{b}{\sin B} = \frac{c}{\sin C}.$$

4. Law of cosines:
$$a^2 = c^2 + b^2 - 2bc\cos A, \qquad \cos A = \frac{c^2 + b^2 - a^2}{2bc}.$$
$$b^2 = a^2 + c^2 - 2ca\cos B, \qquad \cos B = \frac{a^2 + c^2 - b^2}{2ca}.$$
$$c^2 = b^2 + a^2 - 2ab\cos C, \qquad \cos C = \frac{b^2 + a^2 - c^2}{2ab}.$$

5. Triangle sides in terms of other components:
$$a = b\cos C + c\cos B,$$
$$b = a\cos C + c\cos A,$$
$$c = b\cos A + a\cos B.$$

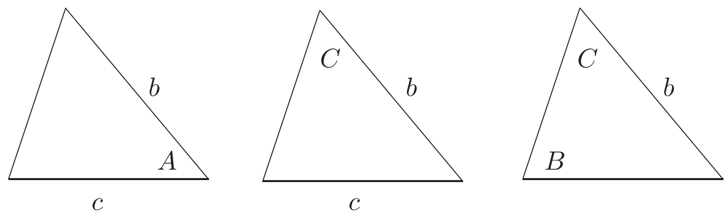
6. Law of tangents:
$$\frac{a+b}{a-b} = \frac{\tan\frac{A+B}{2}}{\tan\frac{A-B}{2}}, \qquad \frac{b+c}{b-c} = \frac{\tan\frac{B+C}{2}}{\tan\frac{B-C}{2}},$$
$$\frac{a+c}{a-c} = \frac{\tan\frac{A+C}{2}}{\tan\frac{A-C}{2}}.$$

7. Area of general triangle:
$$\begin{aligned} \text{Area} &= \frac{bc\sin A}{2} = \frac{ac\sin B}{2} = \frac{ab\sin C}{2}, \\ &= \frac{c^2\sin A\sin B}{2\sin C} = \frac{b^2\sin A\sin C}{2\sin B} = \frac{a^2\sin B\sin C}{2\sin A}, \\ &= rs = \frac{abc}{4R} = \sqrt{s(s-a)(s-b)(s-c)} \qquad \text{(Heron's formula).} \end{aligned}$$

8. Mollweide's formulas:
$$\frac{b-c}{a} = \frac{\sin\frac{1}{2}(B-C)}{\cos\frac{1}{2}A}, \qquad \frac{c-a}{b} = \frac{\sin\frac{1}{2}(C-A)}{\cos\frac{1}{2}B},$$
$$\frac{a-b}{c} = \frac{\sin\frac{1}{2}(A-B)}{\cos\frac{1}{2}C}.$$

9. Newton's formulas:
$$\frac{b+c}{a} = \frac{\cos\frac{1}{2}(B-C)}{\sin\frac{1}{2}A}, \qquad \frac{c+a}{b} = \frac{\cos\frac{1}{2}(C-A)}{\sin\frac{1}{2}B},$$
$$\frac{a+b}{c} = \frac{\cos\frac{1}{2}(A-B)}{\sin\frac{1}{2}C}.$$

FIGURE 6.5
Different triangles requiring solution.



6.6.3 SOLUTION OF TRIANGLES

A triangle is totally described by specifying any side and two additional parameters: either the remaining two sides (if they satisfy the triangle inequality), another side and the included angle, or two specified angles. If two sides are given and an angle that is not the included angle, then there might be 0, 1, or 2 such triangles. Two angles alone specify the shape of a triangle, but not its size, which requires specification of a side.

6.6.3.1 Three sides given

Formulas for any one of the angles:

$$\cos A = \frac{c^2+b^2-a^2}{2bc}, \qquad \sin A = \frac{2}{bc}\sqrt{s(s-a)(s-b)(s-c)},$$

$$\sin\frac{A}{2} = \sqrt{\frac{(s-b)(s-c)}{bc}}, \qquad \cos\frac{A}{2} = \sqrt{\frac{s(s-a)}{bc}},$$

$$\tan\frac{A}{2} = \sqrt{\frac{(s-b)(s-c)}{s(s-a)}} = \frac{r}{s-a}.$$

6.6.3.2 Given two sides (b, c) and the included angle (A)

See Figure 6.5, left. The remaining side and angles can be determined by repeated use of the law of cosines. For example,

1. Non-logarithmic solution; perform these steps sequentially:
   (a) $a^2 = b^2 + c^2 - 2bc\cos A$
   (b) $\cos B = (a^2 + c^2 - b^2)/2ca$
   (c) $\cos C = (a^2 + b^2 - c^2)/2ba$
2. Logarithmic solution; perform these steps sequentially:
   (a) $B + C = \pi - A$

(b) $\tan\dfrac{(B-C)}{2} = \dfrac{b-c}{b+c}\tan\dfrac{(B+C)}{2}$

(c) $B = \dfrac{B+C}{2} + \dfrac{B-C}{2}$

(d) $C = \dfrac{B+C}{2} - \dfrac{B-C}{2}$

(e) $a = \dfrac{b\sin A}{\sin B}$

6.6.3.3 Given two sides (b, c) and an angle (C), not the included angle

See Figure 6.5, middle. The remaining angles and side are determined by use of the law of sines and the fact that the sum of the angles is π, $A+B+C=\pi$.

$$\sin B = \frac{b\sin C}{c}, \qquad A = \pi - B - C, \qquad a = \frac{b\sin A}{\sin B}. \tag{6.6.2}$$

6.6.3.4 Given one side (b) and two angles (B, C)

See Figure 6.5, right. The third angle is specified by $A = \pi - B - C$. The remaining sides are found by

$$a = \frac{b\sin A}{\sin B}, \qquad c = \frac{b\sin C}{\sin B}. \tag{6.6.3}$$

6.6.4 HALF ANGLE FORMULAS

$$\tan\frac{A}{2} = \frac{r}{s-a} \qquad \tan\frac{B}{2} = \frac{r}{s-b} \qquad \tan\frac{C}{2} = \frac{r}{s-c}$$

$$\sin\frac{A}{2} = \sqrt{\frac{(s-b)(s-c)}{bc}} \qquad \cos\frac{A}{2} = \sqrt{\frac{s(s-a)}{bc}}$$

$$\sin\frac{B}{2} = \sqrt{\frac{(s-c)(s-a)}{ca}} \qquad \cos\frac{B}{2} = \sqrt{\frac{s(s-b)}{ca}}$$

$$\sin\frac{C}{2} = \sqrt{\frac{(s-a)(s-b)}{ab}} \qquad \cos\frac{C}{2} = \sqrt{\frac{s(s-c)}{ab}}$$

6.7 TABLES OF TRIGONOMETRIC FUNCTIONS

6.7.1 CIRCULAR FUNCTIONS OF SPECIAL ANGLES

Angle	$0 = 0°$	$\pi/12 = 15°$	$\pi/6 = 30°$	$\pi/4 = 45°$	$\pi/3 = 60°$
sin	0	$\frac{\sqrt{2}}{4}(\sqrt{3}-1)$	$1/2$	$\sqrt{2}/2$	$\sqrt{3}/2$
cos	1	$\frac{\sqrt{2}}{4}(\sqrt{3}+1)$	$\sqrt{3}/2$	$\sqrt{2}/2$	$1/2$
tan	0	$2-\sqrt{3}$	$\sqrt{3}/3$	1	$\sqrt{3}$
csc	∞	$\sqrt{2}(\sqrt{3}+1)$	2	$\sqrt{2}$	$2\sqrt{3}/3$
sec	1	$\sqrt{2}(\sqrt{3}-1)$	$2\sqrt{3}/3$	$\sqrt{2}$	2
cot	∞	$2+\sqrt{3}$	$\sqrt{3}$	1	$\sqrt{3}/3$

Angle	$5\pi/12 = 75°$	$\pi/2 = 90°$	$7\pi/12 = 105°$	$2\pi/3 = 120°$
sin	$\frac{\sqrt{2}}{4}(\sqrt{3}+1)$	1	$\frac{\sqrt{2}}{4}(\sqrt{3}+1)$	$\sqrt{3}/2$
cos	$\frac{\sqrt{2}}{4}(\sqrt{3}-1)$	0	$-\frac{\sqrt{2}}{4}(\sqrt{3}-1)$	$-1/2$
tan	$2+\sqrt{3}$	∞	$-(2+\sqrt{3})$	$-\sqrt{3}$
csc	$\sqrt{2}(\sqrt{3}-1)$	1	$\sqrt{2}(\sqrt{3}-1)$	$2\sqrt{3}/3$
sec	$\sqrt{2}(\sqrt{3}+1)$	∞	$-\sqrt{2}(\sqrt{3}+1)$	-2
cot	$2-\sqrt{3}$	0	$-(2-\sqrt{3})$	$-\sqrt{3}/3$

Angle	$3\pi/4 = 135°$	$5\pi/6 = 150°$	$11\pi/12 = 165°$	$\pi = 180°$
sin	$\sqrt{2}/2$	$1/2$	$\frac{\sqrt{2}}{4}(\sqrt{3}-1)$	0
cos	$-\sqrt{2}/2$	$-\sqrt{3}/2$	$-\frac{\sqrt{2}}{4}(\sqrt{3}+1)$	-1
tan	-1	$-\sqrt{3}/3$	$-(2-\sqrt{3})$	0
csc	$\sqrt{2}$	2	$\sqrt{2}(\sqrt{3}+1)$	∞
sec	$-\sqrt{2}$	$-2\sqrt{3}/3$	$-\sqrt{2}(\sqrt{3}-1)$	-1
cot	-1	$-\sqrt{3}$	$-(2+\sqrt{3})$	∞

6.7.2 EVALUATING SINES AND COSINES AT MULTIPLES OF π

	n an integer	n even	n odd	$n/2$ odd	$n/2$ even
$\sin n\pi$	0	0	0	0	0
$\cos n\pi$	$(-1)^n$	$+1$	-1	$+1$	$+1$
$\sin n\pi/2$		0	$(-1)^{(n-1)/2}$	0	0
$\cos n\pi/2$		$(-1)^{n/2}$	0	-1	$+1$

	n odd	$n/2$ odd	$n/2$ even
$\sin n\pi/4$	$(-1)^{(n^2+4n+11)/8}/\sqrt{2}$	$(-1)^{(n-2)/4}$	0

Note the useful formulas (where $i^2 = -1$)

$$\sin\frac{n\pi}{2} = \frac{i^{n+1}}{2}\left[(-1)^n - 1\right], \qquad \cos\frac{n\pi}{2} = \frac{i^n}{2}\left[(-1)^n + 1\right]. \tag{6.7.1}$$

6.7.3 TRIGONOMETRIC FUNCTIONS FOR DEGREE ARGUMENTS

(degrees) x	$\sin x$	$\cos x$	$\tan x$	$\cot x$	$\sec x$	$\csc x$
0	0	1	0	$\pm\infty$	1	$\pm\infty$
5	0.0872	0.9962	0.0875	11.4300	1.0038	11.4737
10	0.1736	0.9848	0.1763	5.6713	1.0154	5.7588
15	0.2588	0.9659	0.2680	3.7321	1.0353	3.8637
20	0.3420	0.9397	0.3640	2.7475	1.0642	2.9238
25	0.4226	0.9063	0.4663	2.1445	1.1034	2.3662
30	0.5000	0.8660	0.5774	1.7321	1.1547	2.0000
35	0.5736	0.8192	0.7002	1.4282	1.2208	1.7434
40	0.6428	0.7660	0.8391	1.1918	1.3054	1.5557
45	0.7071	0.7071	1.0000	1.0000	1.4142	1.4142
50	0.7660	0.6428	1.1918	0.8391	1.5557	1.3054
55	0.8192	0.5736	1.4282	0.7002	1.7434	1.2208
60	0.8660	0.5000	1.7321	0.5774	2.0000	1.1547
65	0.9063	0.4226	2.1445	0.4663	2.3662	1.1034
70	0.9397	0.3420	2.7475	0.3640	2.9238	1.0642
75	0.9659	0.2588	3.7321	0.2680	3.8637	1.0353
80	0.9848	0.1736	5.6713	0.1763	5.7588	1.0154
85	0.9962	0.0872	11.4300	0.0875	11.4737	1.0038
90	1	0	$\pm\infty$	0	$\pm\infty$	1
105	0.9659	−0.2588	−3.7321	−0.2680	−3.8637	1.0353
120	0.8660	−0.5000	−1.7321	−0.5774	−2.0000	1.1547
135	0.7071	−0.7071	−1.0000	−1.0000	−1.4142	1.4142
150	0.5000	−0.8660	−0.5774	−1.7321	−1.1547	2.0000
165	0.2588	−0.9659	−0.2680	−3.7321	−1.0353	3.8637
180	0	−1	0	$\pm\infty$	−1	$\pm\infty$
195	−0.2588	−0.9659	0.2680	3.7321	−1.0353	−3.8637
210	−0.5000	−0.8660	0.5774	1.7321	−1.1547	−2.0000
225	−0.7071	−0.7071	1.0000	1.0000	−1.4142	−1.4142
240	−0.8660	−0.5000	1.7321	0.5774	−2.0000	−1.1547
255	−0.9659	−0.2588	3.7321	0.2680	−3.8637	−1.0353
270	−1	0	$\pm\infty$	0	$\pm\infty$	−1
285	−0.9659	0.2588	−3.7321	−0.2680	3.8637	−1.0353
300	−0.8660	0.5000	−1.7321	−0.5774	2.0000	−1.1547
315	−0.7071	0.7071	−1.0000	−1.0000	1.4142	−1.4142
330	−0.5000	0.8660	−0.5774	−1.7321	1.1547	−2.0000
345	−0.2588	0.9659	−0.2680	−3.7321	1.0353	−3.8637
360	0	1	0	$\pm\infty$	1	$\pm\infty$

6.7.4 TRIGONOMETRIC FUNCTIONS FOR RADIAN ARGUMENTS

(radians) x	$\sin x$	$\cos x$	$\tan x$	$\cot x$	$\sec x$	$\csc x$
0	0	1	0	$\pm\infty$	1	$\pm\infty$
0.1	0.0998	0.9950	0.1003	9.9666	1.0050	10.0167
0.2	0.1987	0.9801	0.2027	4.9332	1.0203	5.0335
0.3	0.2955	0.9553	0.3093	3.2327	1.0468	3.3839
0.4	0.3894	0.9211	0.4228	2.3652	1.0857	2.5679
0.5	0.4794	0.8776	0.5463	1.8305	1.1395	2.0858
0.6	0.5646	0.8253	0.6841	1.4617	1.2116	1.7710
0.7	0.6442	0.7648	0.8423	1.1872	1.3075	1.5523
0.8	0.7174	0.6967	1.0296	0.9712	1.4353	1.3940
0.9	0.7833	0.6216	1.2602	0.7936	1.6087	1.2766
1.0	0.8415	0.5403	1.5574	0.6421	1.8508	1.1884
1.1	0.8912	0.4536	1.9648	0.5090	2.2046	1.1221
1.2	0.9320	0.3624	2.5722	0.3888	2.7597	1.0729
1.3	0.9636	0.2675	3.6021	0.2776	3.7383	1.0378
1.4	0.9854	0.1700	5.7979	0.1725	5.8835	1.0148
1.5	0.9975	0.0707	14.1014	0.0709	14.1368	1.0025
$\pi/2$	1	0	$\pm\infty$	0	$\pm\infty$	1
1.6	0.9996	−0.0292	−34.2325	−0.0292	−34.2471	1.0004
1.7	0.9917	−0.1288	−7.6966	−0.1299	−7.7613	1.0084
1.8	0.9738	−0.2272	−4.2863	−0.2333	−4.4014	1.0269
1.9	0.9463	−0.3233	−2.9271	−0.3416	−3.0932	1.0567
2.0	0.9093	−0.4161	−2.1850	−0.4577	−2.4030	1.0998
2.1	0.8632	−0.5048	−1.7098	−0.5848	−1.9808	1.1585
2.2	0.8085	−0.5885	−1.3738	−0.7279	−1.6992	1.2369
2.3	0.7457	−0.6663	−1.1192	−0.8935	−1.5009	1.3410
2.4	0.6755	−0.7374	−0.9160	−1.0917	−1.3561	1.4805
2.5	0.5985	−0.8011	−0.7470	−1.3386	−1.2482	1.6709
2.6	0.5155	−0.8569	−0.6016	−1.6622	−1.1670	1.9399
2.7	0.4274	−0.9041	−0.4727	−2.1154	−1.1061	2.3398
2.8	0.3350	−0.9422	−0.3555	−2.8127	−1.0613	2.9852
2.9	0.2392	−0.9710	−0.2464	−4.0584	−1.0299	4.1797
3.0	0.1411	−0.9900	−0.1425	−7.0153	−1.0101	7.0862
3.1	0.0416	−0.9991	−0.0416	−24.0288	−1.0009	24.0496
π	0	−1	0	$\pm\infty$	−1	$\pm\infty$

6.8 ANGLE CONVERSION

Degrees	Radians
1°	0.0174533
2°	0.0349066
3°	0.0523599
4°	0.0698132
5°	0.0872665
6°	0.1047198
7°	0.1221730
8°	0.1396263
9°	0.1570796
10°	0.1745329

Minutes	Radians
1′	0.00029089
2′	0.00058178
3′	0.00087266
4′	0.00116355
5′	0.00145444
6′	0.00174533
7′	0.00203622
8′	0.00232711
9′	0.00261799
10′	0.00290888

Seconds	Radians
1″	0.0000048481
2″	0.0000096963
3″	0.0000145444
4″	0.0000193925
5″	0.0000242407
6″	0.0000290888
7″	0.0000339370
8″	0.0000387851
9″	0.0000436332
10″	0.0000484814

Radians	Deg.	Min.	Sec.	Degrees
1	57°	17′	44.8″	57.2958
2	114°	35′	29.6″	114.5916
3	171°	53′	14.4″	171.8873
4	229°	10′	59.2″	229.1831
5	286°	28′	44.0″	286.4789
6	343°	46′	28.8″	343.7747
7	401°	4′	13.6″	401.0705
8	458°	21′	58.4″	458.3662
9	515°	39′	43.3″	515.6620

Radians	Deg.	Min.	Sec.	Degrees
0.1	5°	43′	46.5″	5.7296
0.2	11°	27′	33.0″	11.4592
0.3	17°	11′	19.4″	17.1887
0.4	22°	55′	5.9″	22.9183
0.5	28°	38′	52.4″	28.6479
0.6	34°	22′	38.9″	34.3775
0.7	40°	6′	25.4″	40.1070
0.8	45°	50′	11.8″	45.8366
0.9	51°	33′	58.3″	51.5662

Radians	Deg.	Min.	Sec.	Degrees
0.01	0°	34′	22.6″	0.5730
0.02	1°	8′	45.3″	1.1459
0.03	1°	43′	7.9″	1.7189
0.04	2°	17′	30.6″	2.2918
0.05	2°	51′	53.2″	2.8648
0.06	3°	26′	15.9″	3.4377
0.07	4°	0′	38.5″	4.0107
0.08	4°	35′	1.2″	4.5837
0.09	5°	9′	23.8″	5.1566

Radians	Deg.	Min.	Sec.	Degrees
0.001	0°	3′	26.3″	0.0573
0.002	0°	6′	52.5″	0.1146
0.003	0°	10′	18.8″	0.1719
0.004	0°	13′	45.1″	0.2292
0.005	0°	17′	11.3″	0.2865
0.006	0°	20′	37.6″	0.3438
0.007	0°	24′	3.9″	0.4011
0.008	0°	27′	30.1″	0.4584
0.009	0°	30′	56.4″	0.5157

6.9 INVERSE CIRCULAR FUNCTIONS

6.9.1 DEFINITION IN TERMS OF AN INTEGRAL

$$\begin{aligned} \text{arc sin}(z) &= \sin^{-1} z = \int_0^z \frac{dt}{\sqrt{1-t^2}}, \\ \text{arc cos}(z) &= \cos^{-1} z = \int_z^1 \frac{dt}{\sqrt{1-t^2}} = \frac{\pi}{2} - \sin^{-1} z, \\ \text{arc tan}(z) &= \tan^{-1} z = \int_0^z \frac{dt}{1+t^2} = \frac{\pi}{2} - \cot^{-1} z, \end{aligned} \tag{6.9.1}$$

where z can be complex. The path of integration must not cross the real axis in the first two cases and the imaginary axis in the third case except possibly inside the unit circle. If $-1 \leq x \leq 1$, then $\sin^{-1} x$ and $\cos^{-1} x$ are real, $-\frac{\pi}{2} \leq \sin^{-1} x \leq \frac{\pi}{2}$, and $0 \leq \cos^{-1} x \leq \pi$.

$$\begin{aligned} \csc^{-1} z &= \sin^{-1}(1/z), \\ \sec^{-1} z &= \cos^{-1}(1/z), \\ \cot^{-1} z &= \tan^{-1}(1/z), \\ \sec^{-1} z + \csc^{-1} z &= \pi/2. \end{aligned} \tag{6.9.2}$$

6.9.2 PRINCIPAL VALUES OF THE INVERSE CIRCULAR FUNCTIONS

The general solutions of $\{\sin t = z, \cos t = z, \tan t = z\}$ are, respectively:

$$\begin{aligned} t &= \sin^{-1} z = (-1)^k t_0 + k\pi, && \text{with } \sin t_0 = z, \\ t &= \cos^{-1} z = \pm t_1 + 2k\pi, && \text{with } \cos t_1 = z, \\ t &= \tan^{-1} z = t_2 + k\pi, && \text{with } \tan t_2 = z, \end{aligned}$$

where k is an arbitrary integer. While "$\sin^{-1} x$" can denote, as above, any angle whose $\sin$ is x; the function $\sin^{-1} x$ usually denotes the *principal value*. The principal values of the inverse trigonometric functions are defined as follows:

1. When $-1 \leq x \leq 1$, then $-\pi/2 \leq \sin^{-1} x \leq \pi/2$.
2. When $-1 \leq x \leq 1$, then $0 \leq \cos^{-1} x \leq \pi$.
3. When $-\infty \leq x \leq \infty$, then $-\pi/2 \leq \tan^{-1} x \leq \pi/2$.
4. When $1 \leq x$, then $0 \leq \csc^{-1} x \leq \pi/2$.
   When $x \leq -1$, then $-\pi/2 \leq \csc^{-1} x \leq 0$.
5. When $1 \leq x$, then $0 \leq \sec^{-1} x \leq \pi/2$.
   When $x \leq -1$, then $\pi/2 \leq \sec^{-1} x \leq \pi$.
6. When $-\infty \leq x \leq \infty$, then $0 \leq \cot^{-1} x \leq \pi$.

6.9.3 FUNDAMENTAL IDENTITIES

$$\sin^{-1} x + \cos^{-1} x = \pi/2 \qquad \tan^{-1} x + \cot^{-1} x = \pi/2$$

If $\alpha = \sin^{-1} x$, then

$$\sin\alpha = x, \qquad \cos\alpha = \sqrt{1-x^2}, \qquad \tan\alpha = \frac{x}{\sqrt{1-x^2}},$$

$$\csc\alpha = \frac{1}{x}, \qquad \sec\alpha = \frac{1}{\sqrt{1-x^2}}, \qquad \cot\alpha = \frac{\sqrt{1-x^2}}{x}.$$

If $\alpha = \cos^{-1} x$, then

$$\sin\alpha = \sqrt{1-x^2}, \qquad \cos\alpha = x, \qquad \tan\alpha = \frac{\sqrt{1-x^2}}{x},$$

$$\csc\alpha = \frac{1}{\sqrt{1-x^2}}, \qquad \sec\alpha = \frac{1}{x}, \qquad \cot\alpha = \frac{x}{\sqrt{1-x^2}}.$$

If $\alpha = \tan^{-1} x$, then

$$\sin\alpha = \frac{x}{\sqrt{1+x^2}}, \qquad \cos\alpha = \frac{1}{\sqrt{1+x^2}}, \qquad \tan\alpha = x,$$

$$\csc\alpha = \frac{\sqrt{1+x^2}}{x}, \qquad \sec\alpha = \sqrt{1+x^2}, \qquad \cot\alpha = \frac{1}{x}.$$

6.9.4 FUNCTIONS OF NEGATIVE ARGUMENTS

$$\sin^{-1}(-z) = -\sin^{-1} z, \qquad \sec^{-1}(-z) = \pi - \sec^{-1} z,$$

$$\cos^{-1}(-z) = \pi - \cos^{-1} z, \qquad \csc^{-1}(-z) = -\csc^{-1} z,$$

$$\tan^{-1}(-z) = -\tan^{-1} z, \qquad \cot^{-1}(-z) = \pi - \cot^{-1} z.$$

6.9.5 SUM AND DIFFERENCE OF TWO INVERSE TRIGONOMETRIC FUNCTIONS

$$\sin^{-1} z_1 \pm \sin^{-1} z_2 = \sin^{-1}\left(z_1\sqrt{1-z_2^2} \pm z_2\sqrt{1-z_1^2}\right).$$

$$\cos^{-1} z_1 \pm \cos^{-1} z_2 = \cos^{-1}\left(z_1 z_2 \mp \sqrt{(1-z_2^2)(1-z_1^2)}\right).$$

$$\tan^{-1} z_1 \pm \tan^{-1} z_2 = \tan^{-1}\left(\frac{z_1 \pm z_2}{1 \mp z_1 z_2}\right).$$

$$\sin^{-1} z_1 \pm \cos^{-1} z_2 = \sin^{-1}\left(z_1 z_2 \pm \sqrt{(1-z_1^2)(1-z_2^2)}\right),$$

$$= \cos^{-1}\left(z_2\sqrt{1-z_1^2} \mp z_1\sqrt{1-z_2^2}\right).$$

$$\tan^{-1} z_1 \pm \cot^{-1} z_2 = \tan^{-1}\left(\frac{z_1 z_2 \pm 1}{z_2 \mp z_1}\right) = \cot^{-1}\left(\frac{z_2 \mp z_1}{z_1 z_2 \pm 1}\right).$$

These formulas shown presume principal values of the inverse functions.

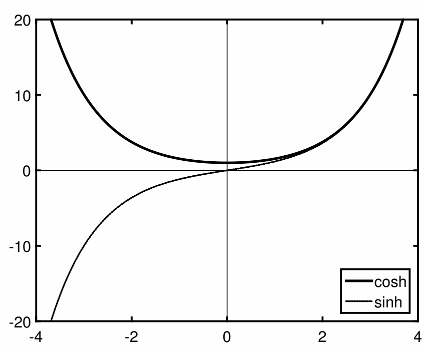
6.10 HYPERBOLIC FUNCTIONS

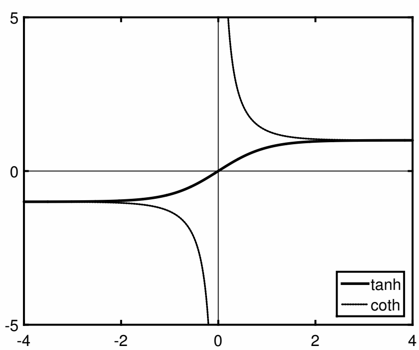
$$\sinh z = \frac{e^z - e^{-z}}{2}, \qquad \operatorname{csch} z = \frac{1}{\sinh z},$$
$$\cosh z = \frac{e^z + e^{-z}}{2}, \qquad \operatorname{sech} z = \frac{1}{\cosh z},$$
$$\tanh z = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{\sinh z}{\cosh z}, \qquad \coth z = \frac{1}{\tanh z}.$$

When $z = x + iy$,

$$\sinh z = \sinh x \cos y + i \cosh x \sin y, \qquad \tanh z = \frac{\sinh 2x + i \sin 2y}{\cosh 2x + \cos 2y},$$
$$\cosh z = \cosh x \cos y + i \sinh x \sin y, \qquad \coth z = \frac{\sinh 2x - i \sin 2y}{\cosh 2x - \cos 2y}.$$

Function	Domain (interval of u)	Range (interval of function)	Remarks
$\sinh u$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	
$\cosh u$	$(-\infty, +\infty)$	$[1, +\infty)$	
$\tanh u$	$(-\infty, +\infty)$	$(-1, +1)$	
$\operatorname{csch} u$	$(-\infty, 0)$	$(0, -\infty)$	Two branches,
	$(0, +\infty)$	$(+\infty, 0)$	pole at $u = 0$.
$\operatorname{sech} u$	$(-\infty, +\infty)$	$(0, 1]$	
$\coth u$	$(-\infty, 0)$	$(-1, -\infty)$	Two branches,
	$(0, +\infty)$	$(+\infty, 1)$	pole at $u = 0$.





6.10.1 RELATIONSHIP TO CIRCULAR FUNCTIONS

$$\cosh z = \cos iz, \qquad \sinh z = -i \sin iz, \qquad \tanh z = -i \tan iz.$$

6.10.2 HYPERBOLIC FUNCTIONS IN TERMS OF ONE ANOTHER

Function	$\sinh x$	$\cosh x$	$\tanh x$
$\sinh x =$	$\sinh x$	$\pm\sqrt{(\cosh x)^2 - 1}$	$\dfrac{\tanh x}{\sqrt{1-(\tanh x)^2}}$
$\cosh x =$	$\sqrt{1+(\sinh x)^2}$	$\cosh x$	$\dfrac{1}{\sqrt{1-(\tanh x)^2}}$
$\tanh x =$	$\dfrac{\sinh x}{\sqrt{1+(\sinh x)^2}}$	$\pm\dfrac{\sqrt{(\cosh x)^2-1}}{\cosh x}$	$\tanh x$
$\operatorname{csch} x =$	$\dfrac{1}{\sinh x}$	$\pm\dfrac{1}{\sqrt{(\cosh x)^2-1}}$	$\dfrac{\sqrt{1-(\tanh x)^2}}{\tanh x}$
$\operatorname{sech} x =$	$\dfrac{1}{\sqrt{1+(\sinh x)^2}}$	$\dfrac{1}{\cosh x}$	$\sqrt{1-(\tanh x)^2}$
$\coth x =$	$\dfrac{\sqrt{1+(\sinh x)^2}}{\sinh x}$	$\pm\dfrac{\cosh x}{\sqrt{(\cosh x)^2-1}}$	$\dfrac{1}{\tanh x}$

Function	$\operatorname{csch} x$	$\operatorname{sech} x$	$\coth x$
$\sinh x =$	$\dfrac{1}{\operatorname{csch} x}$	$\pm\dfrac{\sqrt{1-(\operatorname{sech} x)^2}}{\operatorname{sech} x}$	$\pm\dfrac{1}{\sqrt{(\coth x)^2-1}}$
$\cosh x =$	$\pm\dfrac{\sqrt{(\operatorname{csch} x)^2+1}}{\operatorname{csch} x}$	$\dfrac{1}{\operatorname{sech} x}$	$\pm\dfrac{\coth x}{\sqrt{(\coth x)^2-1}}$
$\tanh x =$	$\dfrac{1}{\sqrt{(\operatorname{csch} x)^2+1}}$	$\pm\sqrt{1-(\operatorname{sech} x)^2}$	$\dfrac{1}{\coth x}$
$\operatorname{csch} x =$	$\operatorname{csch} x$	$\pm\dfrac{\operatorname{sech} x}{\sqrt{1-(\operatorname{sech} x)^2}}$	$\pm\sqrt{(\coth x)^2-1}$
$\operatorname{sech} x =$	$\pm\dfrac{\operatorname{csch} x}{\sqrt{(\operatorname{csch} x)^2+1}}$	$\operatorname{sech} x$	$\pm\dfrac{\sqrt{(\coth x)^2-1}}{\coth x}$
$\coth x =$	$\sqrt{(\operatorname{csch} x)^2+1}$	$\pm\dfrac{1}{\sqrt{1-(\operatorname{sech} x)^2}}$	$\coth x$

6.10.3 RELATIONS AMONG HYPERBOLIC FUNCTIONS

$$e^z = \cosh z + \sinh z, \qquad e^{-z} = \cosh z - \sinh z,$$

$$(\cosh z)^2 - (\sinh z)^2 = (\tanh z)^2 + (\operatorname{sech} z)^2 = (\coth z)^2 - (\operatorname{csch} z)^2 = 1.$$

6.10.4 SYMMETRY RELATIONSHIPS

$$\cosh(-z) = +\cosh z, \qquad \sinh(-z) = -\sinh z, \qquad \tanh(-z) = -\tanh z.$$

6.10.5 SERIES EXPANSIONS

$$\cosh z = 1 + \frac{z^2}{2!} + \frac{z^4}{4!} + \frac{z^6}{6!} + \dots, \qquad |z| < \infty.$$
$$\sinh z = z + \frac{z^3}{3!} + \frac{z^5}{5!} + \frac{z^7}{7!} + \dots, \qquad |z| < \infty.$$
$$\tanh z = z - \frac{z^3}{3} + \frac{2z^5}{15} - \frac{17z^7}{315} + \dots, \qquad |z| < \frac{\pi}{2}.$$

6.10.6 PRODUCTS OF FUNCTIONS

$$\sinh u \sinh w = \frac{1}{2}\left(\cosh(u+w) - \cosh(u-w)\right),$$
$$\sinh u \cosh w = \frac{1}{2}\left(\sinh(u+w) + \sinh(u-w)\right),$$
$$\cosh u \cosh w = \frac{1}{2}\left(\cosh(u+w) + \cosh(u-w)\right).$$

6.10.7 SUM AND DIFFERENCE FORMULAS

$$\cosh(z_1 \pm z_2) = \cosh z_1 \cosh z_2 \pm \sinh z_1 \sinh z_2,$$
$$\sinh(z_1 \pm z_2) = \sinh z_1 \cosh z_2 \pm \cosh z_1 \sinh z_2,$$
$$\tanh(z_1 \pm z_2) = \frac{\tanh z_1 \pm \tanh z_2}{1 \pm \tanh z_1 \tanh z_2} = \frac{\sinh 2z_1 \pm \sinh 2z_2}{\cosh 2z_1 \pm \cosh 2z_2},$$
$$\coth(z_1 \pm z_2) = \frac{1 \pm \coth z_1 \coth z_2}{\coth z_1 \pm \coth z_2} = \frac{\sinh 2z_1 \mp \sinh 2z_2}{\cosh 2z_1 - \cosh 2z_2}.$$

6.10.8 SUMS OF FUNCTIONS

$$\sinh u \pm \sinh w = 2\sinh\frac{u \pm w}{2}\cosh\frac{u \mp w}{2},$$
$$\cosh u + \cosh w = 2\cosh\frac{u + w}{2}\cosh\frac{u - w}{2},$$
$$\cosh u - \cosh w = 2\sinh\frac{u + w}{2}\sinh\frac{u - w}{2},$$
$$\tanh u \pm \tanh w = \frac{\sinh u \pm w}{\cosh u \cosh w},$$
$$\coth u \pm \coth w = \frac{\sinh u \pm w}{\sinh u \sinh w}.$$

6.10.9 HALF–ARGUMENT FORMULAS

$$\sinh\frac{z}{2} = \pm\sqrt{\frac{\cosh z - 1}{2}}, \qquad \cosh\frac{z}{2} = +\sqrt{\frac{\cosh z + 1}{2}},$$

$$\tanh\frac{z}{2} = \pm\sqrt{\frac{\cosh z - 1}{\cosh z + 1}} = \frac{\sinh z}{\cosh z + 1}, \quad \coth\frac{z}{2} = \pm\sqrt{\frac{\cosh z + 1}{\cosh z - 1}} = \frac{\sinh z}{\cosh z - 1}.$$

6.10.10 MULTIPLE ARGUMENT RELATIONS

$$\begin{aligned}
\sinh 2\alpha &= 2\sinh\alpha\cosh\alpha = \frac{2\tanh\alpha}{1-\tanh^2\alpha}.\\
\sinh 3\alpha &= 3\sinh\alpha + 4\sinh^3\alpha = \sinh\alpha(4\cosh^2\alpha - 1).\\
\cosh 2\alpha &= \cosh^2\alpha + \sinh^2\alpha = 2\cosh^2\alpha - 1,\\
&= 1 + 2\sinh^2\alpha = \frac{1+\tanh^2\alpha}{1-\tanh^2\alpha}.\\
\cosh 3\alpha &= -3\cosh\alpha + 4\cosh^3\alpha = \cosh\alpha(4\sinh^2\alpha + 1).\\
\tanh 2\alpha &= \frac{2\tanh\alpha}{1+\tanh^2\alpha}.\\
\tanh 3\alpha &= \frac{3\tanh\alpha + \tanh^3\alpha}{1+3\tanh^2\alpha}.\\
\coth 2\alpha &= \frac{1+\coth^2\alpha}{2\coth\alpha}.\\
\coth 3\alpha &= \frac{3\coth\alpha + \coth^3\alpha}{1+3\coth^2\alpha}.
\end{aligned}$$

6.10.11 SUM AND DIFFERENCE OF INVERSE FUNCTIONS

$$\begin{aligned}
\sinh^{-1}x \pm \sinh^{-1}y &= \sinh^{-1}\left(x\sqrt{1+y^2} \pm y\sqrt{1+x^2}\right),\\
\cosh^{-1}x \pm \cosh^{-1}y &= \cosh^{-1}\left(xy \pm \sqrt{(y^2-1)(x^2-1)}\right),\\
\tanh^{-1}x \pm \tanh^{-1}y &= \tanh^{-1}\left(\frac{x \pm y}{xy \pm 1}\right),\\
\sinh^{-1}x \pm \cosh^{-1}y &= \sinh^{-1}\left(xy \pm \sqrt{(1+x^2)(y^2-1)}\right),\\
&= \cosh^{-1}\left(y\sqrt{1+x^2} \pm x\sqrt{y^2-1}\right),\\
\tanh^{-1}x \pm \coth^{-1}y &= \tanh^{-1}\left(\frac{xy \pm 1}{y \pm x}\right) = \coth^{-1}\left(\frac{y \pm x}{xy \pm 1}\right).
\end{aligned}$$

6.11 INVERSE HYPERBOLIC FUNCTIONS

$$\cosh^{-1} z = \int_0^z \frac{dt}{\sqrt{t^2-1}}, \quad \sinh^{-1} z = \int_0^z \frac{dt}{\sqrt{1+t^2}}, \quad \tanh^{-1} z = \int_0^z \frac{dt}{1-t^2}.$$

Function	Domain	Range	Remarks
$\sinh^{-1} u$	$(-\infty, +\infty)$	$(-\infty, +\infty)$	Odd function
$\cosh^{-1} u$	$[1, +\infty)$	$(-\infty, +\infty)$	Even function, double valued
$\tanh^{-1} u$	$(-1, +1)$	$(-\infty, +\infty)$	Odd function
$\operatorname{csch}^{-1} u$	$(-\infty, 0), (0, \infty)$	$(0, -\infty), (\infty, 0)$	Odd function, two branches, pole at $u = 0$
$\operatorname{sech}^{-1} u$	$(0, 1]$	$(-\infty, +\infty)$	Double valued
$\coth^{-1} u$	$(-\infty, -1), (1, \infty)$	$(-\infty, 0), (\infty, 0)$	Odd function, two branches

6.11.1 RELATIONSHIPS WITH LOGARITHMIC FUNCTIONS

$$\sinh^{-1} x = \log\left(x + \sqrt{x^2+1}\right), \qquad \operatorname{csch}^{-1} x = \log\left(\frac{1 \pm \sqrt{1+x^2}}{x}\right),$$

$$\cosh^{-1} x = \log\left(x \pm \sqrt{x^2-1}\right), \qquad \operatorname{sech}^{-1} x = \log\left(\frac{1 \pm \sqrt{1-x^2}}{x}\right),$$

$$\tanh^{-1} x = \frac{1}{2}\log\left(\frac{1+x}{1-x}\right), \qquad \coth^{-1} x = \frac{1}{2}\log\left(\frac{x+1}{x-1}\right).$$

6.11.2 RELATIONSHIPS WITH CIRCULAR FUNCTIONS

$$\sinh^{-1} x = -i \sin^{-1} ix \qquad \sinh^{-1} ix = +i \sin^{-1} x$$

$$\tanh^{-1} x = -i \tan^{-1} ix \qquad \tanh^{-1} ix = +i \tan^{-1} x$$

$$\operatorname{csch}^{-1} x = +i \csc^{-1} ix \qquad \operatorname{csch}^{-1} ix = -i \csc^{-1} x$$

$$\coth^{-1} x = +i \cot^{-1} ix \qquad \coth^{-1} ix = -i \cot^{-1} x$$

6.11.3 DIFFERENTIATION FORMULAS

$$\frac{d\sinh z}{dz} = \cosh z, \qquad \frac{d\cosh z}{dz} = \sinh z,$$

$$\frac{d\tanh z}{dz} = (\operatorname{sech} z)^2, \qquad \frac{d\operatorname{csch} z}{dz} = -\operatorname{csch} z \coth z,$$

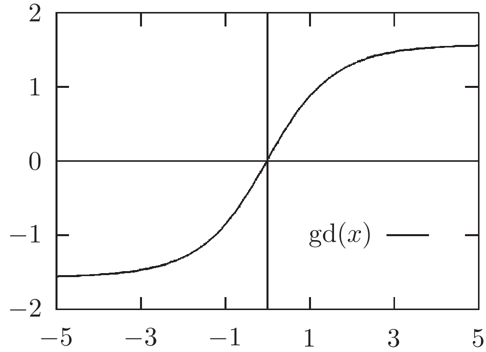
$$\frac{d\operatorname{sech} z}{dz} = -\operatorname{sech} z \tanh z, \qquad \frac{d\coth z}{dz} = -(\operatorname{csch} z)^2.$$

6.11.4 RELATIONSHIPS AMONG INVERSE HYPERBOLIC FUNCTIONS

Function	$\sinh^{-1} x$	$\cosh^{-1} x$	$\tanh^{-1} x$
$\sinh^{-1} x =$	$\sinh^{-1} x$	$\pm\cosh^{-1}\sqrt{x^2+1}$	$\tanh^{-1}\dfrac{x}{\sqrt{1+x^2}}$
$\cosh^{-1} x =$	$\pm\sinh^{-1}\sqrt{x^2-1}$	$\cosh^{-1} x$	$\pm\tanh^{-1}\dfrac{\sqrt{x^2-1}}{x}$
$\tanh^{-1} x =$	$\sinh^{-1}\dfrac{x}{\sqrt{1-x^2}}$	$\pm\cosh^{-1}\dfrac{1}{\sqrt{1-x^2}}$	$\tanh^{-1} x$
$\operatorname{csch}^{-1} x =$	$\sinh^{-1}\dfrac{1}{x}$	$\pm\cosh^{-1}\dfrac{\sqrt{1+x^2}}{x}$	$\tanh^{-1}\dfrac{1}{\sqrt{1+x^2}}$
$\operatorname{sech}^{-1} x =$	$\pm\sinh^{-1}\dfrac{\sqrt{1-x^2}}{x}$	$\cosh^{-1}\dfrac{1}{x}$	$\pm\tanh^{-1}\sqrt{1-x^2}$
$\coth^{-1} x =$	$\sinh^{-1}\dfrac{1}{\sqrt{x^2-1}}$	$\pm\cosh^{-1}\dfrac{x}{\sqrt{x^2-1}}$	$\tanh^{-1}\dfrac{1}{x}$

Function	$\operatorname{csch}^{-1} x$	$\operatorname{sech}^{-1} x$	$\coth^{-1} x$
$\sinh^{-1} x =$	$\operatorname{csch}^{-1}\dfrac{1}{x}$	$\pm\operatorname{sech}^{-1}\dfrac{1}{\sqrt{1+x^2}}$	$\coth^{-1}\dfrac{\sqrt{1+x^2}}{x}$
$\cosh^{-1} x =$	$\pm\operatorname{csch}^{-1}\dfrac{1}{\sqrt{x^2-1}}$	$\operatorname{sech}^{-1}\dfrac{1}{x}$	$\pm\coth^{-1}\dfrac{x}{\sqrt{x^2-1}}$
$\tanh^{-1} x =$	$\operatorname{csch}^{-1}\dfrac{\sqrt{1-x^2}}{x}$	$\pm\operatorname{sech}^{-1}\sqrt{1-x^2}$	$\coth^{-1}\dfrac{1}{x}$
$\operatorname{csch}^{-1} x =$	$\operatorname{csch}^{-1} x$	$\pm\operatorname{sech}^{-1}\dfrac{x}{\sqrt{1+x^2}}$	$\coth^{-1}\sqrt{1+x^2}$
$\operatorname{sech}^{-1} x =$	$\pm\operatorname{csch}^{-1}\dfrac{x}{\sqrt{1-x^2}}$	$\operatorname{sech}^{-1} x$	$\pm\coth^{-1}\dfrac{1}{\sqrt{1-x^2}}$
$\coth^{-1} x =$	$\operatorname{csch}^{-1}\sqrt{x^2-1}$	$\operatorname{sech}^{-1}\dfrac{\sqrt{x^2-1}}{x}$	$\coth^{-1} x$

6.12 GUDERMANNIAN FUNCTION



This function relates circular and hyperbolic functions without the use of functions of imaginary argument. The Gudermannian is a monotonic odd function which is asymptotic to $\pm\frac{\pi}{2}$ as $x \to \pm\infty$. It is zero at the origin.

$$\begin{aligned} \operatorname{gd} x &= \text{the Gudermannian of } x \\ &= \int_0^x \frac{dt}{\cosh t} = 2\tan^{-1}\left(\tanh\frac{x}{2}\right) = 2\tan^{-1} e^x - \frac{\pi}{2}. \\ \operatorname{gd}^{-1} x &= \text{the inverse Gudermannian of } x \\ &= \int_0^x \frac{dt}{\cos t} = \log\left[\tan\left(\frac{\pi}{4} + \frac{x}{2}\right)\right] = \log\left(\sec x + \tan x\right). \end{aligned} \tag{6.12.1}$$

If $\operatorname{gd}(x+iy) = \alpha + i\beta$, then

$$\tan\alpha = \frac{\sinh x}{\cos y}, \qquad \tanh\beta = \frac{\sin y}{\cosh x},$$

$$\tanh x = \frac{\sin\alpha}{\cosh\beta}, \qquad \tan y = \frac{\sin\beta}{\cosh\alpha}.$$

6.12.1 FUNDAMENTAL IDENTITIES

$$\begin{aligned} \tanh\left(\frac{x}{2}\right) &= \tan\left(\frac{\operatorname{gd} x}{2}\right), \\ e^x &= \cosh x + \sinh x = \sec\operatorname{gd} x + \tan\operatorname{gd} x, \\ &= \tan\left(\frac{\pi}{4} + \frac{\operatorname{gd} x}{2}\right) = \frac{1+\sin(\operatorname{gd} x)}{\cos(\operatorname{gd} x)}, \\ i\operatorname{gd}^{-1} x &= \operatorname{gd}^{-1}(ix), \qquad \text{where } i = \sqrt{-1}. \end{aligned}$$

6.12.2 DERIVATIVES OF GUDERMANNIAN

$$\frac{d(\operatorname{gd} x)}{dx} = \operatorname{sech} x \qquad \frac{d(\operatorname{gd}^{-1} x)}{dx} = \sec x.$$

6.12.3 HYPERBOLIC TO CIRCULAR FUNCTIONS

$$\sinh x = \tan(\mathrm{gd}\, x), \qquad \mathrm{csch}\, x = \cot(\mathrm{gd}\, x),$$
$$\cosh x = \sec(\mathrm{gd}\, x), \qquad \mathrm{sech}\, x = \cos(\mathrm{gd}\, x),$$
$$\tanh x = \sin(\mathrm{gd}\, x), \qquad \coth x = \csc(\mathrm{gd}\, x).$$

6.12.4 NUMERICAL VALUES OF HYPERBOLIC FUNCTIONS

x	e^x	$\ln x$	$\mathrm{gd}\, x$	$\sinh x$	$\cosh x$	$\tanh x$
0	1	$-\infty$	0	0	1	0
0.1	1.1052	−2.3026	0.0998	0.1002	1.0050	0.0997
0.2	1.2214	−1.6094	0.1987	0.2013	1.0201	0.1974
0.3	1.3499	−1.2040	0.2956	0.3045	1.0453	0.2913
0.4	1.4918	−0.9163	0.3897	0.4108	1.0811	0.3799
0.5	1.6487	−0.6931	0.4804	0.5211	1.1276	0.4621
0.6	1.8221	−0.5108	0.5669	0.6367	1.1855	0.5370
0.7	2.0138	−0.3567	0.6490	0.7586	1.2552	0.6044
0.8	2.2255	−0.2231	0.7262	0.8881	1.3374	0.6640
0.9	2.4596	−0.1054	0.7985	1.0265	1.4331	0.7163
1.0	2.7183	−0.0000	0.8658	1.1752	1.5431	0.7616
1.1	3.0042	0.0953	0.9281	1.3356	1.6685	0.8005
1.2	3.3201	0.1823	0.9857	1.5095	1.8107	0.8337
1.3	3.6693	0.2624	1.0387	1.6984	1.9709	0.8617
1.4	4.0552	0.3365	1.0872	1.9043	2.1509	0.8854
1.5	4.4817	0.4055	1.1317	2.1293	2.3524	0.9051
1.6	4.9530	0.4700	1.1724	2.3756	2.5775	0.9217
1.7	5.4739	0.5306	1.2094	2.6456	2.8283	0.9354
1.8	6.0496	0.5878	1.2432	2.9422	3.1075	0.9468
1.9	6.6859	0.6419	1.2739	3.2682	3.4177	0.9562
2.0	7.3891	0.6931	1.3018	3.6269	3.7622	0.9640
2.1	8.1662	0.7419	1.3271	4.0219	4.1443	0.9705
2.2	9.0250	0.7885	1.3501	4.4571	4.5679	0.9757
2.3	9.9742	0.8329	1.3709	4.9370	5.0372	0.9801
2.4	11.0232	0.8755	1.3899	5.4662	5.5569	0.9837
2.5	12.1825	0.9163	1.4070	6.0502	6.1323	0.9866
2.6	13.4637	0.9555	1.4225	6.6947	6.7690	0.9890
2.7	14.8797	0.9933	1.4366	7.4063	7.4735	0.9910
2.8	16.4446	1.0296	1.4493	8.1919	8.2527	0.9926
2.9	18.1741	1.0647	1.4609	9.0596	9.1146	0.9940
3.0	20.0855	1.0986	1.4713	10.0179	10.0677	0.9951
3.1	22.1980	1.1314	1.4808	11.0765	11.1215	0.9959
3.2	24.5325	1.1632	1.4893	12.2459	12.2866	0.9967
3.3	27.1126	1.1939	1.4971	13.5379	13.5748	0.9973
3.4	29.9641	1.2238	1.5041	14.9654	14.9987	0.9978
3.5	33.1155	1.2528	1.5104	16.5426	16.5728	0.9982
3.6	36.5982	1.2809	1.5162	18.2855	18.3128	0.9985
3.7	40.4473	1.3083	1.5214	20.2113	20.2360	0.9988
3.8	44.7012	1.3350	1.5261	22.3394	22.3618	0.9990
3.9	49.4024	1.3610	1.5303	24.6911	24.7113	0.9992
4.0	54.5982	1.3863	1.5342	27.2899	27.3082	0.9993

6.13 ORTHOGONAL POLYNOMIALS

Orthogonal polynomials are classes of polynomials, $\{p_n(x)\}$, which obey an orthogonality relationship of the form

$$\int_I w(x)\, p_n(x)\, p_m(x)\, dx = c_n \delta_{nm} \tag{6.13.1}$$

for a given *weight function* $w(x)$ and interval I.

6.13.1 HERMITE POLYNOMIALS

Symbol: $H_n(x)$

Interval: $(-\infty, \infty)$

Differential Equation: $y'' - 2xy' + 2ny = 0$

Explicit Expression: $H_n(x) = \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m n!(2x)^{n-2m}}{m!(n-2m)!}$

Recurrence Relation: $H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$

Weight: e^{-x^2}

Standardization: $H_n(x) = 2^n x^n + \dots$

Norm: $\int_{-\infty}^{\infty} e^{-x^2} \left[H_n(x)\right]^2 \, dx = 2^n n! \sqrt{\pi}$

Rodrigues' Formula: $H_n(x) = (-1)^n e^{x^2} \frac{d^n}{dx^n}(e^{-x^2})$

Generating Function: $\sum_{n=0}^{\infty} H_n(x) \frac{z^n}{n!} = e^{-z^2+2zx}$

Inequality: $|H_n(x)| < \sqrt{2^n e^{x^2} n!}$

6.13.2 JACOBI POLYNOMIALS

Symbol: $P_n^{(\alpha,\beta)}(x)$

Interval: $[-1, 1]$

Parameter Range: $\alpha, \beta > -1$

Differential Equation:

$$(1-x^2)y'' + [\beta - \alpha - (\alpha+\beta+2)x]y' + n(n+\alpha+\beta+1)y = 0$$

Explicit Expression:

$$P_n^{(\alpha,\beta)}(x) = \frac{1}{2^n} \sum_{m=0}^{n} \binom{n+\alpha}{m} \binom{n+\beta}{n-m} (x-1)^{n-m}(x+1)^m$$

Recurrence Relation:

$$\begin{aligned} &2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)P_{n+1}^{(\alpha,\beta)}(x) \\ &= (2n+\alpha+\beta+1)[(\alpha^2-\beta^2) + (2n+\alpha+\beta+2)(2n+\alpha+\beta)x]P_n^{(\alpha,\beta)}(x) \\ &\qquad -2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2)P_{n-1}^{(\alpha,\beta)}(x) \end{aligned}$$

Weight: $(1-x)^\alpha (1+x)^\beta$

Standardization: $P_n^{(\alpha,\beta)}(1) = \binom{n+\alpha}{n}$

Norm: $\displaystyle\int_{-1}^{1}(1-x)^\alpha(1+x)^\beta\left[P_n^{(\alpha,\beta)}(x)\right]^2\,dx = \frac{2^{\alpha+\beta+1}\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{(2n+\alpha+\beta+1)n!\Gamma(n+\alpha+\beta+1)}$

Rodrigues' Formula:

$$P_n^{(\alpha,\beta)}(x) = \frac{(-1)^n}{2^n n!(1-x)^\alpha(1+x)^\beta}\frac{d^n}{dx^n}\left[(1-x)^{n+\alpha}(1+x)^{n+\beta}\right]$$

Generating Function:

$$\sum_{n=0}^{\infty} P_n^{(\alpha,\beta)}(x)z^n = 2^{\alpha+\beta}R^{-1}(1-z+R)^{-\alpha}(1+z+R)^{-\beta},$$

where $R=\sqrt{1-2xz+z^2}$ and $|z|<1$

Inequality:

$$\max_{-1\le x\le 1}\left|P_n^{(\alpha,\beta)}(x)\right| = \begin{cases}\binom{n+q}{n}\sim n^q, & \text{if } q=\max(\alpha,\beta)\ge -\frac{1}{2},\\ \left|P_n^{(\alpha,\beta)}(x')\right|\sim n^{-1/2}, & \text{if } q=\max(\alpha,\beta)< -\frac{1}{2},\end{cases}$$

where $\alpha,\beta>1$ and x' (in the second result) is one of the two maximum points nearest $(\beta-\alpha)/(\alpha+\beta+1)$.

6.13.3 LAGUERRE POLYNOMIALS

Symbol: $L_n(x)$.
Interval: $[0,\infty)$.

$L_n(x)$ is the same as $L_n^{(0)}(x)$ (see the generalized Laguerre polynomials).

6.13.4 GENERALIZED LAGUERRE POLYNOMIALS

Symbol: $L_n^{(\alpha)}(x)$

Interval: $[0,\infty)$

Differential Equation: $xy''+(\alpha+1-x)y'+ny=0$

Explicit Expression: $\displaystyle L_n^{(\alpha)}(x)=\sum_{m=0}^{n}\frac{(-1)^m}{m!}\binom{n+\alpha}{n-m}x^m$

Recurrence Relation:

$$(n+1)L_{n+1}^{(\alpha)}(x)=[(2n+\alpha+1)-x]L_n^{(\alpha)}(x)-(n+\alpha)L_{n-1}^{(\alpha)}(x)$$

Weight: $x^\alpha e^{-x}$

Standardization: $\displaystyle L_n^{(\alpha)}(x)=\frac{(-1)^n}{n!}x^n+\dots$

Norm: $\displaystyle\int_0^\infty x^\alpha e^{-x}\left[L_n^{(\alpha)}(x)\right]^2\,dx=\frac{\Gamma(n+\alpha+1)}{n!}$

Rodrigues' Formula: $\displaystyle L_n^{(\alpha)}(x)=\frac{1}{n!x^\alpha e^{-x}}\frac{d^n}{dx^n}[x^{n+\alpha}e^{-x}]$

Generating Function: $\displaystyle\sum_{n=0}^{\infty}L_n^{(\alpha)}(x)z^n=(1-z)^{-\alpha-1}\exp\left(\frac{xz}{z-1}\right)$

Inequality:

$$\left|L_n^{(\alpha)}(x)\right|\le\begin{cases}\frac{\Gamma(n+\alpha+1)}{n!\Gamma(\alpha+1)}e^{x/2}, & \text{if } x\ge 0 \text{ and } \alpha>0,\\ \left[2-\frac{\Gamma(n+\alpha+1)}{n!\Gamma(\alpha+1)}\right]e^{x/2}, & \text{if } x\ge 0 \text{ and } -1<\alpha<0.\end{cases}$$

Note that $\alpha>-1$ and $\displaystyle L_n^{(m)}(x)=(-1)^m\frac{d^m}{dx^m}\left[L_{n+m}(x)\right]$.

6.13.5 LEGENDRE POLYNOMIALS

Symbol: $P_n(x)$
Interval: $[-1, 1]$
Differential Equation: $(1-x^2)y'' - 2xy' + n(n+1)y = 0$
Explicit Expression: $P_n(x) = \dfrac{1}{2^n}\displaystyle\sum_{m=0}^{\lfloor n/2 \rfloor}(-1)^m \binom{n}{m}\binom{2n-2m}{n} x^{n-2m}$
Recurrence Relation: $(n+1)P_{n+1}(x) = (2n+1)xP_n(x) - nP_{n-1}(x)$
Weight: 1
Standardization: $P_n(1) = 1$
Norm: $\displaystyle\int_{-1}^{1} [P_n(x)]^2\, dx = \frac{2}{2n+1}$
Rodrigues' Formula: $P_n(x) = \dfrac{(-1)^n}{2^n n!}\dfrac{d^n}{dx^n}[(1-x^2)^n]$
Generating Function:

$$\sum_{n=0}^{\infty} P_n(x)z^n = (1-2xz+z^2)^{-1/2}, \text{ for } -1 < x < 1 \text{ and } |z| < 1$$

Inequality: $|P_n(x)| \le 1$ for $-1 \le x \le 1$

See Section 6.24.5 on page 486.

6.13.6 CHEBYSHEV POLYNOMIALS, FIRST KIND

Symbol: $T_n(x)$
Interval: $[-1, 1]$
Differential Equation: $(1-x^2)y'' - xy' + n^2 y = 0$
Explicit Expression:

$$T_n(x) = \cos\left(n \cos^{-1} x\right) = \frac{n}{2}\sum_{m=0}^{\lfloor n/2 \rfloor}(-1)^m \frac{(n-m-1)!}{m!(n-2m)!}(2x)^{n-2m}$$

Recurrence Relation: $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$
Weight: $(1-x^2)^{-1/2}$
Standardization: $T_n(1) = 1$
Norm: $\int_{-1}^{1}(1-x^2)^{-1/2}\,[T_n(x)]^2\, dx = \begin{cases} \pi, & n = 0, \\ \pi/2, & n \neq 0. \end{cases}$
Rodrigues' Formula: $T_n(x) = \dfrac{\sqrt{\pi(1-x^2)}}{(-2)^n\Gamma(n+\frac{1}{2})}\dfrac{d^n}{dx^n}\left[(1-x^2)^{n-1/2}\right]$
Generating Function:

$$\sum_{n=0}^{\infty} T_n(x)z^n = \frac{1-xz}{1-2xz+z^2}, \text{ for } -1 < x < 1 \text{ and } |z| < 1$$

Inequality: $|T_n(x)| \le 1$ for $-1 \le x \le 1$

Note that $T_n(x) = \dfrac{n!\sqrt{\pi}}{\Gamma(n+\frac{1}{2})}P_n^{(-1/2,-1/2)}(x)$.

6.13.7 CHEBYSHEV POLYNOMIALS, SECOND KIND

Symbol: $U_n(x)$
Interval: $[-1, 1]$
Differential Equation: $(1-x^2)y'' - 3xy' + n(n+2)y = 0$
Explicit Expression: $U_n(x) = \sum_{m=0}^{\lfloor n/2 \rfloor} \frac{(-1)^m (n-m)!}{m!(n-2m)!}(2x)^{n-2m}$

and $U_n(\cos\theta) = \frac{\sin[(n+1)\theta]}{\sin\theta}$

Recurrence Relation: $U_{n+1}(x) = 2xU_n(x) - U_{n-1}(x)$
Weight: $(1-x^2)^{1/2}$
Standardization: $U_n(1) = n+1$
Norm: $\int_{-1}^{1} (1-x^2)^{1/2} \left[U_n(x)\right]^2 \, dx = \frac{\pi}{2}$
Rodrigues' Formula:

$$U_n(x) = \frac{(-1)^n (n+1)\sqrt{\pi}}{(1-x^2)^{1/2} 2^{n+1} \Gamma(n+\frac{3}{2})} \frac{d^n}{dx^n}[(1-x^2)^{n+(1/2)}]$$

Generating Function:

$$\sum_{n=0}^{\infty} U_n(x) z^n = \frac{1}{1-2xz+z^2}, \text{ for } -1 < x < 1 \text{ and } |z| < 1$$

Inequality: $|U_n(x)| \le n+1$ for $-1 \le x \le 1$

Note that $U_n(x) = \frac{(n+1)!\sqrt{\pi}}{2\Gamma(n+\frac{3}{2})} P_n^{(1/2,1/2)}(x)$.

6.13.8 TABLES OF ORTHOGONAL POLYNOMIALS

6.13.8.1 Table of Jacobi polynomials

Notation: $(m)_n = m(m+1)\ldots(m+n-1)$.

$$P_0^{(\alpha,\beta)}(x) = 1.$$

$$P_1^{(\alpha,\beta)}(x) = \frac{1}{2}\Big(2(\alpha+1) + (\alpha+\beta+2)(x-1)\Big).$$

$$P_2^{(\alpha,\beta)}(x) = \frac{1}{8}\Big(4(\alpha+1)_2 + 4(\alpha+\beta+3)(\alpha+2)(x-1) + (\alpha+\beta+3)_2(x-1)^2\Big).$$

$$P_3^{(\alpha,\beta)}(x) = \frac{1}{48}\Big(8(\alpha+1)_3 + 12(\alpha+\beta+4)(\alpha+2)_2(x-1) + 6(\alpha+\beta+4)_2(\alpha+3)(x-1)^2 + (\alpha+\beta+4)_3(x-1)^3\Big).$$

$$P_4^{(\alpha,\beta)}(x) = \frac{1}{384}\Big(16(\alpha+1)_4 + 32(\alpha+\beta+5)(\alpha+2)_3(x-1) + 24(\alpha+\beta+5)_2(\alpha+3)_2(x-1)^2 + 8(\alpha+\beta+5)_3(\alpha+4)(x-1)^3 + (\alpha+\beta+5)_4(x-1)^4\Big).$$

6.13.9 TABLES OF ORTHOGONAL POLYNOMIALS (H, L, P, T, U)

$H_0 = 1$	$x^8 = (1680H_0 + 3360H_2 + 840H_4 + 56H_6 + H_8)/256$
$H_1 = 2x$	$x^7 = (840H_1 + 420H_3 + 42H_5 + H_7)/128$
$H_4 = 16x^4 - 48x^2 + 12$	$x^6 = (120H_0 + 180H_2 + 30H_4 + H_6)/64$
$H_2 = 4x^2 - 2$	$x^5 = (60H_1 + 20H_3 + H_5)/32$
$H_3 = 8x^3 - 12x$	$x^4 = (12H_0 + 12H_2 + H_4)/16$
$H_5 = 32x^5 - 160x^3 + 120x$	$x^3 = (6H_1 + H_3)/8$
$H_6 = 64x^6 - 480x^4 + 720x^2 - 120$	$x^2 = (2H_0 + H_2)/4$
$H_7 = 128x^7 - 1344x^5 + 3360x^3 - 1680x$	$x = (H_1)/2$
$H_8 = 256x^8 - 3584x^6 + 13440x^4 - 13440x^2 + 1680$	$1 = H_0$

$L_0 = 1$	$x^6 = 720L_0 - 4320L_1 + 10800L_2 - 14400L_3 + 10800L_4 - 4320L_5 + 720L_6$
$L_1 = -x + 1$	$x^5 = 120L_0 - 600L_1 + 1200L_2 - 1200L_3 + 600L_4 - 120L_5$
$L_2 = (x^2 - 4x + 2)/2$	$x^4 = 24L_0 - 96L_1 + 144L_2 - 96L_3 + 24L_4$
$L_3 = (-x^3 + 9x^2 - 18x + 6)/6$	$x^3 = 6L_0 - 18L_1 + 18L_2 - 6L_3$
$L_4 = (x^4 - 16x^3 + 72x^2 - 96x + 24)/24$	$x^2 = 2L_0 - 4L_1 + 2L_2$
$L_5 = (-x^5 + 25x^4 - 200x^3 + 600x^2 - 600x + 120)/120$	$x = L_0 - L_1$
$L_6 = (x^6 - 36x^5 + 450x^4 - 2400x^3 + 5400x^2 - 4320x + 720)/720$	$1 = L_0$

$P_0 = 1$	$x^8 = (715P_0 + 2600P_2 + 2160P_4 + 832P_6 + 128P_8)/6435$
$P_1 = x$	$x^7 = (143P_1 + 182P_3 + 88P_5 + 16P_7)/429$
$P_2 = (3x^2 - 1)/2$	$x^6 = (33P_0 + 110P_2 + 72P_4 + 16P_6)/231$
$P_3 = (5x^3 - 3x)/2$	$x^5 = (27P_1 + 28P_3 + 8P_5)/63$
$P_4 = (35x^4 - 30x^2 + 3)/8$	$x^4 = (7P_0 + 20P_2 + 8P_4)/35$
$P_5 = (63x^5 - 70x^3 + 15x)/8$	$x^3 = (3P_1 + 2P_3)/5$
$P_6 = (231x^6 - 315x^4 + 105x^2 - 5)/16$	$x^2 = (P_0 + 2P_2)/3$
$P_7 = (429x^7 - 693x^5 + 315x^3 - 35x)/16$	$x = P_1$
$P_8 = (6435x^8 - 12012x^6 + 6930x^4 - 1260x^2 + 35)/128$	$1 = P_0$

$T_0 = 1$	$x^8 = (35T_0 + 56T_2 + 28T_4 + 8T_6 + T_8)/128$
$T_1 = x$	$x^7 = (35T_1 + 21T_3 + 7T_5 + T_7)/64$
$T_2 = 2x^2 - 1$	$x^6 = (10T_0 + 15T_2 + 6T_4 + T_6)/32$
$T_3 = 4x^3 - 3x$	$x^5 = (10T_1 + 5T_3 + T_5)/16$
$T_4 = 8x^4 - 8x^2 + 1$	$x^4 = (3T_0 + 4T_2 + T_4)/8$
$T_5 = 16x^5 - 20x^3 + 5x$	$x^3 = (3T_1 + T_3)/4$
$T_6 = 32x^6 - 48x^4 + 18x^2 - 1$	$x^2 = (T_0 + T_2)/2$
$T_7 = 64x^7 - 112x^5 + 56x^3 - 7x$	$x = T_1$
$T_8 = 128x^8 - 256x^6 + 160x^4 - 32x^2 + 1$	$1 = T_0$

$U_0 = 1$	$x^8 = (14U_0 + 28U_2 + 20U_4 + 7U_6 + U_8)/256$
$U_1 = 2x$	$x^7 = (14U_1 + 14U_3 + 6U_5 + U_7)/128$
$U_2 = 4x^2 - 1$	$x^6 = (5U_0 + 9U_2 + 5U_4 + U_6)/64$
$U_3 = 8x^3 - 4x$	$x^5 = (5U_1 + 4U_3 + U_5)/32$
$U_4 = 16x^4 - 12x^2 + 1$	$x^4 = (2U_0 + 3U_2 + U_4)/16$
$U_5 = 32x^5 - 32x^3 + 6x$	$x^3 = (2U_1 + U_3)/8$
$U_6 = 64x^6 - 80x^4 + 24x^2 - 1$	$x^2 = (U_0 + U_2)/4$
$U_7 = 128x^7 - 192x^5 + 80x^3 - 8x$	$x = (U_1)/2$
$U_8 = 256x^8 - 448x^6 + 240x^4 - 40x^2 + 1$	$1 = U_0$

6.13.10 ZERNIKE POLYNOMIALS

The *circle polynomials* or *Zernike polynomials* form a complete orthogonal set over the interior of the unit circle. They are $U_n^m(r,\theta) = R_n^m(r)e^{im\theta}$ where $R_n^m(r)$ are *radial polynomials*, n and m are integers with $n-|m|$ even, and $0 \le |m| \le n$.

1. Orthogonality $$\int_0^{2\pi}\int_0^1 \overline{U_n^m(r,\theta)}U_{n'}^{m'}(r,\theta)\,r\,dr\,d\theta = \frac{\pi}{n+1}\delta_{nn'}\delta_{mm'}$$
$$\int_0^1 R_n^m(r)R_{n'}^m(r)\,r\,dr = \frac{1}{2(n+1)}\delta_{nn'}.$$
2. Explicit formula for the radial polynomials
$$R_n^{\pm m}(r) = \frac{1}{\left(\frac{n-|m|}{2}\right)!\,r^{|m|}}\left\{\left(\frac{\partial}{\partial(r^2)}\right)^{\frac{n-|m|}{2}}\left[\left(r^2\right)^{\frac{n+|m|}{2}}\left(r^2-1\right)^{\frac{n-|m|}{2}}\right]\right\}$$
$$= \sum_{s=0}^{\frac{n-|m|}{2}} \frac{(-1)^s(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!}r^{n-2s}. \tag{6.13.2}$$
3. Expansions in Zernike polynomials
   (a) If $f(r,\theta)$ is a piecewise continuous function then
$$f(r,\theta) = \sum_{n=0}^{\infty}\sum_{m=-n}^{n} A_n^m U_n^m(r,\theta) \qquad \text{where } n-|m| \text{ is even.}$$
$$A_n^m = \overline{A_n^{-m}} = \frac{n+1}{\pi}\int_0^{2\pi}\int_0^1 \overline{U_n^m(r,\theta)}f(r,\theta)\,r\,dr\,d\theta.$$
   (b) If $f(r,\theta)$ is a real piecewise continuous function then
$$f(r,\theta) = \sum_{n=0}^{\infty}\sum_{m=0}^{n}\left[C_n^m\cos(m\theta) + S_n^m\sin(m\theta)\right]R_n^m(r)$$
where $n-|m|$ is even and $\epsilon_m = \begin{cases}1 & \text{if } m=0\\ 2 & \text{otherwise.}\end{cases}$
$$\begin{bmatrix}C_n^m\\ S_n^m\end{bmatrix} = \frac{\epsilon_m(n+1)}{\pi}\int_0^{2\pi}\int_0^1 f(r,\theta)R_n^m(r)\begin{bmatrix}\cos(m\theta)\\ \sin(m\theta)\end{bmatrix}r\,dr\,d\theta$$

n	$m=0$	2	4
0	1		
2	$2r^2-1$	r^2	
4	$6r^4-6r^2+1$	$4r^4-3r^2$	r^4
6	$20r^6-30r^4+12r^2-1$	$15r^6-20r^4+6r^2$	$6r^6-5r^4$

n	$m=1$	3	5
1	r		
3	$3r^3-2r$	r^3	
5	$10r^5-12r^3+3r$	$5r^5-4r^3$	r^5

6.13.11 SPHERICAL HARMONICS

The *spherical harmonics* are defined by

$$Y_{lm}(\theta,\phi) = \sqrt{\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}}P_l^m(\cos\theta)e^{im\phi} \tag{6.13.3}$$

for l an integer and $m = -l, -l+1, \dots, l-1, l$. They satisfy

$$\begin{aligned}
Y_{l,-m}(\theta,\phi) &= (-1)^m\overline{Y_{l,m}(\theta,\phi)},\\
Y_{l0}(\theta,\phi) &= \sqrt{\frac{2l+1}{4\pi}}P_l(\cos\theta),\\
Y_{lm}\left(\frac{\pi}{2},\phi\right) &= \begin{cases}\sqrt{\frac{(2l+1)(l-m)!(l+m)!}{4\pi}}\frac{(-1)^{(l+m)/2}e^{im\phi}}{2^l\left(\frac{l-m}{2}\right)!\left(\frac{l+m}{2}\right)!}, & \frac{l+m}{2}\text{ integral,}\\ 0, & \frac{l+m}{2}\text{ not integral.}\end{cases}
\end{aligned} \tag{6.13.4}$$

The normalization and orthogonality conditions are

$$\begin{aligned}
&\int_0^{2\pi} d\phi\int_0^{\pi}\sin\theta\,d\theta\,\overline{Y_{l'm'}(\theta,\phi)}Y_{lm}(\theta,\phi) = \delta_{ll'}\delta_{mm'},\\
&\int_0^{2\pi} d\phi\int_0^{\pi}\sin\theta\,d\theta\,\overline{Y_{l_1m_2}(\theta,\phi)}Y_{l_2m_2}(\theta,\phi)Y_{l_3m_3}(\theta,\phi),\\
&\qquad\qquad = \sqrt{\frac{(2l_2+1)(2l_3+1)}{4\pi(2l_1+1)}}\left(\begin{matrix}l_1 & l_3\\ m_2 & m_3\end{matrix}\middle|\begin{matrix}l_1\\ m_1\end{matrix}\right)\left(\begin{matrix}l_1 & l_3\\ 0 & 0\end{matrix}\middle|\begin{matrix}l_1\\ 0\end{matrix}\right),
\end{aligned} \tag{6.13.5}$$

where the terms on the right-hand side are Clebsch–Gordan coefficients.

Because of the (distributional) completeness relation,

$$\sum_{l=0}^{\infty}\sum_{m=-l}^{l}Y_{lm}(\theta,\phi)\overline{Y_{lm}(\theta',\phi')} = \delta(\phi-\phi')\delta(\cos\theta-\cos\theta'), \tag{6.13.6}$$

an arbitrary function $g(\theta,\phi)$ can be expanded in spherical harmonics as

$$g(\theta,\phi) = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}A_{lm}Y_{lm}(\theta,\phi), \qquad A_{lm} = \int\overline{Y_{lm}(\theta,\phi)}g(\theta,\phi)\,d\Omega. \tag{6.13.7}$$

Some spherical harmonics

- $l=0$ $\quad Y_{00} = \frac{1}{\sqrt{4\pi}}$
- $l=1$ $\quad Y_{11} = -\sqrt{\frac{3}{8\pi}}\sin\theta\,e^{i\phi}$ $\quad Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta$
- $l=2$ $\quad Y_{22} = \frac{1}{4}\sqrt{\frac{15}{2\pi}}\sin^2\theta\,e^{2i\phi}$ $\quad Y_{21} = -\sqrt{\frac{15}{8\pi}}\sin\theta\cos\theta\,e^{i\phi}$

  $Y_{20} = \frac{1}{2}\sqrt{\frac{5}{4\pi}}(3\cos^2\theta-1)$
- $l=3$ $\quad Y_{33} = -\frac{1}{4}\sqrt{\frac{35}{4\pi}}\sin^3\theta\,e^{3i\phi}$ $\quad Y_{32} = \frac{1}{4}\sqrt{\frac{105}{2\pi}}\sin^2\theta\cos\theta\,e^{2i\phi}$ $\quad \dots$

6.14 CLEBSCH–GORDAN COEFFICIENTS

The Clebsch–Gordan coefficients arise in the integration of three spherical harmonic functions (see Equation (6.13.5) on page 457).

$$\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix} = \delta_{m,m_1+m_2}\sqrt{\frac{(j_1+j_2-j)!(j+j_1-j_2)!(j+j_2-j_1)!(2j+1)}{(j+j_1+j_2+1)!}}$$
$$\times \sum_{0\le k<\infty} \frac{(-1)^k\sqrt{(j_1+m_1)!(j_1-m_1)!(j_2+m_2)!(j_2-m_2)!(j+m)!(j-m)!}}{k!(j_1+j_2-j-k)!(j_1-m_1-k)!(j_2+m_2-k)!(j-j_2+m_1+k)!(j-j_1-m_2+k)!}.$$

1. Conditions:
   (a) Each of $\{j_1, j_2, j, m_1, m_2, m\}$ may be an integer, or half an integer.
   (b) $j_1+j_2-j \ge 0$
   (c) $j_1-j_2+j \ge 0$
   (d) $-j_1+j_2+j \ge 0$
   (e) $j>0$, $j_1>0$, $j_2>0$
   (f) $j+j_1+j_2$ is an integer
   (g) j_1+m_1 is an integer
   (h) j_2+m_2 is an integer
   (i) $|m_1|\le j_1$, $|m_2|\le j_2$, $|m|\le j$
2. Special values:
   (a) $\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix} = 0$ if $m_1+m_2 \ne m$.
   (b) $\begin{pmatrix} j_1 & 0 & j \\ m_1 & 0 & m \end{pmatrix} = \delta_{j_1,j}\delta_{m_1,m}$.
   (c) $\begin{pmatrix} j_1 & j_2 & j \\ 0 & 0 & 0 \end{pmatrix} = 0$ when j_1+j_2+j is an odd integer.
   (d) $\begin{pmatrix} j_1 & j_1 & j \\ m_1 & m_1 & m \end{pmatrix} = 0$ when $2j_1+j$ is an odd integer.
3. Symmetry relations: all of the following are equal to $\begin{pmatrix} j_1 & j_2 & j \\ m_1 & m_2 & m \end{pmatrix}$:
   (a) $\begin{pmatrix} j_2 & j_1 & j \\ -m_2 & -m_1 & -m \end{pmatrix}$,
   (b) $(-1)^{j_1+j_2-j}\begin{pmatrix} j_2 & j_1 & j \\ m_1 & m_2 & m \end{pmatrix}$,
   (c) $(-1)^{j_1+j_2-j}\begin{pmatrix} j_1 & j_2 & j \\ -m_1 & -m_2 & -m \end{pmatrix}$,
   (d) $\sqrt{\frac{2j+1}{2j_1+1}}(-1)^{j_2+m_2}\begin{pmatrix} j & j_2 & j_1 \\ -m & m_2 & -m_1 \end{pmatrix}$,
   (e) $\sqrt{\frac{2j+1}{2j_1+1}}(-1)^{j_1-m_1+j-m}\begin{pmatrix} j & j_2 & j_1 \\ m & -m_2 & m_1 \end{pmatrix}$,
   (f) $\sqrt{\frac{2j+1}{2j_1+1}}(-1)^{j-m+j_1-m_1}\begin{pmatrix} j_2 & j & j_1 \\ m_2 & -m & -m_1 \end{pmatrix}$,
   (g) $\sqrt{\frac{2j+1}{2j_2+1}}(-1)^{j_1-m_1}\begin{pmatrix} j_1 & j & j_2 \\ m_1 & -m & -m_2 \end{pmatrix}$,
   (h) $\sqrt{\frac{2j+1}{2j_2+1}}(-1)^{j_1-m_1}\begin{pmatrix} j & j_1 & j_2 \\ m & -m_1 & m_2 \end{pmatrix}$.

By use of symmetry relations, Clebsch–Gordan coefficients $\begin{pmatrix} j_1 & j_2 & \vert & j \\ m_1 & m_2 & \vert & m \end{pmatrix}$ may be put in the standard form $j_1 \le j_2 \le j$ and $m \ge 0$.

j_1	j_2	m_1	m_2	j	m	$\begin{pmatrix} 1/2 & 1/2 & \vert & j \\ m_1 & m_2 & \vert & m \end{pmatrix}$
1/2	1/2	−1/2	−1/2	1	−1	1
1/2	1/2	−1/2	1/2	0	0	$-\left(\frac{1}{2}\right)^{1/2} = -0.70711$
1/2	1/2	−1/2	1/2	1	0	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1/2	1/2	1/2	−1/2	0	0	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1/2	1/2	1/2	−1/2	1	0	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1/2	1/2	1/2	1/2	1	1	1

j_1	j_2	m_1	m_2	j	m	$\begin{pmatrix} 1 & 1 & \vert & 1 \\ m_1 & m_2 & \vert & m \end{pmatrix}$
1	1	−1	0	1	−1	$-\left(\frac{1}{2}\right)^{1/2} = -0.70711$
1	1	−1	1	1	0	$-\left(\frac{1}{2}\right)^{1/2} = -0.70711$
1	1	0	−1	1	−1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	0	1	1	1	$-\left(\frac{1}{2}\right)^{1/2} = -0.70711$
1	1	1	−1	1	0	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	1	0	1	1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$

j_1	j_2	m_1	m_2	j	m	$\begin{pmatrix} 1 & 1 & \vert & 2 \\ m_1 & m_2 & \vert & m \end{pmatrix}$
1	1	−1	−1	2	−2	1
1	1	−1	0	2	−1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	−1	1	2	0	$\left(\frac{1}{6}\right)^{1/2} = 0.40825$
1	1	0	−1	2	−1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	0	0	2	0	$\left(\frac{2}{3}\right)^{1/2} = 0.81650$
1	1	0	1	2	1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	1	−1	2	0	$\left(\frac{1}{6}\right)^{1/2} = 0.40825$
1	1	1	0	2	1	$\left(\frac{1}{2}\right)^{1/2} = 0.70711$
1	1	1	1	2	2	1

j_1	j_2	m_1	m_2	j	m	$\begin{pmatrix} 1 & 1/2 & \vert & 2 \\ m_1 & m_2 & \vert & m \end{pmatrix}$
1	1/2	−1	−1/2	3/2	−3/2	1
1	1/2	−1	1/2	1/2	−1/2	$-\left(\frac{2}{3}\right)^{1/2} = -0.816497$
1	1/2	−1	1/2	3/2	−1/2	$\left(\frac{1}{3}\right)^{1/2} = 0.577350$
1	1/2	0	−1/2	3/2	−1/2	$\left(\frac{2}{3}\right)^{1/2} = 0.816497$
1	1/2	0	−1/2	1/2	−1/2	$\left(\frac{1}{3}\right)^{1/2} = 0.577350$
1	1/2	0	1/2	1/2	1/2	$-\left(\frac{1}{3}\right)^{1/2} = -0.577350$
1	1/2	0	1/2	3/2	1/2	$\left(\frac{2}{3}\right)^{1/2} = 0.816497$
1	1/2	1	−1/2	1/2	1/2	$\left(\frac{2}{3}\right)^{1/2} = 0.816497$
1	1/2	1	−1/2	3/2	1/2	$\left(\frac{1}{3}\right)^{1/2} = 0.577350$
1	1/2	1	1/2	3/2	3/2	1

6.15 BESSEL FUNCTIONS

6.15.1 DIFFERENTIAL EQUATION

The *Bessel differential equation*,

$$z^2y'' + zy' + (z^2 - \nu^2)y = 0.$$

The solutions are denoted with

$$J_\nu(z), \quad Y_\nu(z) \quad \text{(the ordinary \textit{Bessel functions})}$$

and

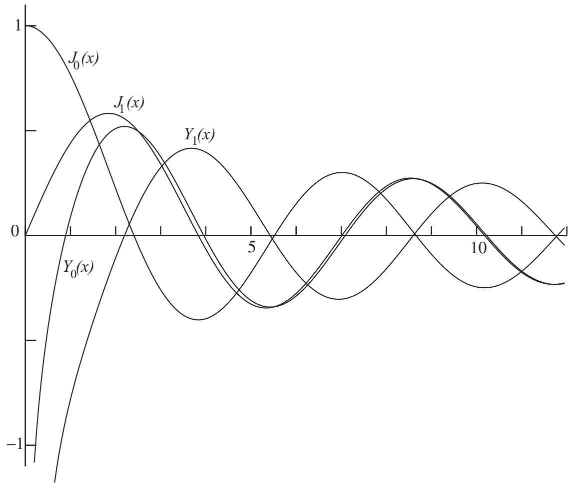
$$H_\nu^{(1)}(z), \quad H_\nu^{(2)}(z) \quad \text{(the \textit{Hankel functions})}.$$

Further solutions are

$$J_{-\nu}(z), \quad Y_{-\nu}(z), \quad H_{-\nu}^{(1)}(z), \quad H_{-\nu}^{(2)}(z).$$

When ν is an integer,

$$J_{-n}(z) = (-1)^n J_n(z), \quad n = 0, 1, 2, \ldots.$$



Bessel functions $J_0(x),\ J_1(x),\ Y_0(x),\ Y_1(x),\ 0 \le x \le 12.$ (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)

6.15.2 SINGULAR POINTS

The Bessel differential equation has a regular singularity at $z = 0$ and an irregular singularity at $z = \infty$.

6.15.3 RELATIONSHIPS

$$H_\nu^{(1)}(z) = J_\nu(z) + iY_\nu(z), \quad H_\nu^{(2)}(z) = J_\nu(z) - iY_\nu(z).$$

Neumann function: If $\nu \neq 0, \pm 1, \pm 2, \ldots$

$$Y_\nu(z) = \frac{\cos \nu\pi \; J_\nu(z) - J_{-\nu}(z)}{\sin \nu\pi}.$$

When $\nu = n$ (integer) then the limit $\nu \to n$ should be taken in the right-hand side of this equation. Complete solutions to Bessel's equation may be written as

$$\begin{array}{ll} c_1 J_\nu(z) + c_2 J_{-\nu}(z), & \text{if } \nu \text{ is not an integer,} \\ c_1 J_\nu(z) + c_2 Y_\nu(z), & \text{for any value of } \nu, \\ c_1 H_\nu^{(1)}(z) + c_2 H_\nu^{(2)}(z), & \text{for any value of } \nu. \end{array}$$

6.15.4 SERIES EXPANSIONS

For any complex z,

$$\begin{aligned} J_\nu(z) &= (\tfrac{1}{2}z)^\nu \sum_{n=0}^{\infty} \frac{(-1)^n (\tfrac{1}{2}z)^{2n}}{\Gamma(n+\nu+1)\, n!}, \\ J_0(z) &= 1 - (\tfrac{1}{2}z)^2 + \frac{1}{2!\,2!}(\tfrac{1}{2}z)^4 - \frac{1}{3!\,3!}(\tfrac{1}{2}z)^6 + \ldots, \\ J_1(z) &= \frac{1}{2} z \left[1 - \frac{1}{1!\,2!}(\tfrac{1}{2}z)^2 + \frac{1}{2!\,3!}(\tfrac{1}{2}z)^4 - \frac{1}{3!\,4!}(\tfrac{1}{2}z)^6 + \ldots \right], \\ Y_n(z) &= \frac{2}{\pi} J_n(z) \ln(\tfrac{1}{2}z) - \frac{(\tfrac{1}{2}z)^{-n}}{\pi} \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!} (\tfrac{1}{2}z)^{2k} - \\ & \quad \frac{(\tfrac{1}{2}z)^n}{\pi} \sum_{k=0}^{\infty} [\psi(k+1) + \psi(n+k+1)] \frac{(-1)^k (\tfrac{1}{2}z)^{2k}}{k!\,(n+k)!}, \end{aligned}$$

where ψ is the logarithmic derivative of the gamma function.

6.15.5 RECURRENCE RELATIONSHIPS

$$\begin{aligned} &C_{\nu-1}(z) + C_{\nu+1}(z) = \frac{2\nu}{z} C_\nu(z), \\ &C_{\nu-1}(z) - C_{\nu+1}(z) = 2C'_\nu(z), \\ &C'_\nu(z) = C_{\nu-1}(z) - \frac{\nu}{z} C_\nu(z), \\ &C'_\nu(z) = -C_{\nu+1}(z) + \frac{\nu}{z} C_\nu(z), \end{aligned}$$

where $C_\nu(z)$ denotes one of the functions $J_\nu(z),\ Y_\nu(z),\ H_\nu^{(1)}(z),\ H_\nu^{(2)}(z)$.

6.15.6 BEHAVIOR AS $z \to 0$

Let Re $\nu > 0$, then

$$J_\nu(z) \sim \frac{(\frac{1}{2}z)^\nu}{\Gamma(\nu+1)}, \qquad Y_\nu(z) \sim -\frac{1}{\pi}\Gamma(\nu)\left(\frac{2}{z}\right)^\nu,$$
$$H_\nu^{(1)}(z) \sim \frac{1}{\pi i}\Gamma(\nu)\left(\frac{2}{z}\right)^\nu, \qquad H_\nu^{(2)}(z) \sim -\frac{1}{\pi i}\Gamma(\nu)\left(\frac{2}{z}\right)^\nu.$$

The same relations hold as Re $\nu \to \infty$, with z fixed.

6.15.7 INTEGRAL REPRESENTATIONS

Let Re $z > 0$ and ν be any complex number.

$$\begin{aligned}
J_\nu(z) &= \frac{1}{\pi}\int_0^\pi \cos(\nu\theta - z\sin\theta)\,d\theta - \frac{\sin\nu\pi}{\pi}\int_0^\infty e^{-\nu t - z\sinh t}\,dt \\
&= \frac{(z/2)^\nu}{\sqrt{\pi}\Gamma(\nu+\frac{1}{2})}\int_{-1}^1 \left(1-t^2\right)^{\nu-\frac{1}{2}}\cos zt\,dt, \quad \text{Re}\,\nu > -\tfrac{1}{2},\ z \text{ complex}, z \neq 0, \\
&= \frac{2\left(\frac{1}{2}x\right)^{-\nu}}{\sqrt{\pi}\Gamma\left(\frac{1}{2}-\nu\right)}\int_1^\infty \frac{\sin xt}{\left(t^2-1\right)^{\nu+\frac{1}{2}}}\,dt, \qquad x > 0,\ |\text{Re}\,\nu| < -\tfrac{1}{2}, \\
Y_\nu(z) &= \frac{1}{\pi}\int_0^\pi \sin(z\sin\theta - \nu\theta)\,d\theta - \int_0^\infty \left(e^{\nu t} + e^{-\nu t}\cos\nu\pi\right)e^{-z\sinh t}\,dt \\
&= -\frac{2\left(\frac{1}{2}x\right)^{-\nu}}{\sqrt{\pi}\Gamma\left(\frac{1}{2}-\nu\right)}\int_1^\infty \frac{\cos xt}{\left(t^2-1\right)^{\nu+\frac{1}{2}}}\,dt, \qquad x > 0,\ |\text{Re}\,\nu| < -\tfrac{1}{2}.
\end{aligned}$$

When $\nu = n$ (an integer), the second integral in the first relation disappears.

6.15.8 FOURIER EXPANSION

For any complex z,

$$e^{-iz\sin t} = \sum_{n=-\infty}^{\infty} e^{-int} J_n(z),$$

with Parseval relation

$$\sum_{n=-\infty}^{\infty} J_n^2(z) = 1.$$

6.15.9 AUXILIARY FUNCTIONS

Let $\chi = z - (\frac{1}{2}\nu + \frac{1}{4})\pi$ and define

$$\begin{aligned}
P(\nu, z) &= \sqrt{\pi z/2}[\quad J_\nu(z)\cos\chi + Y_\nu(z)\sin\chi], \\
Q(\nu, z) &= \sqrt{\pi z/2}[-J_\nu(z)\sin\chi + Y_\nu(z)\cos\chi].
\end{aligned}$$

6.15.10 INVERSE RELATIONSHIPS

$$J_\nu(z) = \sqrt{2/(\pi z)}\,[P(\nu,z)\cos\chi - Q(\nu,z)\sin\chi],$$
$$Y_\nu(z) = \sqrt{2/(\pi z)}\,[P(\nu,z)\sin\chi + Q(\nu,z)\cos\chi].$$

For the Hankel functions,

$$H_\nu^{(1)}(z) = \sqrt{2/(\pi z)}\,[P(\nu,z) + iQ(\nu,z)]e^{i\chi},$$
$$H_\nu^{(2)}(z) = \sqrt{2/(\pi z)}\,[P(\nu,z) - iQ(\nu,z)]e^{-i\chi}.$$

The functions $P(\nu,z), Q(\nu,z)$ are the slowly varying components in the asymptotic expansions of the oscillatory Bessel and Hankel functions.

6.15.11 ASYMPTOTIC EXPANSIONS

Let (α,n) be defined by

$$\begin{aligned}(\alpha,n) &= \frac{2^{-2n}}{n!}\{(4\alpha^2-1)(4\alpha^2-3^2)\cdots(4\alpha^2-(2n-1)^2)\}\\ &= \frac{\Gamma(\frac12+\alpha+n)}{n!\,\Gamma(\frac12+\alpha-n)}, \qquad n=0,1,2,\ldots,\\ &= \frac{(-1)^n\cos(\pi\alpha)}{\pi n!}\Gamma(\tfrac12+\alpha+n)\Gamma(\tfrac12-\alpha+n),\end{aligned}$$

with recursion

$$(\alpha,n+1) = -\frac{(n+\frac12)^2-\alpha^2}{n+1}\,(\alpha,n), \quad n=1,2,3,\ldots, \quad (\alpha,0)=1.$$

Then, for $z\to\infty$,

$$P(\nu,z) \sim \sum_{n=0}^{\infty}(-1)^n\frac{(\nu,2n)}{(2z)^{2n}}, \quad Q(\nu,z) \sim \sum_{n=0}^{\infty}(-1)^n\frac{(\nu,2n+1)}{(2z)^{2n+1}}.$$

With $\mu = 4\nu^2$,

$$P(\nu,z) \sim 1 - \frac{(\mu-1)(\mu-9)}{2!\,(8z)^2} + \frac{(\mu-1)(\mu-9)(\mu-25)(\mu-49)}{4!\,(8z)^4} - \ldots,$$
$$Q(\nu,z) \sim \frac{\mu-1}{8z} - \frac{(\mu-1)(\mu-9)(\mu-25)}{3!\,(8z)^3} + \ldots .$$

For large positive values of x,

$$J_\nu(x) = \sqrt{2/(\pi x)}\,\left[\cos(x-\frac12\,\nu\pi-\frac14\,\pi)+O(x^{-1})\right],$$
$$Y_\nu(x) = \sqrt{2/(\pi x)}\,\left[\sin(x-\frac12\,\nu\pi-\frac14\,\pi)+O(x^{-1})\right].$$

6.15.12 ZEROS OF BESSEL FUNCTIONS

For $\nu \geq 0$, the zeros $j_{\nu,k}$ (and $y_{\nu,k}$) of $J_\nu(x)$ (and $Y_\nu(x)$) can be arranged as sequences

$$0 < j_{\nu,1} < j_{\nu,2} < \cdots < j_{\nu,n} < \cdots, \qquad \lim_{n\to\infty} j_{\nu,n} = \infty,$$
$$0 < y_{\nu,1} < y_{\nu,2} < \cdots < y_{\nu,n} < \cdots, \qquad \lim_{n\to\infty} y_{\nu,n} = \infty.$$

Between two consecutive positive zeros of $J_\nu(x)$, there is exactly one zero of $J_{\nu+1}(x)$. Conversely, between two consecutive positive zeros of $J_{\nu+1}(x)$, there is exactly one zero of $J_\nu(x)$. The same holds for the zeros of $Y_\nu(z)$. Moreover, between each pair of consecutive positive zeros of $J_\nu(x)$, there is exactly one zero of $Y_\nu(x)$, and conversely.

6.15.12.1 Asymptotics of the zeros

When ν is fixed, $s \gg \nu$, and $\mu = 4\nu^2$,

$$j_{\nu,s} \sim \alpha - \frac{\mu-1}{8\alpha}\left[1 - \frac{4(7\mu^2-31)}{3(8\alpha)^2} - \frac{32(83\mu^2-982\mu+3779)}{15(8\alpha)^4} + \dots\right]$$

where $\alpha = (s + \frac{1}{2}\nu - \frac{1}{4})\pi$; $y_{\nu,s}$ has the same asymptotic expansion with $\alpha = (s + \frac{1}{2}\nu - \frac{3}{4})\pi$.

n	$j_{0,n}$	$j_{1,n}$	$y_{0,n}$	$y_{1,n}$
1	2.40483	3.83171	0.89358	2.19714
2	5.52008	7.01559	3.95768	5.42968
3	8.65373	10.17347	7.08605	8.59601
4	11.79153	13.32369	10.22235	11.74915
5	14.93092	16.47063	13.36110	14.89744
6	18.07106	19.61586	16.50092	18.04340
7	21.21164	22.76008	19.64131	21.18807

Positive zeros $j_{\nu,n}, y_{\nu,n}$ of Bessel functions $J_\nu(x), Y_\nu(x),\ \nu = 0, 1$.

6.15.13 HALF ORDER BESSEL FUNCTIONS

For integer values of n, let

$$j_n(z) = \sqrt{\pi/(2z)}\, J_{n+\frac{1}{2}}(z), \quad y_n(z) = \sqrt{\pi/(2z)}\, Y_{n+\frac{1}{2}}(z).$$

Then

$$j_0(z) = y_{-1}(z) = \frac{\sin z}{z}, \quad y_0(z) = -j_{-1}(z) = -\frac{\cos z}{z},$$

and, for $n = 0, 1, 2, \ldots,$

$$j_n(z) = (-z)^n \left[\frac{1}{z}\frac{d}{dz}\right]^n \frac{\sin z}{z}, \quad y_n(z) = -(-z)^n \left[\frac{1}{z}\frac{d}{dz}\right]^n \frac{\cos z}{z}.$$

- Recursion relationships: The functions $j_n(z), y_n(z)$ both satisfy

$$z[f_{n-1}(z) + f_{n+1}(z)] = (2n+1)f_n(z),$$
$$nf_{n-1}(z) - (n+1)f_{n+1}(z) = (2n+1)f_n'(z).$$

- Differential equation

$$z^2 f'' + 2zf' + [z^2 - n(n+1)]f = 0.$$

6.15.14 MODIFIED BESSEL FUNCTIONS

1. Differential equation

$$z^2 y'' + zy' - (z^2 + \nu^2)y = 0.$$

2. Solutions $I_\nu(z), K_\nu(z)$,

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{n=0}^{\infty} \frac{(z/2)^{2n}}{\Gamma(n+\nu+1)\, n!},$$
$$K_\nu(z) = \frac{\pi}{2} \frac{I_{-\nu}(z) - I_\nu(z)}{\sin \nu\pi},$$

where the right-hand side should be determined by l'Hôpital's rule when ν assumes integer values. When $n = 0, 1, 2, \ldots,$

$$K_n(z) = (-1)^{n+1} I_n(z) \ln \frac{z}{2} + \frac{1}{2}\left(\frac{2}{z}\right)^n \sum_{k=0}^{n-1} \frac{(n-k-1)!}{k!}\left(-\frac{z^2}{4}\right)^k$$
$$+ \frac{(-1)^n}{2}\left(\frac{z}{2}\right)^n \sum_{k=0}^{\infty} [\psi(k+1) + \psi(n+k+1)] \frac{(z/2)^{2k}}{k!\,(n+k)!}.$$

3. Relations with the ordinary Bessel functions

$$I_\nu(z) = e^{-\frac{1}{2}\nu\pi i} J_\nu\left(z e^{\frac{1}{2}\pi i}\right), \qquad -\pi < \arg z \le \frac{\pi}{2},$$
$$I_\nu(z) = e^{\frac{3}{2}\nu\pi i} J_\nu\left(z e^{-\frac{3}{2}\pi i}\right), \qquad \frac{\pi}{2} < \arg z \le \pi,$$
$$K_\nu(z) = \frac{\pi}{2} i e^{\frac{1}{2}\nu\pi i} H_\nu^{(1)}\left(z e^{\frac{1}{2}\pi i}\right), \qquad -\pi < \arg z \le \frac{\pi}{2},$$
$$K_\nu(z) = -\frac{\pi}{2} i e^{-\frac{1}{2}\nu\pi i} H_\nu^{(2)}\left(z e^{-\frac{1}{2}\pi i}\right), \qquad -\frac{\pi}{2} < \arg z \le \pi,$$
$$Y_\nu\left(z e^{\frac{1}{2}\pi i}\right) = e^{\frac{1}{2}(\nu+1)\pi i} I_\nu(z) - \frac{2}{\pi} e^{-\frac{1}{2}\nu\pi i} K_\nu(z), \qquad -\pi < \arg z \le \frac{\pi}{2}.$$

For $n = 0, 1, 2, \ldots,$

$$I_n(z) = i^{-n} J_n(iz), \quad Y_n(iz) = i^{n+1} I_n(z) - \frac{2}{\pi}\, i^{-n} K_n(z),$$
$$I_{-n}(z) = I_n(z), \quad K_{-\nu}(z) = K_\nu(z), \quad \text{for any } \nu.$$

4. Recursion relationships

$$I_{\nu-1}(z) - I_{\nu+1}(z) = \frac{2\nu}{z} I_\nu(z), \qquad K_{\nu+1}(z) - K_{\nu-1}(z) = \frac{2\nu}{z} K_\nu(z),$$
$$I_{\nu-1}(z) + I_{\nu+1}(z) = 2I'_\nu(z), \qquad K_{\nu-1}(z) + K_{\nu+1}(z) = -2K'_\nu(z).$$

5. Integral Representations

$$\begin{aligned} I_\nu(z) &= \frac{1}{\pi}\int_0^\pi e^{z\cos\theta}\cos(\nu\theta)\,d\theta - \frac{\sin\nu\pi}{\pi}\int_0^\infty e^{-\nu t - z\cosh t}\,dt \\ &= \frac{(2z)^\nu e^z}{\sqrt{\pi}\Gamma\left(\nu+\frac{1}{2}\right)}\int_0^1 e^{-2zt}[t(1-t)]^{\nu-\frac{1}{2}}\,dt, \\ &\qquad \text{Re}\,\nu > -\tfrac{1}{2},\ z \text{ complex}, z \neq 0. \end{aligned}$$

$$\begin{aligned} K_\nu(z) &= \int_0^\infty e^{-z\cosh t}\cosh(\nu t)\,dt \\ &= \frac{\sqrt{\pi}\left(\frac{2}{z}\right)^\nu e^{-z}}{\Gamma\left(\nu+\frac{1}{2}\right)}\int_0^\infty e^{-2zt} t^{\nu-\frac{1}{2}}(t+1)^{\nu-\frac{1}{2}}\,dt, \\ &\qquad \text{Re}\,\nu > -\tfrac{1}{2},\ \text{Re}\,z > 0, \end{aligned}$$

$$K_\nu(xz) = \frac{\Gamma\left(\nu+\frac{1}{2}\right)(2z)^\nu}{\sqrt{\pi}x^\nu}\int_0^\infty \frac{\cos xt\,dt}{(t^2+z^2)^{\nu+\frac{1}{2}}}\,dt,$$
$$\text{Re}\,\nu > -\tfrac{1}{2},\ x > 0,\ |\arg z| < \tfrac{1}{2}\pi.$$

When $\nu = n$ (an integer), the second integral in the first relation disappears.

6.15.15 AIRY FUNCTIONS

1. Differential equation: $y'' - zy = 0.$
2. Solutions are $\text{Ai}(z)$ and $\text{Bi}(z)$:

$$\text{Ai}(z) = c_1 f(z) - c_2 g(z),$$
$$\text{Bi}(z) = \sqrt{3}\,[c_1 f(z) + c_2 g(z)]$$

where

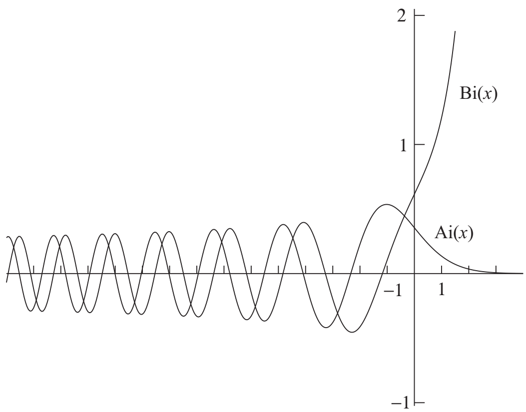
$$\begin{aligned} f(z) &= 1 + \frac{1}{3!}z^3 + \frac{1\cdot 4}{6!}z^6 + \frac{1\cdot 4\cdot 7}{9!}z^9 + \dots, \\ g(z) &= z + \frac{2}{4!}z^4 + \frac{2\cdot 5}{7!}z^7 + \frac{2\cdot 5\cdot 8}{10!}z^{10} + \dots, \\ c_1 &= \text{Ai}(0) = \frac{\text{Bi}(0)}{\sqrt{3}} = \frac{3^{-2/3}}{\Gamma(\frac{2}{3})} = 0.35502\,80538\,87817\,, \\ c_2 &= -\text{Ai}'(0) = \frac{\text{Bi}'(0)}{\sqrt{3}} = \frac{3^{-1/3}}{\Gamma(\frac{1}{3})} = 0.25881\,94037\,92807\,. \end{aligned}$$

3. Wronskian relation:

$$\text{Ai}(z)\,\text{Bi}'(z) - \text{Ai}'(z)\,\text{Bi}(z) = \frac{1}{\pi}.$$

FIGURE 6.6

Graphs of the Airy functions $\mathrm{Ai}(x)$ *and* $\mathrm{Bi}(x)$, x *real.* (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)



4. Relations with the Bessel functions: Let $\zeta = \frac{2}{3}z^{\frac{3}{2}}$, then

$$\mathrm{Ai}(z) = \frac{1}{3}\sqrt{z}\left[I_{-\frac{1}{3}}(\zeta) - I_{\frac{1}{3}}(\zeta)\right] = \frac{1}{\pi}\sqrt{\frac{z}{3}}K_{\frac{1}{3}}(\zeta).$$

$$\mathrm{Ai}(-z) = \frac{1}{3}\sqrt{z}\left[J_{\frac{1}{3}}(\zeta) + J_{-\frac{1}{3}}(\zeta)\right].$$

$$\mathrm{Bi}(z) = \sqrt{\frac{z}{3}}\left[I_{-\frac{1}{3}}(\zeta) + I_{\frac{1}{3}}(\zeta)\right].$$

$$\mathrm{Bi}(-z) = \sqrt{\frac{z}{3}}\left[J_{-\frac{1}{3}}(\zeta) - J_{\frac{1}{3}}(\zeta)\right].$$

5. Asymptotic behavior: Let $\zeta = \frac{2}{3}z^{\frac{3}{2}}$. Then, for $z \to \infty$,

$$\mathrm{Ai}(z) = \frac{1}{2\sqrt{\pi}}z^{-\frac{1}{4}}e^{-\zeta}\left[1 + O\left(\zeta^{-1}\right)\right], \qquad |\arg z| < \pi,$$

$$\mathrm{Bi}(z) = \frac{1}{\sqrt{\pi}}z^{-\frac{1}{4}}e^{\zeta}\left[1 + O\left(\zeta^{-1}\right)\right], \qquad |\arg z| < \tfrac{1}{3}\pi,$$

$$\mathrm{Ai}(-z) = \frac{1}{\sqrt{\pi}}z^{-\frac{1}{4}}\left[\sin\left(\zeta + \tfrac{1}{4}\pi\right) + O\left(\zeta^{-1}\right)\right], \qquad |\arg z| < \tfrac{2}{3}\pi,$$

$$\mathrm{Bi}(-z) = -\frac{1}{\sqrt{\pi}}z^{-\frac{1}{4}}\left[\cos\left(\zeta + \tfrac{1}{4}\pi\right) + O\left(\zeta^{-1}\right)\right], \qquad |\arg z| < \tfrac{2}{3}\pi.$$

6. Integrals for real x:

$$\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos\left(\tfrac{1}{3}t^3 + xt\right)\,dt,$$

$$\mathrm{Bi}(x) = \frac{1}{\pi}\int_0^\infty e^{-\frac{1}{3}t^3 + xt}\,dt + \frac{1}{\pi}\int_0^\infty \sin\left(\tfrac{1}{3}t^3 + xt\right)\,dt.$$

6.15.16 NUMERICAL VALUES FOR THE BESSEL FUNCTIONS

x	$J_0(x)$	$J_1(x)$	$Y_0(x)$	$Y_1(x)$
0.0	1.00000000	0.00000000	$-\infty$	$-\infty$
0.2	0.99002497	0.09950083	−1.08110532	−3.32382499
0.4	0.96039823	0.19602658	−0.60602457	−1.78087204
0.6	0.91200486	0.28670099	−0.30850987	−1.26039135
0.8	0.84628735	0.36884205	−0.08680228	−0.97814418
1.0	0.76519769	0.44005059	0.08825696	−0.78121282
1.2	0.67113274	0.49828906	0.22808350	−0.62113638
1.4	0.56685512	0.54194771	0.33789513	−0.47914697
1.6	0.45540217	0.56989594	0.42042690	−0.34757801
1.8	0.33998641	0.58151695	0.47743171	−0.22366487
2.0	0.22389078	0.57672481	0.51037567	−0.10703243
2.2	0.11036227	0.55596305	0.52078429	0.00148779
2.4	0.00250768	0.52018527	0.51041475	0.10048894
2.6	−0.09680495	0.47081827	0.48133059	0.18836354
2.8	−0.18503603	0.40970925	0.43591599	0.26354539
3.0	−0.26005195	0.33905896	0.37685001	0.32467442
4.0	−0.39714981	−0.06604333	−0.01694074	0.39792571
5.0	−0.17759677	−0.32757914	−0.30851763	0.14786314

x	$e^{-x}I_0(x)$	$e^{-x}I_1(x)$	$e^{x}K_0(x)$	$e^{x}K_1(x)$
0.0	1.00000000	0.00000000	∞	∞
0.2	0.82693855	0.08228312	2.14075732	5.83338603
0.4	0.69740217	0.13676322	1.66268209	3.25867388
0.6	0.59932720	0.17216442	1.41673762	2.37392004
0.8	0.52414894	0.19449869	1.25820312	1.91793030
1.0	0.46575961	0.20791042	1.14446308	1.63615349
1.2	0.41978208	0.21525686	1.05748453	1.44289755
1.4	0.38306252	0.21850759	0.98807000	1.30105374
1.6	0.35331500	0.21901949	0.93094598	1.19186757
1.8	0.32887195	0.21772628	0.88283353	1.10480537
2.0	0.30850832	0.21526929	0.84156822	1.03347685
2.2	0.29131733	0.21208773	0.80565398	0.97377017
2.4	0.27662232	0.20848109	0.77401814	0.92291367
2.6	0.26391400	0.20465225	0.74586824	0.87896728
2.8	0.25280553	0.20073741	0.72060413	0.84053006
3.0	0.24300035	0.19682671	0.69776160	0.80656348
4.0	0.20700192	0.17875084	0.60929767	0.68157595
5.0	0.18354081	0.16397227	0.54780756	0.60027386

6.16 BETA FUNCTION

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt, \quad \text{Re}\, p > 0, \quad \text{Re}\, q > 0. \tag{6.16.1}$$

1. Relations:

$$B(p,q) = B(q,p), \qquad B(p,q+1) = \frac{q}{p}\,B(p+1,q) = \frac{q}{p+q}\,B(p,q).$$

2. Relation with the gamma function:

$$B(p,q) = \frac{\Gamma(p)\,\Gamma(q)}{\Gamma(p+q)}, \qquad B(p,q)\,B(p+q,r) = \frac{\Gamma(p)\,\Gamma(q)\,\Gamma(r)}{\Gamma(p+q+r)}.$$

3. Other integrals: (in all cases Re $p > 0$ and Re $q > 0$)

$$\begin{aligned} B(p,q) &= 2\int_0^{\pi/2} \sin^{2p-1}\theta\, \cos^{2q-1}\theta\, d\theta \\ &= \int_0^\infty \frac{t^{p-1}}{(t+1)^{p+q}}\,dt = \int_0^\infty e^{-pt}\left(1-e^{-t}\right)^{q-1}\,dt \\ &= r^q(r+1)^p \int_0^1 \frac{t^{p-1}(1-t)^{q-1}}{(r+t)^{p+q}}\,dt, \quad r > 0. \end{aligned}$$

6.16.1 NUMERICAL VALUES OF THE BETA FUNCTION

p	$q = 0.100$	0.200	0.300	0.400	0.500	0.600	0.700	0.800	0.900	1.000
0.1	19.715	14.599	12.831	11.906	11.323	10.914	10.607	10.365	10.166	10.000
0.2	14.599	9.502	7.748	6.838	6.269	5.872	5.576	5.345	5.157	5.000
0.3	12.831	7.748	6.010	5.112	4.554	4.169	3.883	3.661	3.482	3.333
0.4	11.906	6.838	5.112	4.226	3.679	3.303	3.027	2.813	2.641	2.500
0.5	11.323	6.269	4.554	3.679	3.142	2.775	2.506	2.299	2.135	2.000
0.6	10.914	5.872	4.169	3.303	2.775	2.415	2.154	1.954	1.796	1.667
0.7	10.607	5.576	3.883	3.027	2.506	2.154	1.899	1.705	1.552	1.429
0.8	10.365	5.345	3.661	2.813	2.299	1.954	1.705	1.517	1.369	1.250
0.9	10.166	5.157	3.482	2.641	2.135	1.796	1.552	1.369	1.226	1.111
1.0	10.000	5.000	3.333	2.500	2.000	1.667	1.429	1.250	1.111	1.000
1.2	9.733	4.751	3.099	2.279	1.791	1.468	1.239	1.069	0.938	0.833
1.4	9.525	4.559	2.921	2.113	1.635	1.321	1.101	0.938	0.813	0.714
1.6	9.355	4.404	2.779	1.982	1.513	1.208	0.994	0.837	0.718	0.625
1.8	9.213	4.276	2.663	1.875	1.415	1.117	0.909	0.758	0.644	0.556
2.0	9.091	4.167	2.564	1.786	1.333	1.042	0.840	0.694	0.585	0.500
2.2	8.984	4.072	2.480	1.710	1.264	0.979	0.783	0.641	0.536	0.455
2.4	8.890	3.989	2.406	1.644	1.205	0.925	0.734	0.597	0.495	0.417
2.6	8.805	3.915	2.340	1.586	1.153	0.878	0.692	0.558	0.460	0.385
2.8	8.728	3.848	2.282	1.534	1.107	0.837	0.655	0.525	0.430	0.357
3.0	8.658	3.788	2.230	1.488	1.067	0.801	0.622	0.496	0.403	0.333

6.17 ELLIPTIC INTEGRALS

Any integral of the type $\int R(x,y)\,dx$, where $R(x,y)$ is a rational function of x and y, with y^2 being a polynomial of the third or fourth degree in x (that is $y^2 = a_0x^4 + a_1x^3 + a_2x^2 + a_3x + a_4$ with $|a_0| + |a_1| > 0$) is called an *elliptic integral*. All elliptic integrals can be reduced to three basic types:

1. *Elliptic integral of the first kind*

$$\begin{aligned} F(\phi,k) &= \int_0^\phi \frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \\ &= \int_0^x \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}, \quad x = \sin\phi, \quad k^2 < 1. \end{aligned}$$

2. *Elliptic integral of the second kind*

$$\begin{aligned} E(\phi,k) &= \int_0^\phi \sqrt{1-k^2\sin^2\theta}\,d\theta \\ &= \int_0^x \frac{\sqrt{1-k^2t^2}}{\sqrt{1-t^2}}\,dt, \quad x = \sin\phi, \quad k^2 < 1. \end{aligned}$$

3. *Elliptic integral of the third kind*

$$\begin{aligned} \Pi(n;\phi,k) &= \int_0^\phi \frac{1}{1+n\sin^2\theta}\frac{d\theta}{\sqrt{1-k^2\sin^2\theta}} \\ &= \int_0^x \frac{1}{1+nt^2}\frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}, \quad x = \sin\phi, \quad k^2 < 1. \end{aligned}$$

where for $n < -1$ the integral should be interpreted as a Cauchy principal value integral.

6.17.1 COMPLETE ELLIPTIC INTEGRALS

1. The *complete elliptic integrals* of the first and second kinds are

$$\begin{aligned} K = & K(k) = F\left(\frac{\pi}{2},k\right) = \int_0^{\pi/2} \left(1-k^2\sin^2 t\right)^{-1/2} dt = \frac{\pi}{2}F\left(\frac{1}{2},\frac{1}{2};1;k^2\right), \\ & K(\alpha) = \int_0^{\pi/2} \left(1-\sin^2\alpha\sin^2 t\right)^{-1/2} dt, \\ E = & E(k) = E\left(\frac{\pi}{2},k\right) = \int_0^{\pi/2} \left(1-k^2\sin^2 t\right)^{1/2} dt = \frac{\pi}{2}F\left(-\frac{1}{2},\frac{1}{2};1;k^2\right), \\ & E(\alpha) = \int_0^{\pi/2} \left(1-\sin^2\alpha\sin^2 t\right)^{1/2} dt, \end{aligned}$$

where $F\left(\pm\frac{1}{2},\frac{1}{2};1;k^2\right)$ is the Gauss hypergeometric function.

2. *Complementary integrals*

   In these expressions, primes do not mean derivatives. k is called the *modulus*, $k' = \sqrt{1-k^2}$ is called the *complementary modulus.*

$$K' = K'(k) = K(k') = \int_0^{\pi/2} \left(1-(1-k^2)\sin^2 t\right)^{-1/2}\, dt = F\left(\frac{\pi}{2}, k'\right),$$
$$E' = E'(k) = E(k') = \int_0^{\pi/2} \left(1-(1-k^2)\sin^2 t\right)^{1/2}\, dt = E\left(\frac{\pi}{2}, k'\right),$$

3. The *Legendre relation* is,

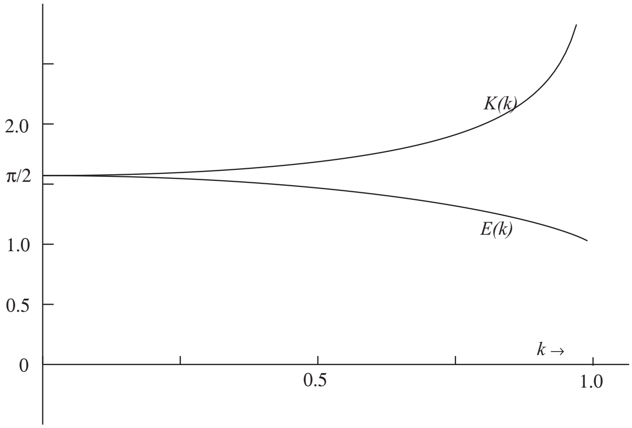
$$K\,E' + E\,K' - K\,K' = \frac{\pi}{2}.$$

4. Extension of the range of ϕ

$$F(\pi, k) = 2K, \qquad E(\pi, k) = 2E,$$

   and, for $m = 0, 1, 2, \ldots,$

$$F(\phi + m\pi, k) = mF(\pi, k) + F(\phi, k) = 2mK + F(\phi, k),$$
$$E(\phi + m\pi, k) = mE(\pi, k) + E(\phi, k) = 2mE + E(\phi, k).$$



The complete elliptic integrals $E(k)$ and $K(k)$, $0 \le k \le 1$. (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)

6.17.2 NUMERICAL VALUES OF THE ELLIPTIC INTEGRALS

$$F(\phi,\alpha) = \int_0^\phi \left(1-\sin^2\alpha\sin^2 t\right)^{-1/2}\,dt \qquad \text{(note that } k=\sin\alpha)$$

ϕ \ α	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
0°	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10°	0.1745	0.1746	0.1746	0.1748	0.1749	0.1751	0.1752	0.1753	0.1754	0.1754
20°	0.3491	0.3493	0.3499	0.3508	0.3520	0.3533	0.3545	0.3555	0.3561	0.3564
30°	0.5236	0.5243	0.5263	0.5294	0.5334	0.5379	0.5422	0.5459	0.5484	0.5493
40°	0.6981	0.6997	0.7043	0.7116	0.7213	0.7323	0.7436	0.7535	0.7604	0.7629
50°	0.8727	0.8756	0.8842	0.8982	0.9173	0.9401	0.9647	0.9876	1.0044	1.0107
60°	1.0472	1.0519	1.0660	1.0896	1.1226	1.1643	1.2126	1.2619	1.3014	1.3170
70°	1.2217	1.2286	1.2495	1.2853	1.3372	1.4068	1.4944	1.5959	1.6918	1.7354
80°	1.3963	1.4056	1.4344	1.4846	1.5597	1.6660	1.8125	2.0119	2.2653	2.4362
90°	1.5708	1.5828	1.6200	1.6858	1.7868	1.9356	2.1565	2.5046	3.1534	∞

$$E(\phi,\alpha) = \int_0^\phi \left(1-\sin^2\alpha\sin^2 t\right)^{1/2}\,dt \qquad \text{(note that } k=\sin\alpha)$$

ϕ \ α	0°	10°	20°	30°	40°	50°	60°	70°	80°	90°
0°	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
10°	0.1745	0.1745	0.1744	0.1743	0.1742	0.1740	0.1739	0.1738	0.1737	0.1736
20°	0.3491	0.3489	0.3483	0.3473	0.3462	0.3450	0.3438	0.3429	0.3422	0.3420
30°	0.5236	0.5229	0.5209	0.5179	0.5141	0.5100	0.5061	0.5029	0.5007	0.5000
40°	0.6981	0.6966	0.6921	0.6851	0.6763	0.6667	0.6575	0.6497	0.6446	0.6428
50°	0.8727	0.8698	0.8614	0.8483	0.8317	0.8134	0.7954	0.7801	0.7697	0.7660
60°	1.0472	1.0426	1.0290	1.0076	0.9801	0.9493	0.9184	0.8914	0.8728	0.8660
70°	1.2217	1.2149	1.1949	1.1632	1.1221	1.0750	1.0266	0.9830	0.9514	0.9397
80°	1.3963	1.3870	1.3597	1.3161	1.2590	1.1926	1.1225	1.0565	1.0054	0.9848
90°	1.5708	1.5589	1.5238	1.4675	1.3931	1.3055	1.2111	1.1184	1.0401	1.0000

α	$K(\alpha)$	$K'(\alpha)$	$E(\alpha)$	$E'(\alpha)$
0°	$\pi/2$	∞	$\pi/2$	1
5°	1.574	3.832	1.568	1.013
10°	1.583	3.153	1.559	1.040
15°	1.598	2.768	1.544	1.076
20°	1.620	2.505	1.524	1.118
25°	1.649	2.309	1.498	1.164
30°	1.686	2.157	1.467	1.211
35°	1.731	2.035	1.432	1.259
40°	1.787	1.936	1.393	1.306
45°	1.854	1.854	1.351	1.351
50°	1.936	1.787	1.306	1.393
60°	2.157	1.686	1.211	1.467
70°	2.505	1.620	1.118	1.524
80°	3.153	1.583	1.040	1.559
90°	∞	$\pi/2$	1	$\pi/2$

k^2	$K(k)$	$K'(k)$	$E(k)$	$E'(k)$
0	$\pi/2$	∞	$\pi/2$	1
0.05	1.591	2.908	1.551	1.060
0.10	1.612	2.578	1.531	1.105
0.15	1.635	2.389	1.510	1.143
0.20	1.660	2.257	1.489	1.178
0.25	1.686	2.157	1.467	1.211
0.30	1.714	2.075	1.445	1.242
0.35	1.744	2.008	1.423	1.271
0.40	1.778	1.950	1.399	1.298
0.45	1.814	1.899	1.375	1.325
0.50	1.854	1.854	1.351	1.351
0.60	1.950	1.778	1.298	1.399
0.70	2.075	1.714	1.242	1.445
0.80	2.257	1.660	1.178	1.489
0.90	2.578	1.612	1.105	1.531
1	∞	$\pi/2$	1	$\pi/2$

6.18 JACOBIAN ELLIPTIC FUNCTIONS

The *Jacobian Elliptic functions* are the inverses of elliptic integrals. If $u = F(\phi, k)$ (the elliptic integral of the first kind)

$$u = \int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} = \int_0^{\sin\phi} \frac{dt}{\sqrt{(1-t^2)(1-k^2t^2)}}, \tag{6.18.1}$$

with $k^2 < 1$, then the inverse function is

$$\phi = \operatorname{am} u \qquad \text{(the } amplitude \text{ of } u). \tag{6.18.2}$$

(Note that the parameter k is not always explicitly written.) The Jacobian elliptic functions are then defined as

1. $\operatorname{sn} u = \operatorname{sn}(u,k) = \sin\phi = \sin(\operatorname{am} u)$
2. $\operatorname{cn} u = \operatorname{cn}(u,k) = \cos\phi = \cos(\operatorname{am} u) = \sqrt{1 - \operatorname{sn}^2 u}$
3. $\operatorname{dn} u = \operatorname{dn}(u,k) = \sqrt{1 - k^2 \sin^2 \phi} = \sqrt{1 - k^2 \operatorname{sn}^2 u}$

Note that

$$u = \int_1^{\operatorname{cn}(u,k)} \frac{dt}{\sqrt{(1-t^2)(k'^2 + k^2t^2)}} = \int_1^{\operatorname{dn}(u,k)} \frac{dt}{\sqrt{(1-t^2)(t^2 - k'^2)}}. \tag{6.18.3}$$

6.18.1 PROPERTIES

1. Relationships

$$\operatorname{sn}^2(u,k) + \operatorname{cn}^2(u,k) = 1,$$
$$\operatorname{dn}^2(u,k) + k^2 \operatorname{sn}^2(u,k) = 1,$$
$$\operatorname{dn}^2(u,k) - k^2 \operatorname{cn}^2(u,k) = 1 - k^2 = k'^2.$$

2. Special values

   (a) $\operatorname{sn}(0,k) = 0$,
   (b) $\operatorname{cn}(0,k) = 1$,
   (c) $\operatorname{dn}(0,k) = 1$,
   (d) $\operatorname{am}(0,k) = 0$,
   (e) $\operatorname{sn}(u,0) = \sin u$,
   (f) $\operatorname{cn}(u,0) = \cos u$,
   (g) $\operatorname{dn}(u,0) = 1$,
   (h) $\operatorname{sn}(u,1) = \tanh u$,
   (i) $\operatorname{cn}(u,1) = \operatorname{sech} u$,
   (j) $\operatorname{dn}(u,1) = \operatorname{sech} u$.

3. Symmetry properties

   (a) $\operatorname{sn}(-u) = -\operatorname{sn}(u)$,
   (b) $\operatorname{cn}(-u) = \operatorname{cn}(u)$,
   (c) $\operatorname{dn}(-u) = \operatorname{dn}(u)$,
   (d) $\operatorname{am}(-u) = -\operatorname{am}(u)$.

4. Addition formulas
   (a) $\operatorname{sn}(u \pm v) = \dfrac{\operatorname{sn} u \operatorname{cn} v \operatorname{dn} v \pm \operatorname{cn} u \operatorname{sn} v \operatorname{dn} u}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}$,
   (b) $\operatorname{cn}(u \pm v) = \dfrac{\operatorname{cn} u \operatorname{cn} v \mp \operatorname{sn} u \operatorname{dn} u \operatorname{sn} v \operatorname{dn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}$,
   (c) $\operatorname{dn}(u \pm v) = \dfrac{\operatorname{dn} u \operatorname{dn} v \mp k^2 \operatorname{sn} u \operatorname{cn} u \operatorname{sn} v \operatorname{cn} v}{1 - k^2 \operatorname{sn}^2 u \operatorname{sn}^2 v}$.
5. The elliptic functions are *doubly periodic functions* with respect to the variable u. The periods of

$$\begin{array}{llll} \operatorname{sn}(u,k) & \text{are} & 4K & \text{and} & 2iK', \\ \operatorname{cn}(u,k) & \text{are} & 4K & \text{and} & 2K + 2iK', \\ \operatorname{dn}(u,k) & \text{are} & 2K & \text{and} & 4iK'. \end{array}$$

6.18.2 DERIVATIVES AND INTEGRALS

1. $\dfrac{d}{du} \operatorname{sn} u = \operatorname{cn} u \operatorname{dn} u.$
2. $\dfrac{d}{du} \operatorname{cn} u = -\operatorname{sn} u \operatorname{dn} u.$
3. $\dfrac{d}{du} \operatorname{dn} u = -k^2 \operatorname{cn} u \operatorname{sn} u.$
4. $\displaystyle\int \operatorname{sn} u \, du = \frac{1}{k} (\operatorname{dn} u - k \operatorname{cn} u).$
5. $\displaystyle\int \operatorname{cn} u \, du = \frac{1}{k} \cos^{-1}(\operatorname{dn} u).$
6. $\displaystyle\int \operatorname{dn} u \, du = \operatorname{am} u = \sin^{-1}(\operatorname{sn} u).$

6.18.3 SERIES EXPANSIONS

$$\begin{aligned} \operatorname{sn}(u,k) &= u - (1+k^2)\frac{u^3}{3!} + (1 + 14k^2 + k^4)\frac{u^5}{5!} \\ &\qquad - (1 + 135k^2 + 135k^4 + k^6)\frac{u^7}{7!} + \dots, \\ \operatorname{cn}(u,k) &= 1 - \frac{u^2}{2!} + (1+4k^2)\frac{u^4}{4!} - (1 + 44k^2 + 16k^4)\frac{u^6}{6!} + \dots, \\ \operatorname{dn}(u,k) &= 1 - k^2\frac{u^2}{2!} + k^2(4+k^2)\frac{u^4}{4!} - k^2(16 + 44k^2 + k^4)\frac{u^6}{6!} + \dots . \end{aligned} \tag{6.18.4}$$

Let the *nome* q be defined by $q = e^{-\pi K/K'}$ and $v = \pi u/(2K)$. Then

$$\begin{aligned} \operatorname{sn}(u,k) &= \frac{2\pi}{kK} \sum_{n=0}^{\infty} \frac{q^{n+\frac{1}{2}}}{1 - q^{2n+1}} \sin[(2n+1)v], \\ \operatorname{cn}(u,k) &= \frac{2\pi}{kK} \sum_{n=0}^{\infty} \frac{q^{n+\frac{1}{2}}}{1 + q^{2n+1}} \cos[(2n+1)v], \\ \operatorname{dn}(u,k) &= \frac{\pi}{2K} + \frac{2\pi}{K} \sum_{n=1}^{\infty} \frac{q^n}{1 + q^{2n}} \cos(2nv). \end{aligned} \tag{6.18.5}$$

6.19 ERROR FUNCTIONS

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,dt,$$
$$\operatorname{erfc} x = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2}\,dt.$$

The function $\operatorname{erf} x$ is known as the *error function*. The function $\operatorname{erfc} x$ is known as the *complementary error function*.

6.19.1 PROPERTIES

1. Relationships:

$$\operatorname{erf} x + \operatorname{erfc} x = 1, \qquad \operatorname{erf}(-x) = -\operatorname{erf} x, \qquad \operatorname{erfc}(-x) = 2 - \operatorname{erfc} x.$$

2. Relationship with normal probability function:

$$\frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{1}{2}t^2}\,dt = \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right).$$

6.19.2 ERROR FUNCTION OF SPECIAL VALUES

$$\operatorname{erf}(\pm\infty) = \pm 1, \quad \operatorname{erfc}(-\infty) = 2, \quad \operatorname{erfc}\infty = 0,$$
$$\operatorname{erf} x_0 = \operatorname{erfc}\ x_0 = \frac{1}{2} \quad \text{if } x_0 \approx 0.476936.$$

6.19.3 EXPANSIONS

1. Series expansions:

$$\operatorname{erf} x = \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{(-1)^n\, x^{2n+1}}{(2n+1)\,n!} = \frac{2}{\sqrt{\pi}} \left(x - \frac{x^3}{3} + \frac{1}{2!}\frac{x^5}{5} - \frac{1}{3!}\frac{x^7}{7} + \dots \right)$$
$$= \frac{2}{\sqrt{\pi}} \sum_{n=0}^{\infty} \frac{\Gamma\left(\frac{3}{2}\right)\, e^{-x^2}}{\Gamma\left(n+\frac{3}{2}\right)} x^{2n+1} = \frac{2}{\sqrt{\pi}} e^{-x^2} \left(x + \frac{2}{3}x^3 + \frac{4}{15}x^5 \dots \right).$$

2. Asymptotic expansion: For $z \to \infty$, $|\arg z| < \frac{3}{4}\pi$,

$$\operatorname{erfc} z \sim \frac{2}{\sqrt{\pi}} \frac{e^{-z^2}}{2z} \sum_{n=0}^{\infty} \frac{(-1)^n\,(2n)!}{n!(2z)^{2n}}$$
$$\sim \frac{2}{\sqrt{\pi}} \frac{e^{-z^2}}{2z} \left(1 - \frac{1}{2\,z^2} + \frac{3}{4\,z^4} - \frac{15}{8\,z^6} + \dots \right).$$

6.19.4 SPECIAL CASES

1. Dawson's integral $\quad F(x) = e^{-x^2} \int_0^x e^{t^2}\, dt = -\frac{1}{2}\, i\sqrt{\pi}\, e^{-x^2}\, \operatorname{erf}(ix).$
2. Plasma dispersion function

$$\begin{aligned} w(z) = e^{-z^2}\operatorname{erfc}(-iz) &= \frac{1}{\pi i}\int_{-\infty}^{\infty} \frac{e^{-t^2}}{t-z}\, dt, \quad \operatorname{Im} z > 0 \\ = 2\, e^{-z^2} - w(-z) &= \sum_{n=0}^{\infty} \frac{(iz)^n}{\Gamma(\frac{n}{2}+1)}. \end{aligned}$$

6.20 FRESNEL INTEGRALS

$$C(z) = \sqrt{\frac{2}{\pi}} \int_0^z \cos t^2\, dt, \qquad S(z) = \sqrt{\frac{2}{\pi}} \int_0^z \sin t^2\, dt.$$

6.20.1 PROPERTIES

1. Relations: $\quad C(z) + iS(z) = \frac{1+i}{2}\, \operatorname{erf} \frac{(1-i)z}{\sqrt{2}}.$
2. Limits: $\quad \lim_{z\to\infty} C(z) = \frac{1}{2}, \quad \lim_{z\to\infty} S(z) = \frac{1}{2}.$
3. Representations:

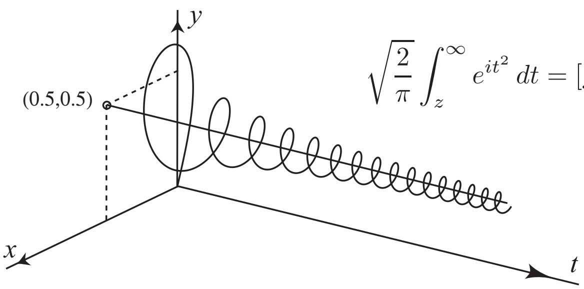
$$\begin{aligned} C(z) &= \frac{1}{2} + f(z)\, \sin(z^2) - g(z)\, \cos(z^2), \\ S(z) &= \frac{1}{2} - f(z)\, \cos(z^2) - g(z)\, \sin(z^2), \end{aligned}$$

where

$$f(z) = \frac{1}{\pi\sqrt{2}} \int_0^{\infty} \frac{e^{-z^2 t}}{\sqrt{t}(t^2+1)}\, dt, \qquad g(z) = \frac{1}{\pi\sqrt{2}} \int_0^{\infty} \frac{\sqrt{t}e^{-z^2 t}}{(t^2+1)}\, dt.$$

4. The figure below showing *Cornu's spiral* (also called an *Euler spiral*) is given by $x = C(t)$ and $y = S(t)$ for $t \geq 0$. (Figure from N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)

$$\sqrt{\frac{2}{\pi}} \int_z^{\infty} e^{it^2}\, dt = [g(z) + if(z)]\, e^{iz^2}.$$



6.20.2 ASYMPTOTIC EXPANSION

And for $z \to \infty,\ |\arg z| < \frac{1}{2}\pi$,

$$f(z) \sim \frac{1}{\pi\sqrt{2}} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(2n+1/2)}{z^{2n+1/2}} = \frac{1}{\sqrt{2\pi z}} \left[1 - \frac{3}{4\,z^2} + \frac{105}{16\,z^4} - \dots\right],$$

$$g(z) \sim \frac{1}{\pi\sqrt{2}} \sum_{n=0}^{\infty} (-1)^n \frac{\Gamma(2n+3/2)}{z^{2n+3/2}} = \frac{1}{2z\sqrt{2\pi z}} \left[1 - \frac{15}{4\,z^2} + \frac{945}{16\,z^4} - \dots\right].$$

6.20.3 NUMERICAL VALUES OF ERROR FUNCTIONS AND FRESNEL INTEGRALS

x	$\operatorname{erf}(x)$	$e^{x^2}\operatorname{erfc}(x)$	$C(x)$	$S(x)$
0.0	0.00000000	1.00000000	0.00000000	0.00000000
0.2	0.22270259	0.80901952	0.15955138	0.00212745
0.4	0.42839236	0.67078779	0.31833776	0.01699044
0.6	0.60385609	0.56780472	0.47256350	0.05691807
0.8	0.74210096	0.48910059	0.61265370	0.13223984
1.0	0.84270079	0.42758358	0.72170592	0.24755829
1.2	0.91031398	0.37853742	0.77709532	0.39584313
1.4	0.95228512	0.33874354	0.75781398	0.55244498
1.6	0.97634838	0.30595299	0.65866707	0.67442706
1.8	0.98909050	0.27856010	0.50694827	0.71289443
2.0	0.99532227	0.25539568	0.36819298	0.64211874
2.2	0.99813715	0.23559296	0.32253723	0.49407286
2.4	0.99931149	0.21849873	0.40704642	0.36532279
2.6	0.99976397	0.20361325	0.55998756	0.36073841
2.8	0.99992499	0.19054888	0.64079292	0.48940140
3.0	0.99997791	0.17900115	0.56080398	0.61721360
3.2	0.99999397	0.16872810	0.41390216	0.58920847
3.4	0.99999848	0.15953536	0.39874249	0.44174492
3.6	0.99999964	0.15126530	0.53845493	0.39648758
3.8	0.99999992	0.14378884	0.60092662	0.52778933
4.0	0.99999998	0.13699946	0.47431072	0.59612656
4.2	1.00000000	0.13080849	0.41041217	0.46899697
4.4	1.00000000	0.12514166	0.54218734	0.41991084
4.6	1.00000000	0.11993626	0.56533023	0.55685845
4.8	1.00000000	0.11513908	0.42894668	0.54293254
5.0	1.00000000	0.11070464	0.48787989	0.42121705

6.21 GAMMA FUNCTION

$$\Gamma(z) = \int_0^\infty t^{z-1}\, e^{-t}\, dt, \quad z = x + iy, \quad x > 0.$$

6.21.1 RECURSION FORMULA

$$\Gamma(z+1) = z\,\Gamma(z).$$

The relation $\Gamma(z) = \Gamma(z+1)/z$ can be used to define the gamma function in the left half plane with z not equal to a non-positive integer (i.e., $z \neq 0, -1, -2, \ldots$).

6.21.2 GAMMA FUNCTION OF SPECIAL VALUES

$$\Gamma(n+1) = n! \quad \text{if } n = 0, 1, 2, \ldots, \text{ where } 0! = 1,$$

$$\Gamma(1) = 1, \quad \Gamma(2) = 1, \quad \Gamma(3) = 2, \quad \Gamma(\tfrac{1}{2}) = \sqrt{\pi},$$

$$\Gamma\left(m + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2m-1)}{2^m}\sqrt{\pi}, \qquad m = 1, 2, 3, \ldots,$$

$$\Gamma\left(-m + \frac{1}{2}\right) = \frac{(-1)^m 2^m}{1 \cdot 3 \cdot 5 \cdots (2m-1)}\sqrt{\pi}, \qquad m = 1, 2, 3, \ldots.$$

$\Gamma\left(\frac{1}{4}\right) = 3.62560\,99082,$ $\quad\Gamma\left(\frac{1}{3}\right) = 2.67893\,85347,$

$\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} = 1.77245\,38509,$ $\quad\Gamma\left(\frac{2}{3}\right) = 1.35411\,79394,$

$\Gamma\left(\frac{3}{4}\right) = 1.22541\,67024,$ $\quad\Gamma\left(\frac{3}{2}\right) = \sqrt{\pi}/2 = 0.88622\,69254.$

6.21.3 PROPERTIES

1. Singular points: The gamma function has simple poles at $z = -n$, (for $n = 0, 1, 2, \ldots$), with the respective residues $(-1)^n/n!$; that is,

$$\lim_{z \to -n} (z+n)\Gamma(z) = \frac{(-1)^n}{n!}.$$

2. Definition by products:

$$\Gamma(z) = \lim_{n \to \infty} \frac{n!\, n^z}{z(z+1)\cdots(z+n)},$$

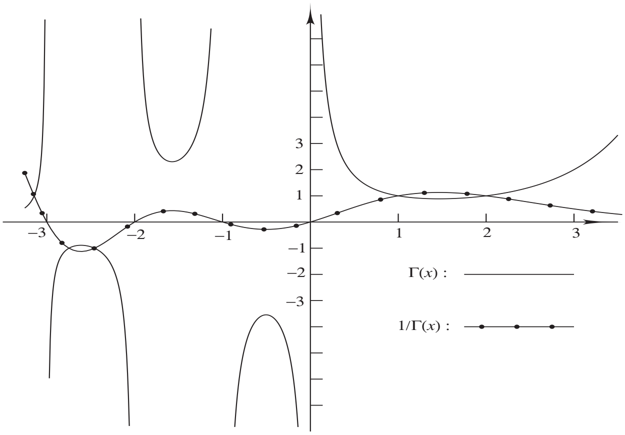
$$\frac{1}{\Gamma(z)} = z\, e^{\gamma z} \prod_{n=1}^{\infty} \left[(1 + z/n)\, e^{-z/n}\right], \quad \gamma \text{ is Euler's constant.}$$

3. Other integrals:

$$\Gamma(z)\, \cos\frac{\pi z}{2} = \int_0^\infty t^{z-1}\, \cos t\, dt, \qquad 0 < \text{Re } z < 1,$$

$$\Gamma(z)\, \sin\frac{\pi z}{2} = \int_0^\infty t^{z-1}\, \sin t\, dt, \quad -1 < \text{Re } z < 1.$$

FIGURE 6.7
Graphs of $\Gamma(x)$ *and* $1/\Gamma(x)$ *for* x *real.* (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)



4. Derivative at $x = 1$: $\Gamma'(1) = \int_0^{\infty} \ln t\, e^{-t}\, dt = -\gamma$

5. Multiplication formula: $\Gamma(2z) = \pi^{-1/2}\, 2^{2z-1}\, \Gamma(z)\, \Gamma\left(z + \frac{1}{2}\right)$

6. Reflection formulas:
$$\Gamma(z)\,\Gamma(1-z) = \frac{\pi}{\sin\,\pi z}, \qquad \Gamma\left(\tfrac{1}{2}+z\right)\,\Gamma\left(\tfrac{1}{2}-z\right) = \frac{\pi}{\cos\,\pi z},$$
$$\Gamma(z-n) = (-1)^n \Gamma(z) \frac{\Gamma(1-z)}{\Gamma(n+1-z)} = \frac{(-1)^n\,\pi}{\sin \pi z\, \Gamma(n+1-z)}.$$

6.21.4 ASYMPTOTIC EXPANSION

For $z \to \infty$, $|\arg z| < \pi$:

$$\Gamma(z) \sim \sqrt{\frac{2\pi}{z}}\, z^z\, e^{-z} \left[1 + \frac{1}{12\,z} + \frac{1}{288\,z^2} - \frac{139}{51\,840\,z^3} + \dots\right].$$

$$\begin{aligned} \ln\Gamma(z) &\sim \ln\left(\sqrt{\frac{2\pi}{z}} z^z e^{-z}\right) + \sum_{n=1}^{\infty} \frac{B_{2n}}{2n\,(2n-1)} \frac{1}{z^{2n-1}} \\ &\sim \ln\left(\sqrt{\frac{2\pi}{z}} z^z e^{-z}\right) + \frac{1}{12z} - \frac{1}{360z^3} + \frac{1}{1\,260z^5} - \frac{1}{1\,680z^7} + \dots, \end{aligned} \tag{6.21.1}$$

where B_n are the Bernoulli numbers. If we let $z = n$ a large positive integer, then a useful approximation for $n!$ is given by *Stirling's formula*,

$$\Gamma(n+1) = n! \sim \sqrt{2\pi n}\, n^n\, e^{-n}, \quad n \to \infty. \tag{6.21.2}$$

6.21.5 LOGARITHMIC DERIVATIVE OF THE GAMMA FUNCTION

1. Definition (for $z \neq 0, -1, -2, \ldots$):

$$\psi(z) = \frac{d}{dz} \ln \Gamma(z) = -\gamma + \sum_{n=0}^{\infty} \left(\frac{1}{n+1} - \frac{1}{z+n} \right).$$

2. Special values: $\psi(1) = -\gamma, \quad \psi\left(\frac{1}{2}\right) = -\gamma - 2\ln 2$
3. Asymptotic expansion: For $z \to \infty$, $|\arg z| < \pi$:

$$\begin{aligned}\psi(z) &\sim \ln z - \frac{1}{2z} - \sum_{n=1}^{\infty} \frac{B_{2n}}{2nz^{2n}} \\ &\sim \ln z - \frac{1}{2\,z} - \frac{1}{12\,z^2} + \frac{1}{120\,z^4} - \frac{1}{252\,z^6} + \ldots.\end{aligned}$$

6.21.6 NUMERICAL VALUES OF GAMMA FUNCTION

x	$\Gamma(x)$	$\ln\Gamma(x)$	$\psi(x)$	$\psi'(x)$
1.00	1.00000000	0.00000000	−0.57721566	1.64493407
1.04	0.97843820	−0.02179765	−0.51327488	1.55371164
1.08	0.95972531	−0.04110817	−0.45279934	1.47145216
1.12	0.94359019	−0.05806333	−0.39545533	1.39695222
1.16	0.92980307	−0.07278247	−0.34095315	1.32920818
1.20	0.91816874	−0.08537409	−0.28903990	1.26737721
1.24	0.90852106	−0.09593721	−0.23949368	1.21074707
1.28	0.90071848	−0.10456253	−0.19211890	1.15871230
1.32	0.89464046	−0.11133336	−0.14674236	1.11075532
1.36	0.89018453	−0.11632650	−0.10321006	1.06643142
1.40	0.88726382	−0.11961291	−0.06138454	1.02535659
1.44	0.88580506	−0.12125837	−0.02114267	0.98719773
1.48	0.88574696	−0.12132396	0.01762627	0.95166466
1.52	0.88703878	−0.11986657	0.05502211	0.91850353
1.56	0.88963920	−0.11693929	0.09113519	0.88749142
1.60	0.89351535	−0.11259177	0.12604745	0.85843189
1.80	0.93138377	−0.07108387	0.28499143	0.73697414
2.00	1.00000000	0.00000000	0.42278434	0.64493407

6.22 HYPERGEOMETRIC FUNCTIONS

Recall the geometric series and binomial expansion ($|z| < 1$),

$$(1-z)^{-1} = \sum_{n=0}^{\infty} z^n, \qquad (1-z)^{-a} = \sum_{n=0}^{\infty} \binom{-a}{n} (-z)^n = \sum_{n=0}^{\infty} \frac{(a)_n}{n!} z^n$$

where the shifted factorial, $(a)_n$, is defined in Section 1.3.4.

The *Gauss hypergeometric function*, F, is defined by:

$$\begin{aligned} F(a,b;c;z) &= \sum_{n=0}^{\infty} \frac{(a)_n\,(b)_n}{(c)_n\; n!}\, z^n \\ &= 1 + \frac{ab}{c} z + \frac{a(a+1)\,b(b+1)}{c(c+1)\,2!} z^2 + \dots, \quad |z| < 1, \\ &= F(b,a;c;z) \end{aligned} \tag{6.22.1}$$

where a, b and c may all assume complex values, $c \neq 0, -1, -2, \ldots$.

6.22.1 SPECIAL CASES

1. $F(a,b;b;z) = (1-z)^{-a}$
2. $F(1,1;2;z) = -\dfrac{\ln(1-z)}{z}$
3. $F\left(\dfrac{1}{2}, 1; \dfrac{3}{2}; z^2\right) = \dfrac{1}{2z} \ln\left(\dfrac{1+z}{1-z}\right)$
4. $F\left(\dfrac{1}{2}, 1; \dfrac{3}{2}; -z^2\right) = \dfrac{\tan^{-1} z}{z}$
5. $F\left(\dfrac{1}{2}, \dfrac{1}{2}; \dfrac{3}{2}; z^2\right) = \dfrac{\sin^{-1} z}{z}$
6. $F\left(\dfrac{1}{2}, \dfrac{1}{2}; \dfrac{3}{2}; -z^2\right) = \dfrac{\ln(z + \sqrt{1+z^2})}{z}$
7. Polynomial case; for $m = 0, 1, 2, \ldots$

$$F(-m,b;c;z) = \sum_{n=0}^{m} \frac{(-m)_n\,(b)_n}{(c)_n\; n!}\, z^n = \sum_{n=0}^{m} (-1)^n \binom{m}{n} \frac{(b)_n}{(c)_n}\, z^n. \tag{6.22.2}$$

6.22.2 PROPERTIES

1. Derivatives:

$$\begin{aligned} \frac{d}{dz} F(a,b;c;z) &= \frac{ab}{c} F(a+1, b+1; c+1; z), \\ \frac{d^n}{dz^n} F(a,b;c;z) &= \frac{(a)_n\,(b)_n}{(c)_n} F(a+n, b+n; c+n; z). \end{aligned}$$

2. Special values; when Re $(c-a-b)>0$:

$$F(a,b;c;1)=\frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)}.$$

3. Integral; when Re $c>$ Re $b>0$:

$$F(a,b;c;z)=\frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 t^{b-1}(1-t)^{c-b-1}(1-tz)^{-a}\,dt.$$

4. Functional relationships:

$$\begin{aligned}F(a,b;c;z)&=(1-z)^{-a}F\bigl(a,c-b;c;\frac{z}{z-1}\bigr)\\&=(1-z)^{-b}F\bigl(c-a,b;c;\frac{z}{z-1}\bigr)\\&=(1-z)^{c-a-b}F(c-a,c-b;c;z).\end{aligned}$$

5. Differential equation:

$$z(1-z)F''+[(c-(a+b+1)z]F'-abF=0,$$

with (regular) singular points $z=0,1,\infty$. (See page 408 for singular points.)

6.22.3 RECURSION FORMULAS

Notation: F is $F\,(a,b;c;z)$; $F(a+), F(a-)$ are $F\,(a+1,b;c;z)$, $F\,(a-1,b;c;z)$, respectively, etc.

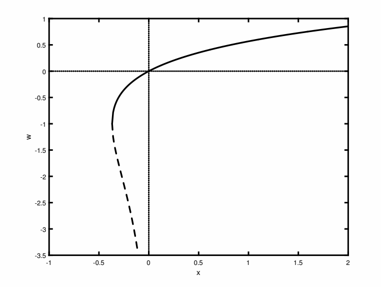
1. $(c-a)F(a-)+(2a-c-az+bz)F+a(z-1)F(a+)=0$
2. $c(c-1)(z-1)F(c-)+c[c-1-(2c-a-b-1)z]F+(c-a)(c-b)zF(c+)=0$
3. $c[a+(b-c)z]F-ac(1-z)F(a+)+(c-a)(c-b)zF(c+)=0$
4. $c(1-z)F-cF(a-)+(c-b)zF(c+)=0$
5. $(b-a)F+aF(a+)-bF(b+)=0$
6. $(c-a-b)F+a(1-z)F(a+)-(c-b)F(b-)=0$
7. $(c-a-1)F+aF(a+)-(c-1)F(c-)=0$
8. $(b-a)(1-z)F-(c-a)F(a-)+(c-b)F(b-)=0$
9. $[a-1+(b+1-c)z]F+(c-a)F(a-)-(c-1)(1-z)F(c-)=0$

6.23 LAMBERT FUNCTION

Lambert's $W(x)$ function is defined implicitly by

$$We^W = x$$

When $x \in [0, \infty]$ there is one real solution $W(x)$; it is nonnegative and increasing. When $x \in \left(-\frac{1}{e}, 0\right)$ there are two real solutions, one increasing and one decreasing.



6.23.1 PROPERTIES

1. Numerical Values

   (a) $W\left(-e^{-1}\right) = -1$

   (b) $W(0) = 0 \quad W'(0) = 1$

   (c) $W(1) = 0.567143\ldots$

   (d) $W(e) = 1$

2. Representations

   (a) $W(x) = \displaystyle\sum_{n=1}^{\infty} \frac{(-n)^{n-1}}{n!} x^n = x - x^2 + \frac{3}{2}x^3 - \frac{8}{3}x^4 + \ldots$

   (b) As $x \to \infty$ we have $W(x) \sim \ln x - \ln\ln x + \frac{\ln\ln x}{\ln x} + \ldots$

   (c) As $x \to 0^-$ we have $W(x) \sim -\ln\left(-\frac{1}{x}\right) - \ln\ln\left(-\frac{1}{x}\right) - \frac{\ln\ln\left(-\frac{1}{x}\right)}{\ln\left(-\frac{1}{x}\right)} + \ldots$

3. Solving equations

   (a) $\Big(\ln(A + Bx) + Cx = \ln D\Big) \iff \left(x = \frac{1}{C} W\left(\frac{CD}{B} e^{AC/B}\right) - \frac{A}{B}\right)$

   (b) $\left(p^{ax+b} = cx + d\right) \iff \left(x = -\frac{W\left(-\frac{a \ln p}{c} p^{b-ad/c}\right)}{a \ln p} - \frac{d}{c}\right)$

   (c) $\left(x^x = z\right) \iff \left(x = \frac{\ln z}{W(\ln z)}\right)$

   (d) $x \uparrow\uparrow \infty = x^{x^{x^{\cdot^{\cdot}}}} = \frac{W(-\ln x)}{-\ln x}$ when the power tower converges

4. Differential equation and integrals

   (a) $\dfrac{dW}{dx} = \dfrac{1}{x + e^W}$

   (b) $\displaystyle\int W(x)\, dx = x(W(x) - 1) + e^{W(x)} + C$

   (c) $\displaystyle\int_0^e W(x)\, dx = e - 1$

   (d) $\displaystyle\int_0^\infty \frac{W(x)}{x^{3/2}}\, dx = 2\sqrt{2\pi}$

6.24 LEGENDRE FUNCTIONS

6.24.1 DIFFERENTIAL EQUATION

The *Legendre differential equation* is

$$(1 - z^2)w'' - 2zw' + \nu(\nu + 1)w = 0. \tag{6.24.1}$$

The solutions $P_\nu(z), Q_\nu(z)$ can be written in terms of Gaussian hypergeometric functions:

$$\begin{aligned}
P_\nu(z) &= F\left(-\nu, \nu + 1; 1; \frac{1}{2} - \frac{1}{2}\,z\right), \\
Q_\nu(z) &= \frac{\sqrt{\pi}\,\Gamma(\nu + 1)}{\Gamma\left(\nu + \frac{3}{2}\right)\,(2z)^{\nu+1}} F\left(\frac{1}{2}\,\nu + 1, \frac{1}{2}\,\nu + \frac{1}{2}; \nu + \frac{3}{2}; z^{-2}\right).
\end{aligned}$$

1. $P_\nu(z)$ has a singular point at $z = -1$ and is analytic in the remaining part of the complex plane, with a branch cut along $(-\infty, -1]$.
2. $Q_\nu(z)$ has singular points at $z = \pm 1$ and is analytic in the remaining part of the complex $z-$plane, with a branch cut along $(-\infty, +1]$. The Q_ν function is not defined if $\nu = -1, -2, \ldots$.

6.24.2 RELATIONSHIPS

$$\begin{aligned}
P_{-\nu-1}(z) &= P_\nu(z), \\
Q_{-\nu-1}(z) &= Q_\nu(z) - \pi \cot \nu\pi P_\nu(z).
\end{aligned}$$

Let $n \geq m$, then

$$\int_{-1}^{1} P_n^m(x) P_k^m(x)\, dx = \begin{cases} 0, & \text{if } k \neq n, \\ \dfrac{2}{2n + 1}\,\dfrac{(n + m)!}{(n - m)!}, & \text{if } k = n. \end{cases}$$

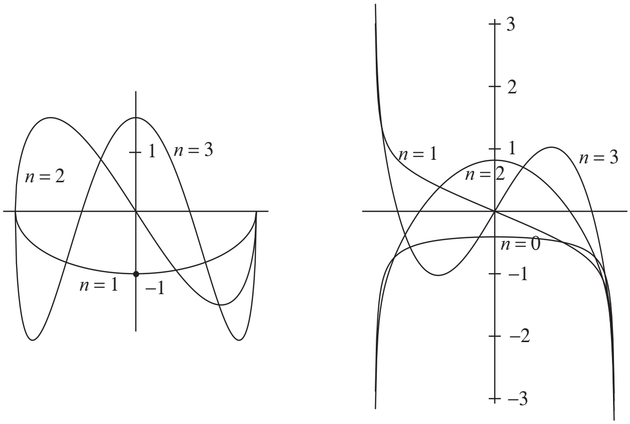
6.24.3 RECURSION RELATIONSHIPS

$$\begin{aligned}
(\nu + 1)P_{\nu+1}(z) &= (2\nu + 1)zP_\nu(z) - \nu P_{\nu-1}(z), \\
(2\nu + 1)P_\nu(z) &= P'_{\nu+1}(z) - P'_{\nu-1}(z), \\
(\nu + 1)P_\nu(z) &= P'_{\nu+1}(z) - zP'_\nu(z), \\
\nu P_\nu(z) &= zP'_\nu(z) - P'_{\nu-1}(z), \\
(1 - z^2)P'_\nu(z) &= \nu P_{\nu-1}(z) - \nu z P_\nu(z).
\end{aligned}$$

The functions $Q_\nu(z)$ satisfy the same relations.

FIGURE 6.8

Legendre functions $P_n(x)$, $n = 1, 2, 3$ *(left) and* $Q_n(x)$, $n = 0, 1, 2, 3$ *(right) on the interval* $[-1, 1]$. (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)



6.24.4 INTEGRAL REPRESENTATIONS

$$\begin{aligned} P_\nu(\cosh\alpha) &= \frac{2}{\pi}\int_0^\alpha \frac{\cosh(\nu+\frac{1}{2})\theta}{\sqrt{2\cosh\alpha - 2\cosh\theta}}\,d\theta \\ &= \frac{1}{\pi}\int_{-\alpha}^\alpha \frac{e^{-(\nu+1/2)\theta}}{\sqrt{2\cosh\alpha - 2\cosh\theta}}\,d\theta \\ &= \frac{1}{\pi}\int_0^\pi \frac{d\psi}{(\cosh\alpha + \sinh\alpha\ \cos\psi)^{\nu+1}} \\ &= \frac{1}{\pi}\int_0^\pi (\cosh\alpha + \sinh\alpha\ \cos\psi)^\nu\, d\psi. \end{aligned} \tag{6.24.2}$$

$$\begin{aligned} P_\nu(\cos\beta) &= \frac{2}{\pi}\int_0^\beta \frac{\cos(\nu+\frac{1}{2})\theta}{\sqrt{2\cos\theta - 2\cos\beta}}\,d\theta \\ &= \frac{1}{\pi}\int_0^\pi \frac{d\psi}{(\cos\beta + i\sin\beta\ \cos\psi)^{\nu+1}} \\ &= \frac{1}{\pi}\int_0^\pi (\cos\beta + i\sin\beta\ \cos\psi)^\nu\, d\psi. \end{aligned} \tag{6.24.3}$$

$$
\begin{aligned}
Q_\nu(z) &= 2^{-\nu-1}\int_{-1}^{1}\frac{(1-t^2)^\nu}{(z-t)^{\nu+1}}\,dt \qquad \operatorname{Re}\nu > -1,\ |\arg z| < \pi,\ z \notin [-1,1] \\
&= \int_0^\infty \left[z+\sqrt{z^2-1}\cosh\phi\right]^{-\nu-1} d\phi \\
&= \int_\alpha^\infty \frac{e^{-(\nu+1/2)\theta}}{\sqrt{2\cosh\theta - 2\cosh\alpha}}\,d\theta, \quad z=\cosh\alpha.
\end{aligned} \tag{6.24.4}
$$

6.24.5 POLYNOMIAL CASE

Legendre polynomials occur when $\nu = n = 0, 1, 2, \ldots$ (see page 453)

$$
\begin{aligned}
P_n(x) &= F\left(-n, n+1; 1; \frac{1}{2} - \frac{x}{2}\right) \\
&= \sum_{k=0}^{m} \frac{(-1)^k (2n-2k)!}{2^n\, k!\, (n-k)!\, (n-2k)!} x^{n-2k}, \ m = \begin{cases} \frac{1}{2}n, & \text{if } n \text{ even}, \\ \frac{1}{2}(n-1), & \text{if } n \text{ odd}. \end{cases}
\end{aligned} \tag{6.24.5}
$$

The Legendre polynomials satisfy $\int_{-1}^{1} P_n(x)P_m(x)\,dx = \frac{2}{2m+1}\delta_{nm}$.
The Legendre series representation is

$$
f(x) = \sum_{n=0}^{\infty} A_n P_n(x), \qquad A_n = \frac{2n+1}{2}\int_{-1}^{1} f(x)P_n(x)\,dx. \tag{6.24.6}
$$

For integer order, we distinguish two cases: $Q_n(x)$ (defined for $x \in (-1,1)$) and $Q_n(z)$ (defined for Re $z \notin [-1,1]$):

$$
Q_0(x) = \frac{1}{2}\ln\frac{1+x}{1-x}, \qquad Q_1(x) = \frac{1}{2}x\ln\frac{1+x}{1-x} - 1, \tag{6.24.7}
$$

and

$$
Q_0(z) = \frac{1}{2}\ln\frac{z+1}{z-1}, \qquad Q_1(z) = \frac{1}{2}z\ln\frac{z+1}{z-1} - 1. \tag{6.24.8}
$$

In both cases

$$
Q_n(y) = P_n(y)Q_0(y) - \sum_{k=0}^{n-1}\frac{(2k+1)[1-(-1)^{n+k}]}{(n+k+1)(n-k)}P_k(y). \tag{6.24.9}
$$

Legendre polynomials $P_n(x)$ and functions $Q_n(x)$, $x \in (-1,1)$.

n	$P_n(x)$	$Q_n(x)$
0	1	$\frac{1}{2}\ln[(1+x)/(1-x)]$
1	x	$P_1(x)Q_0(x) - 1$
2	$\frac{1}{2}(3x^2-1)$	$P_2(x)Q_0(x) - \frac{3}{2}x$
3	$\frac{1}{2}x(5x^2-3)$	$P_3(x)Q_0(x) - \frac{5}{2}x^2 + \frac{2}{3}$
4	$\frac{1}{8}(35x^4 - 30x^2 + 3)$	$P_4(x)Q_0(x) - \frac{35}{8}x^3 + \frac{55}{24}x$
5	$\frac{1}{8}x(63x^4 - 70x^2 + 15)$	$P_5(x)Q_0(x) - \frac{63}{8}x^4 + \frac{49}{8}x^2 - \frac{8}{15}$

6.24.6 ASSOCIATED LEGENDRE FUNCTION

The *associated Legendre differential equation* is

$$(1-z^2)y'' - 2zy' + \left[\nu(\nu+1) - \frac{\mu^2}{1-z^2}\right] y = 0.$$

The solutions $P_\nu^\mu(z), Q_\nu^\mu(z)$, the *associated Legendre functions*, can be given in terms of Gauss hypergeometric functions. We only consider integer values of μ, ν, and replace them with m, n. Then the associated differential equation follows from the Legendre differential equation after it has been differentiated m times.

6.24.7 RELATIONSHIPS BETWEEN THE ASSOCIATED AND ORDINARY LEGENDRE FUNCTIONS

The following relationships are for $z \notin [-1, 1]$

$$\begin{aligned}
P_n^m(z) &= (z^2-1)^{\frac{1}{2}m} \frac{d^m}{dz^m} P_n(z), \\
&= \frac{(-1)^m}{2^n n!}(z^2-1)^{\frac{1}{2}m} \frac{d^{n+m}}{dx^{n+m}}(z^2-1)^n, \\
P_n^{-m}(z) &= \frac{(n-m)!}{(n+m)!} P_n^m(z), \\
Q_n^m(z) &= (z^2-1)^{\frac{1}{2}m} \frac{d^m}{dz^m} Q_n(z), \\
Q_n^{-m}(z) &= \frac{(n-m)!}{(n+m)!} Q_n^m(z), \\
P_n^{-m}(z) &= (z^2-1)^{-\frac{1}{2}m} \underbrace{\int_1^z \cdots \int_1^z}_{m} P_n(z)\,(dz)^m, \\
Q_n^{-m}(z) &= (-1)^m (z^2-1)^{-\frac{1}{2}m} \underbrace{\int_z^\infty \cdots \int_z^\infty}_{m} Q_n(z)\,(dz)^m, \\
P_{-n-1}^m(z) &= P_n^m(z).
\end{aligned}$$

6.24.8 RECURSION RELATIONSHIPS

$$\begin{aligned}
P_n^{m+1}(z) + \frac{2mz}{\sqrt{z^2-1}} P_n^m(z) &= (n-m+1)(n+m) P_n^{m-1}(z), \\
(z^2-1)\frac{dP_n^m(z)}{dz} &= mzP_n^m(z) + \sqrt{z^2-1} P_n^{m+1}(z), \\
(2n+1)zP_n^m(z) &= (n-m+1)P_{n+1}^m(z) + (n+m)P_{n-1}^m(z), \\
(z^2-1)\frac{dP_n^m(z)}{dz} &= (n-m+1)P_{n+1}^m(z) - (n+1)zP_n^m(z), \\
P_{n-1}^m(z) - P_{n+1}^m(z) &= -(2n+1)\sqrt{z^2-1} P_n^{m-1}(z).
\end{aligned}$$

The functions $Q_n^m(z)$ satisfy the same relations.

6.25 POLYLOGARITHMS

$$\mathrm{Li}_1(z) = \int_0^z \frac{dt}{1-t} = -\ln(1-z), \qquad \text{logarithm}$$

$$\mathrm{Li}_2(z) = \int_0^z \frac{\mathrm{Li}_1(t)}{t}\,dt = -\int_0^z \frac{\ln(1-t)}{t}\,dt, \qquad \text{dilogarithm}$$

$$\mathrm{Li}_n(z) = \int_0^z \frac{\mathrm{Li}_{n-1}(t)}{t}\,dt, \quad n \geq 2, \qquad \text{polylogarithm}$$

$$\mathrm{Li}_\nu(z) = \frac{z}{\Gamma(\nu)} \int_0^\infty \frac{t^{\nu-1}}{e^t - z}\,dt, \qquad \mathrm{Re}\,\nu > 0, \quad z \notin \{\mathrm{Re}\,z \in [1,\infty], \mathrm{Im}\,z = 0\}$$

6.25.1 POLYLOGARITHMS OF SPECIAL VALUES

$$\mathrm{Li}_2(1) = \frac{\pi^2}{6}, \quad \mathrm{Li}_2(-1) = -\frac{\pi^2}{12}, \quad \mathrm{Li}_2(\tfrac{1}{2}) = \frac{\pi^2}{12} - \frac{(\ln 2)^2}{2}$$

$$\mathrm{Li}_\nu(1) = \zeta(\nu), \quad \mathrm{Re}\,\nu > 1 \quad \text{(Riemann zeta function)}$$

6.25.2 POLYLOGARITHM PROPERTIES

1. Definition: For any complex ν

$$\mathrm{Li}_\nu(z) = \sum_{k=1}^\infty \frac{z^k}{k^\nu}, \quad |z| < 1.$$

2. Singular points: $z = 1$ is a singular point of $\mathrm{Li}_\nu(z)$.
3. Generating function:

$$\sum_{n=2}^\infty w^{n-1}\,\mathrm{Li}_n(z) = z \int_0^\infty \frac{e^{wt}-1}{e^t - z}\,dt, \quad z \notin [1,\infty).$$

The series converges for $|w| < 1$, the integral is defined for Re $w < 1$.

4. Functional equations for dilogarithms:

$$\mathrm{Li}_2(z) + \mathrm{Li}_2(1-z) = \frac{1}{6}\pi^2 - \ln z\,\ln(1-z),$$

$$\frac{1}{2}\,\mathrm{Li}_2(x^2) = \mathrm{Li}_2(x) + \mathrm{Li}_2(-x),$$

$$\mathrm{Li}_2(-1/x) + \mathrm{Li}_2(-x) = -\frac{1}{6}\pi^2 - \frac{1}{2}(\ln x)^2,$$

$$2\mathrm{Li}_2(x) + 2\mathrm{Li}_2(y) + 2\mathrm{Li}_2(z) = \mathrm{Li}_2(-xy/z) + \mathrm{Li}_2(-yz/x) + \mathrm{Li}_2(-zx/y),$$

where $1/x + 1/y + 1/z = 1$.

6.26 PROLATE SPHEROIDAL WAVE FUNCTIONS

The spheroidal wave functions $\mathbb{P}_n^m(x, \gamma^2)$ are the solutions of

$$\frac{d}{dx}\left(\left(1-x^2\right)\frac{dw}{dx}\right) + \left(\lambda + \gamma^2(1-x^2) - \frac{m^2}{1-x^2}\right) w = 0, \qquad (6.26.1)$$

that are bounded on $(-1, 1)$, with real parameters λ, γ^2, and $m = 0, 1, 2, \ldots$.. Solutions only exist for special values of λ, namely, the eigenvalues $\lambda_n^m(\gamma^2)$. When $\gamma = 0$, they reduce to the associated Legendre functions with $\lambda_n^m(0) = n(n+1)$.

Uaing $\gamma = \tau\sigma$ (for $\tau, \sigma > 0$) the prolate spheroidal wave functions $\varphi_{n,\sigma,\tau}(t)$ are

$$\varphi_{n,\sigma,\tau}(t) = \sqrt{\frac{(2n+1)\lambda_n}{2\tau}}\, \mathbb{P}_n^0\left(\frac{t}{\tau}, \gamma^2\right), \quad \text{for } n = 0, 1, 2, \ldots \qquad (6.26.2)$$

where $\lambda_n = \lambda_{n,\tau,\sigma} > 0$.

1. **Differential Equation** here $\nu_{n,\sigma,\tau}$ are different eigenvalues

$$(\tau^2 - t^2)\frac{d^2\varphi_{n,\sigma,\tau}}{dt^2} - 2t\frac{d\varphi_{n,\sigma,\tau}}{dt} - \sigma^2 t^2 \varphi_{n,\sigma,\tau} = \nu_{n,\sigma,\tau}\varphi_{n,\sigma,\tau}, \qquad (6.26.3)$$

2. **Integral Equations**

$$\begin{aligned} \int_{-\tau}^{\tau} \varphi_{n,\sigma,\tau}(x)\frac{\sin\sigma(t-x)}{\pi(t-x)}\,dx &= \lambda_n\varphi_{n,\sigma,\tau}(t) \\ \int_{-\infty}^{\infty} \varphi_{n,\sigma,\tau}(x)\frac{\sin\sigma(t-x)}{\pi(t-x)}\,dx &= \varphi_{n,\sigma,\tau}(t) \\ \int_{-\tau}^{\tau} \varphi_{n,\sigma,\tau}(t)e^{-i\sigma wt/\tau}\,dt &= \gamma_{n,\sigma,\tau}\varphi_{n,\sigma,\tau}(w) \end{aligned} \qquad (6.26.4)$$

3. **Orthogonality**

$$\begin{aligned} \int_{-\tau}^{\tau} \varphi_{n,\sigma,\tau}(t)\varphi_{m,\sigma,\tau}(t)\,dt &= \lambda_n\delta_{m,n} \\ \int_{-\infty}^{\infty} \varphi_{n,\sigma,\tau}(t)\varphi_{m,\sigma,\tau}(t)\,dt &= \delta_{m,n} \end{aligned} \qquad (6.26.5)$$

4. **Fourier Transform** here $\chi_\sigma(w) = 1$ on $(-\sigma, \sigma)$, and zero otherwise.

$$\int_{-\infty}^{\infty} e^{-itw}\varphi_{n,\sigma,\tau}(t)\,dt = (-i)^n\sqrt{\frac{2\pi\tau}{\sigma\lambda_n}}\,\varphi_{n,\sigma,\tau}\left(\frac{\tau w}{\sigma}\right)\chi_\sigma(w) \qquad (6.26.6)$$

5. **Discrete Orthogonality**

$$\sum_{n=0}^{\infty}\varphi_{n,\sigma,\tau}\left(\frac{k\pi}{\sigma}\right)\varphi_{n,\sigma,\tau}\left(\frac{m\pi}{\sigma}\right) = \delta_{k,m} \qquad (6.26.7)$$

6. **Special Summation**

$$\sum_{n=0}^{\infty}\varphi_{n,\sigma,\tau}(t)\varphi_{n,\sigma,\tau}(x) = \frac{\sin\sigma(t-x)}{\pi(t-x)} \qquad (6.26.8)$$

6.27 SINE, COSINE, AND EXPONENTIAL INTEGRALS

6.27.1 SINE AND COSINE INTEGRALS

$$\mathrm{Si}(z) = \int_0^z \frac{\sin t}{t}\,dt, \qquad \mathrm{Ci}(z) = \gamma + \ln z + \int_0^z \frac{\cos t - 1}{t}\,dt,$$

where γ is Euler's constant.

1. Alternative definitions:

$$\mathrm{Si}(z) = \frac{1}{2}\,\pi - \int_z^\infty \frac{\sin t}{t}\,dt, \quad \mathrm{Ci}(z) = -\int_z^\infty \frac{\cos t}{t}\,dt.$$

2. Limits: $\lim_{z\to\infty} \mathrm{Si}(z) = \frac{1}{2}\,\pi, \quad \lim_{z\to\infty} \mathrm{Ci}(z) = 0.$
3. Representations:

$$\mathrm{Si}(z) = -f(z)\cos z - g(z)\sin z + \frac{1}{2}\,\pi,$$
$$\mathrm{Ci}(z) = +f(z)\sin z - g(z)\cos z,$$

where

$$f(z) = \int_0^\infty \frac{e^{-zt}}{t^2+1}\,dt, \quad g(z) = \int_0^\infty \frac{te^{-zt}}{t^2+1}\,dt.$$

4. Asymptotic expansion: For $z \to \infty$, $|\arg z| < \pi$,

$$f(z) \sim \frac{1}{z}\sum_{n=0}^\infty (-1)^n \frac{(2n)!}{z^{2n}}, \qquad g(z) \sim \frac{1}{z^2}\sum_{n=0}^\infty (-1)^n \frac{(2n+1)!}{z^{2n}}.$$

6.27.2 EXPONENTIAL INTEGRALS

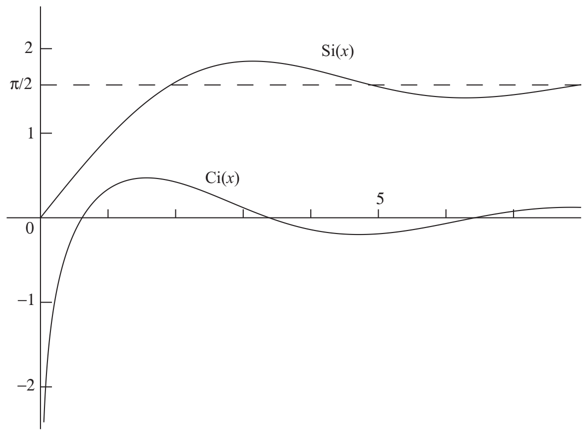
$$E_n(z) = \int_1^\infty \frac{e^{-zt}}{t^n}\,dt, \quad \mathrm{Re}\, z > 0, \quad n = 1, 2, \dots.$$
$$= \frac{z^{n-1}e^{-z}}{\Gamma(n)} \int_0^\infty \frac{e^{-zt}t^{n-1}}{t+1}\,dt, \quad \mathrm{Re}\, z > 0.$$

1. Special case: $E_1(z) = \int_z^\infty \frac{e^{-t}}{t}\,dt, \quad |\arg z| < \pi.$

   For real values $\mathrm{Ei}(x) = \int_{-\infty}^x \frac{e^t}{t}\,dt$, where for $x > 0$ the integral is interpreted as a Cauchy principal value integral.
2. Representations:

$$E_1(z) = -\gamma - \ln z + \int_0^z \frac{1 - e^{-t}}{t}\,dt,$$
$$E_1\left(ze^{\frac{1}{2}\pi i}\right) = -\gamma - \ln z - \mathrm{Ci}(z) + i\left[-\frac{1}{2}\,\pi + \mathrm{Si}(z)\right].$$
$$E_1(x) = -\,\mathrm{Ei}(-x), \qquad x > 0$$

FIGURE 6.9
Sine and cosine integrals $\mathrm{Si}(x)$ *and* $\mathrm{Ci}(x)$*, for* $0 \le x \le 8$*.* (From N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics,* John Wiley & Sons, 1996. Reprinted with permission of John Wiley & Sons, Inc.)



6.27.3 LOGARITHMIC INTEGRAL

$$\mathrm{li}(x) = \int_0^x \frac{dt}{\ln t} = \mathrm{Ei}(\ln x),$$

where for $x > 1$ the integral is interpreted as a Cauchy principal value integral.

6.27.4 NUMERICAL VALUES

x	$\mathrm{Si}(x)$	$\mathrm{Ci}(x)$	$e^x E_1(x)$	$e^{-x}\,\mathrm{Ei}(x)$
0.0	0.00000000	$-\infty$	∞	$-\infty$
0.2	0.19955609	−1.04220560	1.49334875	−0.67280066
0.4	0.39646146	−0.37880935	1.04782801	0.07022623
0.6	0.58812881	−0.02227071	0.82793344	0.42251981
0.8	0.77209579	0.19827862	0.69124540	0.60542430
1.0	0.94608307	0.33740392	0.59634736	0.69717488
2.0	1.60541298	0.42298083	0.36132862	0.67048271
3.0	1.84865253	0.11962979	0.26208374	0.49457640
4.0	1.75820314	−0.14098170	0.20634565	0.35955201
5.0	1.54993124	−0.19002975	0.17042218	0.27076626

6.28 WEIERSTRASS ELLIPTIC FUNCTION

The *Weierstrass elliptic function*, $y = \wp(z) = \wp(z; \omega_1, \omega_3)$, is defined as the inverse of the elliptic integral

$$z = \int_y^\infty \frac{ds}{\sqrt{4s^3 - g_2 s - g_3}}, \tag{6.28.1}$$

where g_2 and g_3 are constants. The result follows from the differential equation $[\wp'(z)]^2 = 4[\wp(z)]^3 - g_2\wp(z) - g_3$. The parameters $\{g_2, g_3\}$ are the *invariants* of $\wp$.

6.28.1 PROPERTIES

1. If $\{e_1, e_2, e_3\}$ are the roots of the cubic $4s^3 - g_2 s - g_3 = 0$, then $g_1 = 0 = 4(e_1 + e_2 + e_3)$, $g_2 = -4(e_1e_2 + e_1e_3 + e_2e_3) = 2\left(e_1^2 + e_2^3 + e_3^2\right)$, and $g_3 = 4e_1e_2e_3$. Define the *periods* of $\wp$ (denoted $\{\omega_i\}$) by

$$\wp\left(\frac{\omega_1}{2}\right) = e_1, \quad \wp\left(\frac{\omega_2}{2}\right) = e_2, \quad \wp\left(\frac{\omega_3}{2}\right) = e_3. \tag{6.28.2}$$

   Then we find $\omega_1 + \omega_2 + \omega_3 = 0$ and

$$\wp(z + \omega_i) = \wp(z), \quad i = 1, 3, \tag{6.28.3}$$

   which means that $\wp$ is *doubly periodic* with periods ω_1 and ω_3.
2. Addition theorem

$$\wp(u + v) = \frac{1}{4}\left[\frac{\wp'(u) - \wp'(v)}{\wp(u) - \wp(v)}\right]^2 - \wp(u) - \wp(v) \tag{6.28.4}$$

3. Changing parameters

$$\wp(z; -\omega_3, \omega_1) = \wp(z; \omega_1, \omega_3). \tag{6.28.5}$$

4. $\wp(z; \omega_1, \omega_3)$ is real valued when z is real, ω_1 is real and ω_3 is pure imaginary.
5. $g_2 = 60 \sum_{m^2+n^2\neq 0} \frac{1}{(m\omega_1 + n\omega_3)^4} \qquad g_3 = 140 \sum_{m^2+n^2\neq 0} \frac{1}{(m\omega_1 + n\omega_3)^6}$

6.28.2 RELATION TO JACOBI ELLIPTIC FUNCTIONS

$$\operatorname{sn}(z) = \frac{1}{\sqrt{\wp(z) - e_3}}, \qquad \operatorname{cn}(z) = \sqrt{\frac{\wp(z) - e_1}{\wp(z) - e_3}}, \qquad \operatorname{dn}(z) = \sqrt{\frac{\wp(z) - e_2}{\wp(z) - e_3}}. \tag{6.28.6}$$

6.28.3 SERIES EXPANSION

$$\begin{aligned}\wp(z; \omega_1, \omega_3) &= \frac{1}{z^2} + \sum_{m^2+n^2\neq 0}\left(\frac{1}{(z + m\omega_1 + n\omega_3)^2} - \frac{1}{(m\omega_1 + n\omega_3)^2}\right)\\ &\sim \frac{1}{z^2} + \frac{g_2}{20}z^2 + \frac{g_3}{28}z^4 + O\left(z^6\right)\end{aligned} \tag{6.28.7}$$

6.29 INTEGRAL TRANSFORMS: LIST

Transform	Definition	Inverse
Abel transform	$\widehat{f}(t) = \int_0^t \frac{f(x)}{(t-x)^\alpha}\,dx$	$f(x) = \frac{\sin(\alpha\pi)}{\pi}\frac{d}{dx}\int_0^x \frac{\widehat{f}(t)}{(x-t)^{1-\alpha}}\,dt$
Fourier transform	$\widehat{f}(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-iux} f(x)\,dx$	$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{iux}\widehat{f}(u)\,du$
Fourier cosine transform	$\widehat{f}(u) = \sqrt{\frac{2}{\pi}}\int_0^\infty \cos(ux) f(x)\,dx$	$f(x) = \sqrt{\frac{2}{\pi}}\int_0^\infty \cos(ux)\widehat{f}(u)\,du$
Fourier sine transform	$\widehat{f}(u) = \sqrt{\frac{2}{\pi}}\int_0^\infty \sin(ux) f(x)\,dx$	$f(x) = \sqrt{\frac{2}{\pi}}\int_0^\infty \sin(ux)\widehat{f}(u)\,du$
Hankel transform	$\widehat{f}_\nu(u) = \int_0^\infty x\,J_\nu(ux) f(x)\,dx$	$f(x) = \int_0^\infty u\,J_\nu(ux)\widehat{f}_\nu(u)\,du$
Hartley transform	$\widehat{f}(u) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} [\cos(ux)+\sin(ux)]\,f(x)\,dx$	$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} [\cos(ux)+\sin(ux)]\,\widehat{f}(u)\,du$
Hilbert transform	$\widehat{f}(y) = \frac{1}{\pi}\,\mathrm{PV}\!\int_{-\infty}^{\infty} \frac{f(x)}{x-y}\,dx$	$f(x) = -\frac{1}{\pi}\,\mathrm{PV}\!\int_{-\infty}^{\infty} \frac{\widehat{f}(y)}{y-x}\,dy$
Kontorovich–Lebedev transform	$\widehat{f}(u) = \int_0^\infty K_{iu}(x) f(x)\,dx$	$f(x) = \frac{2}{\pi^2 x}\int_0^\infty u\sinh(\pi u) K_{iu}(x)\widehat{f}(u)\,du$
Laplace transform	$\widehat{f}(p) = \int_0^\infty e^{-px} f(x)\,dx$	$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widehat{f}(p)\,dp$
Two-sided Laplace transform	$\widehat{f}(p) = \int_{-\infty}^{\infty} e^{-px} f(x)\,dx$	$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{px}\widehat{f}(p)\,dp$

Meijer transform (K-transform)	
$\widehat{f}_\nu(u) = \sqrt{\frac{2}{\pi}}\int_0^\infty \sqrt{ux}K_\nu(ux)f(x)\,dx$	$f(x) = \frac{1}{i\sqrt{2\pi}}\int_{c-i\infty}^{c+i\infty} \sqrt{ux}I_\nu(ux)\widehat{f}_\nu(u)\,du$
Mellin transform	
$\widehat{f}(s) = \int_0^\infty x^{s-1}f(x)\,dx$	$f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} x^{-s}\widehat{f}(s)\,ds$
Y-transform	
$\widehat{f}_\nu(u) = \int_0^\infty \sqrt{ux}\,Y_\nu(ux)f(x)\,dx$	$f(x) = \int_0^\infty \sqrt{ux}\,\mathbf{H}_\nu(ux)\widehat{f}_\nu(u)\,du$

6.30 INTEGRAL TRANSFORMS: PRELIMINARIES

1. $I = (a, b)$ is an interval, where $-\infty \le a < b \le \infty$.
2. $L^1(I)$ is the set of all Lebesgue integrable functions on I. In particular, $L^1(\mathbb{R})$ is the set of all Lebesgue integrable functions on the real line $\mathbb{R}$.
3. $L^2(I)$ is the set of all square integrable functions on I (i.e., $\int_I |f(x)|^2\,dx < \infty$).
4. If f is integrable over every finite closed subinterval of I, but not necessarily on I itself, we say that f is *locally integrable* on I. For example, the function $f(x) = 1/x$ is not integrable on the interval $I = (0, 1)$, yet it is locally integrable on it.
5. A function $f(x)$, defined on a closed interval $[a, b]$, is said to be of *bounded variation* if there is an $M > 0$ such that, for any partition $a = x_0 < x_1 < \cdots < x_n = b$, the following relation holds: $\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \le M$.
6. If f has a derivative f' at every point of $[a, b]$, then by the mean value theorem, for any $a \le x < y \le b$, we have $f(x) - f(y) = f'(z)(x - y)$, for some $x < z < y$. If f' is bounded, then f is of bounded variation.
7. The left limit of a function $f(x)$ at a point t (if it exists) will be denoted by $\lim_{x\to t^-} f(x) = f(t-)$, and likewise the right limit at t will be denoted by $\lim_{x\to t^+} f(x) = f(t+)$.

6.31 FOURIER INTEGRAL TRANSFORM

The origin of the Fourier integral transformation can be traced to Fourier's celebrated work on the *Analytical Theory of Heat*, which appeared in 1822. Fourier's major finding was to show that an "arbitrary" function defined on a finite interval could be expanded in a trigonometric series. In an attempt to extend his results to functions defined on the infinite interval $(-\infty, \infty)$, Fourier introduced what is now known as the Fourier integral transform.

The *Fourier integral transform* of a function $f(t)$ is defined by

$$\mathcal{F}(f)(\omega) = \hat{f}(\omega) = F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(t) e^{it\omega}\, dt, \tag{6.31.1}$$

whenever the integral exists.

There is no universal agreement on the definition of the Fourier integral transform. Some authors take the kernel of the transformation as $e^{-it\omega}$, so that the kernel of the inverse transformation is $e^{it\omega}$. In either case, if we define the Fourier transforms

$$\hat{f}(\omega) = a \int_{-\infty}^{\infty} f(t) e^{\pm it\omega}\, dt,$$

then its inverse is

$$f(t) = b \int_{-\infty}^{\infty} \hat{f}(\omega) e^{\mp it\omega}\, d\omega,$$

for some constants a and b, with $ab = 1/2\pi$. Again there is no agreement on the choice of the constants; sometimes one of them is taken as 1 so that the other is $1/(2\pi)$. For the sake of symmetry, we choose $a = b = 1/\sqrt{2\pi}$. The functions f and $\hat{f}$ are called a *Fourier transform pair.*

A definition popular in the engineering literature is the one in which the kernel of the transform is taken as $e^{2\pi it\omega}$ (or $e^{-2\pi it\omega}$) so that the kernel of the inverse transform is $e^{-2\pi it\omega}$ (or $e^{2\pi it\omega}$). The main advantage of this definition is that the constants a and b disappear and the Fourier transform pair becomes

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t) e^{\pm 2\pi it\omega}\, dt \qquad \text{and} \qquad f(t) = \int_{-\infty}^{\infty} \hat{f}(\omega) e^{\mp 2\pi it\omega}\, d\omega. \tag{6.31.2}$$

The Fourier cosine and sine coefficients of $f(t)$ are defined by

$$a(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \cos \omega t\, dt \quad \text{and} \quad b(\omega) = \frac{1}{\pi} \int_{-\infty}^{\infty} f(t) \sin \omega t\, dt. \tag{6.31.3}$$

The Fourier cosine and sine coefficients are related to the Fourier cosine and sine integral transforms. For example, if f is even, then $a(\omega) = \sqrt{2/\pi} F_c(\omega)$ and, if f is odd, $b(\omega) = \sqrt{2/\pi} F_s(\omega)$ (see Section 6.31.7).

Two other integrals related to the Fourier integral transform are Fourier's repeated integral and the allied integral. *Fourier's repeated integral*, $S(f, t)$, of $f(t)$ is defined by

$$\begin{aligned} S(f,t) &= \int_0^{\infty} \left(a(\omega) \cos t\omega + b(\omega) \sin t\omega \right)\, d\omega, \\ &= \frac{1}{\pi} \int_0^{\infty} d\omega \int_{-\infty}^{\infty} f(x) \cos \omega (t - x)\, dx. \end{aligned} \tag{6.31.4}$$

The *allied Fourier integral*, $\tilde{S}(f,t)$, of f is defined by

$$\begin{aligned}\tilde{S}(f,t) &= \int_0^\infty (b(\omega)\cos t\omega - a(\omega)\sin t\omega)\ d\omega,\\ &= \frac{1}{\pi}\int_0^\infty d\omega \int_{-\infty}^\infty f(x)\sin\omega(x-t)dx.\end{aligned} \tag{6.31.5}$$

6.31.1 PROPERTIES

1. *Linearity*: The Fourier transform is linear,
$$\mathcal{F}\left[af(t)+bg(t)\right](\omega) = a\mathcal{F}\left[f(t)\right](\omega) + b\mathcal{F}\left[g(t)\right](\omega) = a\hat{f}(\omega) + b\hat{g}(\omega),$$
where a and b are complex numbers.
2. *Translation*: $\mathcal{F}\left[f(t-b)\right](\omega) = e^{ib\omega}\hat{f}(\omega)$.
3. *Dilation (scaling)*: $\mathcal{F}\left[f(at)\right](\omega) = \frac{1}{a}\hat{f}(\frac{\omega}{a}), \quad a>0.$
4. *Translation and dilation*: $\mathcal{F}\left[f(at-b)\right](\omega) = \frac{1}{a}e^{ib\omega/a}\hat{f}\left(\frac{\omega}{a}\right), \quad a>0.$
5. *Complex conjugation*: $\mathcal{F}\left[\overline{f(t)}\right](\omega) = \overline{\hat{f}(-\omega)}$.
6. *Modulation*: $\mathcal{F}\left[e^{iat}f(t)\right](\omega) = \hat{f}(\omega+a)$, and
$$\mathcal{F}\left[e^{iat}f(bt)\right](\omega) = \frac{1}{b}\hat{f}\left(\frac{\omega+a}{b}\right), \quad b>0. \tag{6.31.6}$$
7. *Differentiation*: If $f^{(k)} \in L^1(\mathbb{R})$, for $k=0,1,2,\cdots,n$ and $\lim_{|t|\to\infty} f^{(k)}(t) = 0$ for $k=0,1,2,\cdots,n-1$, then
$$\mathcal{F}\left[f^{(n)}(t)\right](\omega) = (-i\omega)^n\hat{f}(\omega). \tag{6.31.7}$$
8. *Integration*: Let $f \in L^1(\mathbb{R})$, and define $g(x) = \int_{-\infty}^x f(t)\,dt$. If $g \in L^1(\mathbb{R})$, then $\hat{g}(\omega) = -\hat{f}(\omega)/(i\omega)$.
9. *Multiplication by polynomials*: If $t^k f(t) \in L^1(\mathbb{R})$ for $k=0,1,\ldots,n$, then
$$\mathcal{F}\left[t^k f(t)\right](\omega) = \frac{1}{(i)^k}\hat{f}^{(k)}(\omega), \tag{6.31.8}$$
and hence,
$$\mathcal{F}\left[\left(\sum_{k=0}^n a_k t^k\right) f(t)\right](\omega) = \sum_{k=0}^n \frac{a_k}{(i)^k}\hat{f}^{(k)}(\omega). \tag{6.31.9}$$
10. *Convolution*: The convolution operation, $\star$, associated with the Fourier transform is defined as
$$h(t) = (f\star g)(t) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^\infty f(x)g(t-x)dx,$$
where f and g are defined over the whole real line.

THEOREM 6.31.1

If f and g belong to $L^1(\mathbb{R})$, then so does h. Moreover, $\hat{h}(\omega) = \hat{f}(\omega)\hat{g}(\omega)$. If $\hat{f}$ and $\hat{g}$ belong to $L^1(\mathbb{R})$, then $(\hat{f}\star\hat{g})(\omega) = \widehat{(fg)}(\omega)$.

11. *Parseval's relation*: If $f, g \in L^2(\mathbb{R})$, and if F and G are the Fourier transforms of f and g, respectively, then Parseval's relation is

$$\int_{-\infty}^{\infty} F(\omega)G(\omega)\,d\omega = \int_{-\infty}^{\infty} f(t)g(-t)\,dt. \tag{6.31.10}$$

Replacing G by $\overline{G}$ (so that $g(-t)$ is replaced by $\overline{g}(t)$) results in a more convenient form of Parseval's relation

$$\int_{-\infty}^{\infty} F(\omega)\overline{G}(\omega)\,d\omega = \int_{-\infty}^{\infty} f(t)\overline{g}(t)\,dt. \tag{6.31.11}$$

In particular, for $f = g$,

$$\int_{-\infty}^{\infty} |F(\omega)|^2\,d\omega = \int_{-\infty}^{\infty} |f(t)|^2\,dt. \tag{6.31.12}$$

6.31.2 EXISTENCE

For the Fourier integral transform to exist, it is sufficient that f be absolutely integrable on $(-\infty, \infty)$, i.e., $f \in L^1(\mathbb{R})$.

THEOREM 6.31.2 *(Riemann–Lebesgue lemma)*
If $f \in L^1(\mathbb{R})$, then its Fourier transform $\hat{f}(\omega)$ is defined everywhere, uniformly continuous, and tends to zero as $\omega \to \pm\infty$.

The uniform continuity follows from the relationship

$$\left|\hat{f}(\omega+h) - \hat{f}(\omega)\right| \le \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} |f(t)| \left|e^{iht} - 1\right| dt,$$

and the tendency toward zero as $\omega \to \pm\infty$ is a consequence of the Riemann–Lebesgue lemma.

THEOREM 6.31.3 *(Generalized Riemann–Lebesgue lemma)*
Let $f \in L^1(I)$, where $I = (a, b)$ is finite or infinite and let ω be real.
Let $a \le a' < b' \le b$ and $\hat{f}_\omega(\lambda, a', b') = \int_{a'}^{b'} f(t)e^{i\lambda\omega t}\,dt$.
Then $\lim_{\omega\to\pm\infty} \hat{f}_\omega(\lambda, a', b') = 0$, and the convergence is uniform in a' and b'.
In particular, $\lim_{\omega\to\pm\infty} \int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt = 0$.

6.31.3 INVERSION FORMULA

Many of the theorems on the inversion of the Fourier transform are based on *Dini's condition* which can be stated as follows:

If $f \in L^1(\mathbb{R})$, then a necessary and sufficient condition for

$$S(f, x) = \lim_{\lambda\to\infty} \frac{1}{\pi} \int_{-\infty}^{\infty} f(t)\frac{\sin\lambda(x-t)}{(x-t)}\,dt = a \tag{6.31.13}$$

is that

$$\lim_{\lambda\to\infty}\int_0^{\delta}\left(f(x+y)+f(x-y)-2a\right)\frac{\sin\lambda y}{y}dy=0, \tag{6.31.14}$$

for any fixed $\delta > 0$. By the Riemann–Lebesgue lemma, this condition is satisfied if

$$\int_0^{\delta}\left|\frac{f(x+y)+f(x-y)-2a}{y}\right|dy<\infty, \tag{6.31.15}$$

for some $\delta > 0$. In particular, condition (6.31.15) holds for $a = f(x)$, if f is differentiable at x, and for $a = [f(x+) + f(x-)]/2$, if f is of bounded variation in a neighborhood of x.

THEOREM 6.31.4 *(Inversion theorem)*
Let f be a locally integrable function, of bounded variation in a neighborhood of the point x. If f satisfies either one of the following conditions:

1. *$f(t) \in L^1(\mathbb{R})$, or*
2. *$f(t)/(1+|t|) \in L^1(\mathbb{R})$, and the integral $\int_{-\infty}^{\infty} f(t)e^{i\omega t}\,dt$ converges uniformly on every finite interval of ω,*

then

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\hat{f}(\omega)e^{-ix\omega}\,d\omega=\lim_{\lambda\to\infty}\frac{1}{\sqrt{2\pi}}\int_{-\lambda}^{\lambda}\hat{f}(\omega)e^{-ix\omega}\,d\omega$$

is equal to $[f(x+) + f(x-)]/2$ whenever the expression has meaning, to $f(x)$ whenever $f(x)$ is continuous at x, and to $f(x)$ almost everywhere. If f is continuous and of bounded variation in the interval (a, b), then the convergence is uniform in any interval interior to (a, b).

6.31.4 UNCERTAINTY PRINCIPLE

Let T and W be two real numbers defined by

$$T^2=\frac{1}{E}\int_{-\infty}^{\infty}t^2|f(t)|^2\,dt \quad \text{and} \quad W^2=\frac{1}{E}\int_{-\infty}^{\infty}\omega^2|\hat{f}(\omega)|^2\,d\omega, \tag{6.31.16}$$

where

$$E=\int_{-\infty}^{\infty}|f(t)|^2\,dt=\int_{-\infty}^{\infty}|\hat{f}(\omega)|^2\,d\omega. \tag{6.31.17}$$

Assuming that f is differentiable and $\lim_{|t|\to\infty} tf^2(t) = 0$, then $2TW \geq 1$, or

$$\left(\int_{-\infty}^{\infty}t^2|f(t)|^2\,dt\right)^{1/2}\left(\int_{-\infty}^{\infty}\omega^2|\hat{f}(\omega)|^2\,d\omega\right)^{1/2}\geq\frac{1}{2}\int_{-\infty}^{\infty}|f(t)|^2\,dt. \tag{6.31.18}$$

This means that f and $\hat{f}$ cannot both be very small. Another related property of the Fourier transform is that, if either one of the functions f or $\hat{f}$ vanishes outside some finite interval, then the other one must trail on to infinity. In other words, they cannot both vanish outside any finite interval.

6.31.5 POISSON SUMMATION FORMULA

The Poisson summation formula may be written in the form

$$\frac{1}{\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} f\left(t + \frac{k\pi}{\sigma}\right) = \frac{\sigma}{\pi} \sum_{k=-\infty}^{\infty} \hat{f}(2k\sigma) e^{-2ikt\sigma}, \quad \sigma > 0, \tag{6.31.19}$$

provided that the two series converge. A sufficient condition for the validity of Equation (6.31.19) is that $f = O(1 + |t|)^{-\alpha}$ as $|t| \to \infty$, and $\hat{f} = O\left((1 + |\omega|)^{-\alpha}\right)$ as $|\omega| \to \infty$ for some $\alpha > 1$.

Another version of the Poisson summation formula is

$$\sum_{k=-\infty}^{\infty} \hat{f}(\omega + k\sigma)\overline{\hat{g}}(\omega + k\sigma) = \frac{1}{\sigma} \sum_{k=-\infty}^{\infty} \left(\int_{-\infty}^{\infty} f(t)\overline{g}\left(t - \frac{2\pi k}{\sigma}\right) dt \right) e^{2\pi i k\omega/\sigma}. \tag{6.31.20}$$

6.31.6 SHANNON'S SAMPLING THEOREM

If f is a function *band-limited* to $[-\sigma, \sigma]$, i.e., $f(t) = \frac{1}{\sqrt{2\pi}} \int_{-\sigma}^{\sigma} F(\omega) e^{it\omega}\, d\omega$, with $F \in L^2(-\sigma, \sigma)$, then it can be reconstructed from its sample values at the points $t_k = (k\pi)/\sigma$, $k = 0, \pm 1, \pm 2, \cdots$, via the formula

$$f(t) = \sum_{k=-\infty}^{\infty} f(t_k) \frac{\sin \sigma(t - t_k)}{\sigma(t - t_k)}, \tag{6.31.21}$$

with the series absolutely and uniformly convergent on compact sets.

The series on the right-hand side of Equation (6.31.21) can be written as $\sin \sigma t \sum_{k=-\infty}^{\infty} f(t_k) \frac{(-1)^k}{(\sigma t - k\pi)}$, which is a special case of a Cardinal series (these series have the form $\sin \sigma t \sum_{k=-\infty}^{\infty} C_k \frac{(-1)^k}{(\sigma t - k\pi)}$).

6.31.7 FOURIER SINE AND COSINE TRANSFORMS

The *Fourier cosine transform*, $F_c(\omega)$, and the *Fourier sine transform*, $F_s(\omega)$, of $f(t)$ are defined for $\omega > 0$ as

$$F_c(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t) \cos \omega t\, dt \quad \text{and} \quad F_s(\omega) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} f(t) \sin \omega t\, dt. \tag{6.31.22}$$

The inverse transforms have the same functional form:

$$f(t) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} F_c(\omega) \cos \omega t\, d\omega = \sqrt{\frac{2}{\pi}} \int_0^{\infty} F_s(\omega) \sin \omega t\, d\omega. \tag{6.31.23}$$

If f is even, i.e., $f(t) = f(-t)$, then $F(\omega) = F_c(\omega)$, and if f is odd, i.e., $f(t) = -f(-t)$, then $F(\omega) = iF_s(\omega)$.

6.32 DISCRETE FOURIER TRANSFORM (DFT)

The *discrete Fourier transform* of the sequence $\{a_n\}_{n=0}^{N-1}$, where $N \geq 1$, is a sequence $\{A_m\}_{m=0}^{N-1}$, defined by

$$A_m = \sum_{n=0}^{N-1} a_n (W_N)^{mn}, \qquad \text{for} \quad m = 0, 1, \cdots, N-1, \tag{6.32.1}$$

where $W_N = e^{2\pi i/N}$. Note that $\sum_{m=0}^{N-1} W_N^{m(k-n)} = N\delta_{kn}$. For example, the DFT of the sequence $\{1, 0, 1, 1\}$ is $\{3, -i, 1, i\}$.

The inversion formula is

$$a_n = \frac{1}{N} \sum_{m=0}^{N-1} A_m W_N^{-mn}, \quad n = 0, 1, \cdots, N-1. \tag{6.32.2}$$

Equations (6.32.1) and (6.32.2) are called a discrete Fourier transform (DFT) pair of order N. The factor $1/N$ and the negative sign in the exponent of W_N that appear in Equation (6.32.2) are sometimes introduced in Equation (6.32.1) instead. We use the notation

$$\mathcal{F}_N[(a_n)] = A_m \qquad \mathcal{F}_N^{-1}[(A_m)] = a_n, \tag{6.32.3}$$

to indicate that the discrete Fourier transform of order N of the sequence $\{a_n\}$ is $\{A_m\}$ and that the inverse transform of $\{A_m\}$ is $\{a_n\}$.

Because $W_N^{\pm(m+N)n} = W_N^{\pm mn}$, Equations (6.32.1) and (6.32.2) can be used to extend the sequences $\{a_n\}_{n=0}^{N-1}$ and $\{A_m\}_{m=0}^{N-1}$, as periodic sequences with period N. This means that $A_{m+N} = A_m$, and $a_{n+N} = a_n$. This will be used in what follows without explicit note. Using this, the summation limits, 0 and $N-1$, can be replaced with n_1 and $n_1 + N - 1$, respectively, where n_1 is any integer. In the special case where $n_1 = -M$ and $N = 2M + 1$, Equations (6.32.1) and (6.32.2) become

$$A_m = \sum_{n=-M}^{M} a_n W_N^{mn}, \qquad \text{for} \quad m = -M, -M+1, \cdots, M-1, M, \tag{6.32.4}$$

and

$$a_n = \frac{1}{2M+1} \sum_{m=-M}^{M} A_m W_N^{-mn}, \qquad \text{for} \quad n = -M, -M+1, \cdots, M-1, M. \tag{6.32.5}$$

6.32.1 PROPERTIES

1. *Linearity:* The discrete Fourier transform is linear, that is

$$\mathcal{F}_N[\alpha(a_n) + \beta(b_n)] = \alpha A_m + \beta B_m,$$

for any complex numbers α and β, where the sum of two sequences is defined as $(a_n) + (b_n) = (a_n + b_n)$.

2. *Translation:* $\mathcal{F}_N[(a_{n-k})] = W_N^{mk} A_m$, or $e^{2\pi imk/N} A_m = \sum_{n=0}^{N-1} a_{n-k} W_N^{mn}$.

3. *Modulation:* $\mathcal{F}_N[(W_N^{nk} a_n)] = A_{m+k}$, or $A_{m+k} = \sum_{n=0}^{N-1} e^{2\pi ink/N} a_n W_N^{mn}$.

4. *Complex Conjugation:* $\mathcal{F}_N[(\overline{a}_{-n})] = \overline{A}_m$, or $\overline{A}_m = \sum_{n=0}^{N-1} \overline{a}_{-n} W_N^{mn}$.

5. *Symmetry:* $\mathcal{F}_N[(a_{-n})] = A_{-m}$, or $A_{-m} = \sum_{n=0}^{N-1} a_{-n} W_N^{mn}$.

6. *Convolution*: The convolution of the sequences $\{a_n\}_{n=0}^{N-1}$ and $\{b_n\}_{n=0}^{N-1}$ is the sequence $\{c_n\}_{n=0}^{N-1}$ given by

$$c_n = \sum_{k=0}^{N-1} a_k b_{n-k}. \tag{6.32.6}$$

The *convolution relation of the DFT* is $\mathcal{F}_N[(c_n)] = \mathcal{F}_N[(a_n)]\mathcal{F}_N[(b_n)]$, or $C_m = A_m B_m$. A consequence of this and Equation (6.32.2), is the relation

$$\sum_{k=0}^{N-1} a_k b_{n-k} = \frac{1}{N} \sum_{m=0}^{N-1} A_m B_m W_N^{-mn}. \tag{6.32.7}$$

7. *Parseval's relation*:

$$\sum_{n=0}^{N-1} a_n \overline{d}_n = \frac{1}{N} \sum_{m=0}^{N-1} A_m \overline{D}_m. \tag{6.32.8}$$

In particular,

$$\sum_{n=0}^{N-1} |a_n|^2 = \frac{1}{N} \sum_{m=0}^{N-1} |A_m|^2. \tag{6.32.9}$$

In (4) and (5), the fact that $\overline{W}_N = W_N^{-1}$ has been used. A sequence $\{a_n\}$ is said to be even if $\{a_{-n}\} = \{a_n\}$ and is said to be odd if $\{a_{-n}\} = \{-a_n\}$. The following are consequences of (4) and (5):

1. If $\{a_n\}$ is a sequence of real numbers (i.e., $\overline{a_n} = a_n$), then $\overline{A_m} = A_{-m}$.
2. $\{a_n\}$ is real and even if and only if $\{A_m\}$ is real and even.
3. $\{a_n\}$ is real and odd if and only if $\{A_m\}$ is pure imaginary and odd.

6.33 FAST FOURIER TRANSFORM (FFT)

To determine A_m for each $m = 0, 1, \cdots, M-1$ (using Equation (6.32.1)), $M-1$ multiplications are required. Hence the total number of multiplications required to determine all the A_m's is $(M-1)^2$. This number can be reduced by using decimation.

Assuming M is even, we define $M = 2N$ and write

$$\mathcal{F}_{2N}[(a_n)] = A_m. \tag{6.33.1}$$

Now split $\{a_n\}$ into two sequences, one consisting of terms with even subscripts ($b_n = a_{2n}$) and one with odd subscripts ($c_n = a_{2n+1}$). Then

$$A_m = B_m + W_{2N}^m C_m. \tag{6.33.2}$$

For the evaluation of B_m and C_m, the total number of multiplications required is $2(N-1)^2$. To determine A_m from Equation (6.33.2), we must calculate the product $W_{2N}^m C_m$, for each fixed m. Therefore, the total number of multiplications required to determine A_m from Equation (6.33.2) is $2(N-1)^2 + 2N - 1 = 2N^2 - 2N + 1$.

But if we had determined A_m from (6.33.1), we would have performed $(2N-1)^2 = 4N^2 - 4N + 1$ multiplications. Thus, splitting the sequence $\{a_n\}$ into two sequences and then applying the discrete Fourier transform reduces the number of multiplications required to evaluate A_m approximately by a factor of 2.

If N is even, this process can be repeated. Split $\{b_n\}$ and $\{c_n\}$ into two sequences, each of length $N/2$. Then B_m and C_m are determined in terms of four discrete Fourier transforms, each of order $N/2$. This process can be repeated $k-1$ times if $M = 2^k$ for some positive integer k.

If we denote the required number of multiplications for the discrete Fourier transform of order $N = 2^k$ by $F(N)$, then $F(2N) = 2F(N) + N$ and $F(2) = 1$, which leads to $F(N) = \frac{N}{2}\log_2 N$.

6.34 MULTIDIMENSIONAL FOURIER TRANSFORMS

If $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ and $\mathbf{u} = (u_1, u_2, \ldots, u_n)$, then (see the table on page 522):

1. Fourier transform $\qquad F(\mathbf{u}) = (2\pi)^{-n/2} \underset{\mathbb{R}^n}{\int\cdots\int} f(\mathbf{x}) e^{i(\mathbf{x}\cdot\mathbf{u})}\, d\mathbf{x}.$
2. Inverse Fourier transform $\qquad f(\mathbf{x}) = (2\pi)^{-n/2} \underset{\mathbb{R}^n}{\int\cdots\int} F(\mathbf{u}) e^{-i(\mathbf{x}\cdot\mathbf{u})}\, d\mathbf{u}.$
3. Parseval's relation $\qquad \underset{\mathbb{R}^n}{\int\cdots\int} f(\mathbf{x})\overline{g(\mathbf{x})}\, d\mathbf{x} = \underset{\mathbb{R}^n}{\int\cdots\int} F(\mathbf{u})\overline{G(\mathbf{u})}\, d\mathbf{u}.$

6.35 HANKEL TRANSFORM

The *Hankel transform* of order ν of a real-valued function $f(x)$ is defined as

$$\mathcal{H}_\nu(f)(y) = F_\nu(y) = \int_0^\infty f(x)\sqrt{xy}J_\nu(yx)dx, \tag{6.35.1}$$

for $y > 0$ and $\nu > -1/2$, where $J_\nu(z)$ is the Bessel function of the first kind of order ν.

The Hankel transforms of order $1/2$ and $-1/2$ are equal to the Fourier sine and cosine transforms, respectively, because

$$J_{1/2}(x) = \sqrt{\frac{2}{\pi x}}\sin x, \qquad J_{-1/2}(x) = \sqrt{\frac{2}{\pi x}}\cos x. \tag{6.35.2}$$

As with the Fourier transform, there are many variations on the definition of the Hankel transform. Some authors define it as

$$G_\nu(y) = \int_0^\infty xg(x)J_\nu(yx)dx; \tag{6.35.3}$$

however, the two definitions are equivalent; we only need to replace $f(x)$ by $\sqrt{x}g(x)$ and $F_\nu(y)$ by $\sqrt{y}G_\nu(y)$.

6.35.1 PROPERTIES

1. *Existence*: Since $\sqrt{x}J_\nu(x)$ is bounded on the positive real axis, the Hankel transform of f exists if $f \in L^1(0,\infty)$.
2. *Multiplication by x^m*:

$$\mathcal{H}_\nu\left(x^m f(x)\right)(y) = y^{1/2-\nu}\left(\frac{1}{y}\frac{d}{dy}\right)^m \left[y^{\nu+m-1/2}F_{\nu+m}(y)\right].$$

3. *Division by x*:

$$\mathcal{H}_\nu\left(\frac{2\nu}{x}f(x)\right)(y) = y\left[F_{\nu-1}(y) + F_{\nu+1}(y)\right],$$
$$\mathcal{H}_\nu\left(\frac{f(x)}{x}\right)(y) = y^{1/2-\nu}\int_0^y t^{\nu-1/2}F_{\nu-1}(t)\,dt.$$

4. *Differentiation*:

$$\mathcal{H}_\nu\left(2\nu f'(x)\right)(y) = (\nu - \tfrac{1}{2})yF_{\nu+1}(y) - (\nu + \tfrac{1}{2})yF_{\nu-1}(y).$$

5. *Differentiation and multiplication by powers of x*:

$$\mathcal{H}_\nu\left[x^{1/2-\nu}\left(\frac{1}{x}\frac{d}{dx}\right)^m \left(x^{\nu+m-1/2}f(x)\right)\right](y) = y^m F_{\nu+m}(y).$$

6. *Parseval's relation*: Let F_ν and G_ν denote the Hankel transforms of order ν of f and g, respectively. Then

$$\int_0^\infty F_\nu(y)G_\nu(y)dy = \int_0^\infty f(x)g(x)dx. \tag{6.35.4}$$

In particular, $\int_0^\infty |F_\nu(y)|^2\,dy = \int_0^\infty |f(x)|^2 dx$.

7. *Inversion formula*: If f is absolutely integrable on $(0,\infty)$ and of bounded variation in a neighborhood of point x, then

$$\int_0^\infty F_\nu(y)\sqrt{xy}J_\nu(yx)dy = \frac{f(x+)+f(x-)}{2}, \tag{6.35.5}$$

whenever the expression on the right-hand side of the equation has a meaning; the integral converges to $f(x)$ whenever f is continuous at x.

6.36 HARTLEY TRANSFORM

Define the function $\operatorname{cas} x = \cos x + \sin x = \sqrt{2}\sin(x+\frac{\pi}{4})$. The *Hartley transform* of the real function $g(t)$ is

$$(Hg)(\omega) = \int_{-\infty}^{\infty} g(t)\operatorname{cas}(2\pi\omega t)\,dt. \tag{6.36.1}$$

Let $(Eg)(\omega)$ and $(Og)(\omega)$ be the even and odd parts of $(Hg)(\omega)$,

$$\begin{aligned}(Eg)(\omega) &= \frac{1}{2}\left((Hg)(\omega) + (Hg)(-\omega)\right),\\ (Og)(\omega) &= \frac{1}{2}\left((Hg)(\omega) - (Hg)(-\omega)\right),\end{aligned} \tag{6.36.2}$$

so that $(Hg)(\omega) = (Eg)(\omega)+(Og)(\omega)$. The Fourier transform of g (using the kernel $e^{2\pi i\omega t}$) can then be written in terms of $(Eg)(\omega)$ and $(Og)(\omega)$ as

$$\int_{-\infty}^{\infty} g(t)e^{2\pi i\omega t}\,dt = (Eg)(\omega) + i(Og)(\omega). \tag{6.36.3}$$

Notes:

1. The Hartley transform, applied twice, returns the original function.
2. Note that $\frac{d}{dx}\operatorname{cas}(x) = \operatorname{cas}(-x)$.
3. There is a fast Discrete Hartley transform (DHT) similar to the DFT.
4. For even functions the Hartley and Fourier transforms are the same.

6.37 HILBERT TRANSFORM

The *Hilbert transform* of f is defined as

$$(\mathcal{H}f)(x) = \tilde{f}(x) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(t)}{t-x}\,dt = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(x+t)}{t}\,dt \tag{6.37.1}$$

where the integral is a Cauchy principal value. See the table on page 524.

Since the definition is given in terms of a singular integral, it is sometimes impractical to use. An alternative definition is given below. First, let f be an integrable function, and define $a(t)$ and $b(t)$ by

$$a(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\cos tx dx, \qquad b(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} f(x)\sin tx\,dx. \tag{6.37.2}$$

Consider the function $F(z)$, defined by the integral

$$F(z) = \int_0^{\infty} (a(t) - ib(t))e^{izt}\,dt = U(z) + i\tilde{U}(z), \tag{6.37.3}$$

where $z = x + iy$. The real and imaginary parts of F are

$$\begin{aligned} U(z) &= \int_0^{\infty} (a(t)\cos xt + b(t)\sin xt)e^{-yt}\,dt, \quad \text{and} \\ \tilde{U}(z) &= \int_0^{\infty} (a(t)\sin xt - b(t)\cos xt)e^{-yt}\,dt. \end{aligned} \tag{6.37.4}$$

Formally,

$$\lim_{y\to 0} U(z) = f(x) = \int_0^{\infty} (a(t)\cos xt + b(t)\sin xt)\,dt, \tag{6.37.5}$$

and

$$\lim_{y\to 0} \tilde{U}(z) = -\tilde{f}(x) = \int_0^{\infty} (a(t)\sin xt - b(t)\cos xt)\,dt, \tag{6.37.6}$$

The Hilbert transform of a function f, given by Equation (6.37.5), is defined as the function $\tilde{f}$ given by Equation (6.37.6).

6.37.1 EXISTENCE

If $f \in L^1(\mathbb{R})$, then its Hilbert transform $(\mathcal{H}f)(x)$ exists for almost all x. For $f \in L^p(\mathbb{R}), p > 1$, there is the following stronger result:

THEOREM 6.37.1

Let $f \in L^p(\mathbb{R})$ for $1 < p < \infty$. Then $(\mathcal{H}f)(x)$ exists for almost all x and defines a function that also belongs to $L^p(\mathbb{R})$ with

$$\int_{-\infty}^{\infty} |(\mathcal{H}f)(x)|^p\,dx \le C_p \int_{-\infty}^{\infty} |f(x)|^p\,dx. \tag{6.37.7}$$

In the special case of $p = 2$*, we have*

$$\int_{-\infty}^{\infty} |(\mathcal{H}f)(x)|^2 \, dx = \int_{-\infty}^{\infty} |f(x)|^2 \, dx. \tag{6.37.8}$$

The theorem is not valid if $p = 1$ because, although it is true that $(\mathcal{H}f)(x)$ is defined almost everywhere, it is not necessarily in $L^1(\mathbb{R})$. The function $f(t) = (t \log^2 t)^{-1} H(t)$ provides a counterexample.

6.37.2 PROPERTIES

1. *Translation*: The Hilbert transform commutes with the translation operator $(\mathcal{H}f)(x + a) = \mathcal{H}(f(t + a))(x)$.
2. *Dilation*: The Hilbert transformation also commutes with the dilation operator

   $$(\mathcal{H}f)(ax) = \mathcal{H}(f(at))(x) \quad a > 0,$$

   but

   $$(\mathcal{H}f)(ax) = -\mathcal{H}(f(at))(x) \quad \text{for } a < 0.$$

3. *Multiplication by* t: $\mathcal{H}(tf(t))(x) = x(\mathcal{H}f)(x) + \dfrac{1}{\pi} \displaystyle\int_{-\infty}^{\infty} f(t) \, dt.$
4. *Differentiation*: $\mathcal{H}(f'(t))(x) = (\mathcal{H}f)'(x)$, provided that $f(t) = O(t)$ as $|t| \to \infty$.
5. *Orthogonality*: The Hilbert transform of $f \in L^2(\mathbb{R})$ is orthogonal to f in the sense $\int_{-\infty}^{\infty} f(x)(\mathcal{H}f)(x)dx = 0$.
6. *Parity*: The Hilbert transform of an even function is odd and that of an odd function is even.
7. *Inversion formula*: If $(\mathcal{H}f)(x) = \dfrac{1}{\pi} \displaystyle\int_{-\infty}^{\infty} \frac{f(t)}{t - x} \, dt$, then

   $f(t) = -\dfrac{1}{\pi} \displaystyle\int_{-\infty}^{\infty} \frac{(\mathcal{H}f)(x)}{x - t} dx$ or, symbolically,

   $$\mathcal{H}\left(\mathcal{H}f\right)(x) = -f(x), \tag{6.37.9}$$

   that is, applying the Hilbert transform twice returns the negative of the original function. Moreover, if $f \in L^1(\mathbb{R})$ has a bounded derivative, then the allied integral (see Equation 6.31.5) equals $(\mathcal{H}f)(x)$.
8. *Meromorphic invariance*:

   $$(\mathcal{H}f)(u(x)) = \tilde{f}(u(x)) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(t)}{t - u(x)} \, dt = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f[u(t)]}{t - x} \, dt$$

   where

   $$u(t) = t - \sum_{n=1}^{\infty} \frac{a_n}{t - b_n},$$

   for arbitrary $a_n \geq 0$ and b_n real.

6.37.3 RELATIONSHIP WITH THE FOURIER TRANSFORM

From Equations (6.37.3)–(6.37.6), we obtain

$$\lim_{y\to 0} F(z) = F(x) = \int_0^\infty (a(t) - ib(t))e^{ixt}\,dt = f(x) - i(\mathcal{H}f)(x),$$

where $a(t)$ and $b(t)$ are given by Equation (6.37.2).

Let g be a real-valued integrable function and consider its Fourier transform $\hat{g}(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(t)e^{ixt}\,dt$. If we denote the real and imaginary parts of $\hat{g}$ by f and $\tilde{f}$, respectively, then

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(t)\cos xt\,dt, \qquad \text{and}$$
$$\tilde{f}(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g(t)\sin xt\,dt.$$

Splitting g into its even and odd parts, g_e and g_o, respectively, we obtain

$$g_e(t) = \frac{g(t)+g(-t)}{2} \quad \text{and} \quad g_o(t) = \frac{g(t)-g(-t)}{2};$$

hence

$$f(x) = \sqrt{\frac{2}{\pi}}\int_0^\infty g_e(t)\cos xt\,dt, \quad \text{and} \quad \tilde{f}(x) = \sqrt{\frac{2}{\pi}}\int_0^\infty g_o(t)\sin xt\,dt, \tag{6.37.10}$$

or

$$f(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_e(t)e^{ixt}\,dt, \quad \text{and} \quad \tilde{f}(x) = \frac{-i}{\sqrt{2\pi}}\int_{-\infty}^{\infty} g_o(t)e^{ixt}\,dt. \tag{6.37.11}$$

This shows that, if the Fourier transform of the even part of a real-valued function represents a function $f(x)$, then the Fourier transform of the odd part represents the Hilbert transform of f (up to multiplication by i).

THEOREM 6.37.2

Let $f \in L^1(\mathbb{R})$ and assume that $\mathcal{H}f$ is also in $L^1(\mathbb{R})$. Then

$$\mathcal{F}(\mathcal{H}f)(\omega) = -i\,\mathrm{sgn}(\omega)\mathcal{F}(f)(\omega), \tag{6.37.12}$$

where $\mathcal{F}$ denotes the Fourier transformation. Similarly, if $f \in L^2(\mathbb{R})$, then $(\mathcal{H}f) \in L^2(\mathbb{R})$, and Equation (6.37.12) remains valid.

6.38 LAPLACE TRANSFORM

The Laplace transformation dates back to the work of the French mathematician, Pierre Simon Marquis de Laplace (1749–1827), who used it in his work on probability theory in the 1780s.

The *Laplace transform* of a function $f(t)$ is defined as

$$F(s) = [\mathcal{L}f](s) = \int_0^\infty f(t)e^{-st}\,dt \tag{6.38.1}$$

(also written as $[\mathcal{L}f(t)]$ and $[\mathcal{L}(f(t))]$), whenever the integral exists for at least one value of s. The transform variable, s, can be taken as a complex number. We say that f is Laplace transformable or the Laplace transformation is applicable to f if $[\mathcal{L}f]$ exists for at least one value of s.

6.38.1 EXISTENCE AND DOMAIN OF CONVERGENCE

Sufficient conditions for the existence of the Laplace transform are

1. f is a locally integrable function on $[0,\infty)$, i.e., $\int_0^a |f(t)|\,dt < \infty$, for any $a > 0$.
2. f is of (real) *exponential type*, i.e., for some constants $M, t_0 > 0$ and real γ, f satisfies
$$|f(t)| \le Me^{\gamma t}, \qquad \text{for all } t \ge t_0. \tag{6.38.2}$$

If f is a locally integrable function on $[0,\infty)$ and of (real) exponential type γ, then the Laplace integral of f, $\int_0^\infty f(t)e^{-st}\,dt$, converges absolutely for Re $s > \gamma$ and uniformly for Re $s \ge \gamma_1$, for any $\gamma_1 > \gamma$. Consequently, $F(s)$ is analytic in the half-plane $\Omega = \{s \in \mathbb{C} : \text{Re } s > \gamma\}$. It can be shown that if $F(s)$ exists for some s_0, then it also exists for any s for which Re $s >$ Re s_0. The actual domain of existence of the Laplace transform may be larger than the one given above. For example, the function $f(t) = \cos e^t$ is of real exponential type zero, but $F(s)$ exists for Re $s > -1$.

If $f(t)$ is a locally integrable function on $[0,\infty)$, not of exponential type, and

$$\int_0^\infty f(t)e^{-s_0 t}\,dt \tag{6.38.3}$$

converges for some complex number s_0, then the Laplace integral

$$\int_0^\infty f(t)e^{-st}\,dt \tag{6.38.4}$$

converges in the region Re $(s) >$ Re (s_0) and converges uniformly in the region $|\arg(s - s_0)| \le \theta < \frac{\pi}{2}$. Moreover, if Equation (6.38.3) diverges, then so does Equation (6.38.4) for Re $s <$ Re s_0.

6.38.2 PROPERTIES

1. *Linearity*: $\mathcal{L}(\alpha f + \beta g) = \alpha\mathcal{L}(f) + \beta\mathcal{L}(g) = \alpha F + \beta G$, for any constants α and β.
2. *Dilation*: $[\mathcal{L}(f(at))](s) = \frac{1}{a}F\left(\frac{s}{a}\right)$, for $a > 0$.
3. *Multiplication by exponential functions*:

$$\left[\mathcal{L}\left(e^{at}f(t)\right)\right](s) = F(s-a).$$

4. *Translation*: $[\mathcal{L}(f(t-a)H(t-a))](s) = e^{-as}F(s)$ for $a > 0$; where H is the Heaviside function. This can be put in the form

$$[\mathcal{L}(f(t)H(t-a))](s) = e^{-as}[\mathcal{L}(f(t+a))](s),$$

Examples:

(a) If

$$g(t) = \begin{cases} 0, & 0 \le t \le a, \\ (t-a)^\nu, & a \le t, \end{cases}$$

then $g(t) = f(t-a)H(t-a)$ where $f(t) = t^\nu$ (for Re $\nu > -1$). Since $\mathcal{L}(t^\nu) = \Gamma(\nu+1)/s^{\nu+1}$, it follows that $(\mathcal{L}g)(s) = e^{-as}\Gamma(\nu+1)/s^{\nu+1}$, for Re $s > 0$.

(b) If

$$g(t) = \begin{cases} t, & 0 \le t \le a, \\ 0, & a < t, \end{cases}$$

we may write $g(t) = t[H(t) - H(t-a)] = tH(t) - (t-a)H(t-a) - aH(t-a)$. Thus by properties (1) and (4),

$$G(s) = \frac{1}{s^2} - \frac{1}{s^2}e^{-as} - \frac{a}{s}e^{-as}.$$

5. *Differentiation of the transformed function*: If f is a differentiable function of exponential type, $\lim_{t\to 0+} f(t) = f(0+)$ exists, and f' is locally integrable on $[0,\infty)$, then the Laplace transform of f' exists, and

$$(\mathcal{L}f')(s) = sF(s) - f(0). \tag{6.38.5}$$

Note that although f is assumed to be of exponential type, f' need not be. For example, $f(t) = \sin e^{t^2}$, but $f'(t) = 2te^{t^2}\cos e^{t^2}$.

6. *Differentiation of higher orders*: Let f be an n times differentiable function so that $f^{(k)}$ (for $k = 0, 1, \ldots, n-1$) are of exponential type with the additional assumption that $\lim_{t\to 0+} f^{(k)}(t) = f^{(k)}(0+)$ exists. If $f^{(n)}$ is locally integrable on $[0,\infty)$, then its Laplace transform exists, and

$$\left[\mathcal{L}\left(f^{(n)}\right)\right](s) = s^nF(s) - s^{n-1}f(0) - s^{n-2}f'(0) - \ldots - f^{(n-1)}(0). \tag{6.38.6}$$

7. *Integration*: If $g(t) = \int_0^t f(x)\,dx$, then (if the transforms exist) $G(s) = F(s)/s$. Repeated applications of this rule result in

$$\left[\mathcal{L}\left(f^{(-n)}\right)\right](s) = \frac{1}{s^n}F(s), \tag{6.38.7}$$

where $f^{(-n)}$ is the n^{th} anti-derivative of f defined by $f^{(-n)}(t) = \int_0^t dt_n \int_0^{t_n} dt_{n-1} \cdots \int_0^{t_2} f(t_1)\,dt_1$. Section 6.38.1 shows that the Laplace transform is an analytic function in a half-plane. Hence it has derivatives of all orders at any point in that half-plane. The next property shows that we can evaluate these derivatives by direct calculation.

8. *Multiplication by powers of* t: Let f be a locally integrable function whose Laplace integral converges absolutely and uniformly for Re $s > \sigma$. Then F is analytic in Re $s > \sigma$, and (for $n = 0, 1, 2, \dots,$ with Re $s > \sigma$)

$$\begin{aligned} \left[\mathcal{L}\left(t^n f(t)\right)\right](s) &= \left(-\frac{d}{ds}\right)^n F(s), \\ \left[\mathcal{L}\left(\left(t\frac{d}{dt}\right)^n f(t)\right)\right](s) &= \left(-\frac{d}{ds}s\right)^n F(s), \end{aligned} \tag{6.38.8}$$

where $\left(t\frac{d}{dt}\right)^n$ is the operator $\left(t\frac{d}{dt}\right)$ applied n times.

9. *Division by powers of* t: If f is a locally integrable function of exponential type such that $f(t)/t$ is a Laplace transformable function, then

$$\left[\mathcal{L}\left(\frac{f(t)}{t}\right)\right](s) = \int_s^\infty F(u)\,du, \tag{6.38.9}$$

or, more generally,

$$\left[\mathcal{L}\left(\frac{f(t)}{t^n}\right)\right](s) = \int_s^\infty \cdots \int_{s_3}^\infty \int_{s_2}^\infty F(s_1)\,ds_1\,ds_2 \,\cdots\, ds_n \tag{6.38.10}$$

is the n^{th} repeated integral. It follows from properties (7) and (9) that

$$\left[\mathcal{L}\left(\int_0^t \frac{f(x)}{x}\,dx\right)\right](s) = \frac{1}{s}\int_s^\infty F(u)\,du. \tag{6.38.11}$$

10. *Periodic functions:* Let f be a locally integrable function that is periodic with period T. Then

$$\left[\mathcal{L}(f)\right](s) = \frac{1}{(1 - e^{-Ts})}\int_0^T f(t)e^{-st}\,dt. \tag{6.38.12}$$

11. *Hardy's theorem*: If $f(t) = \sum_{n=0}^\infty c_n t^n$ for $t \geq 0$ and $\sum_{n=0}^\infty \frac{c_n n!}{s_0^n}$ converges for some $s_0 > 0$, then $\left[\mathcal{L}(f)\right](s) = \sum_{n=0}^\infty \frac{c_n n!}{s^n}$ for Re $s > s_0$.

6.38.3 INVERSION FORMULAS

6.38.3.1 Inversion by integration

If $f(t)$ is a locally integrable function on $[0,\infty)$ such that

1. f is of bounded variation in a neighborhood of a point $t_0 \geq 0$ (a right-hand neighborhood if $t_0 = 0$),
2. The Laplace integral of f converges absolutely on the line Re $s = c$, then

$$\lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} F(s)e^{st_0}\,ds = \begin{cases} 0, & \text{if } t_0 < 0, \\ f(0+)/2 & \text{if } t_0 = 0, \\ \left[f(t_0+) + f(t_0-)\right]/2 & \text{if } t_0 > 0. \end{cases}$$

In particular, if f is differentiable on $(0,\infty)$ and satisfies the above conditions, then

$$\lim_{T\to\infty} \frac{1}{2\pi i} \int_{c-iT}^{c+iT} F(s)e^{st}\,ds = f(t), \quad 0 < t < \infty. \tag{6.38.13}$$

6.38.3.2 Inversion by partial fractions

Suppose that F is a rational function $F(s) = P(s)/Q(s)$ in which the degree of the denominator Q is greater than that of the numerator P. For instance, let F be represented in its most reduced form where P and Q have no common zeros, and assume that Q has only simple zeros at $a_1, \ldots, a_n$, then

$$f(t) = \mathcal{L}^{-1}\left(F(s)\right)(t) = \mathcal{L}^{-1}\left(\frac{P(s)}{Q(s)}\right)(t) = \sum_{k=1}^{n} \frac{P(a_k)}{Q'(a_k)} e^{a_k t}. \tag{6.38.14}$$

EXAMPLE If $P(s) = s-5$ and $Q(s) = s^2+6s+13$, then $a_1 = -3+2i$, $a_2 = -3-2i$, and it follows that

$$\begin{aligned} f(t) = \mathcal{L}^{-1}\left(\frac{s-5}{s^2+6s+13}\right) &= \frac{(2i-8)}{4i}e^{(-3+2i)t} + \frac{(2i+8)}{4i}e^{(-3-2i)t} \\ &= e^{-3t}(\cos 2t - 4\sin 2t). \end{aligned}$$

6.38.4 CONVOLUTION

Let $f(t)$ and $g(t)$ be locally integrable functions on $[0,\infty)$, and assume that their Laplace integrals converge absolutely in some half-plane Re $s > \alpha$. Then the convolution operation, $\star$, associated with the Laplace transform, is defined by

$$h(t) = (f \star g)(t) = \int_0^t f(x)g(t-x)\,dx. \tag{6.38.15}$$

The convolution of f and g is a locally integrable function on $[0,\infty)$ that is continuous if f or g is continuous. Additionally, it has a Laplace transform given by

$$H(s) = (\mathcal{L}h)(s) = F(s)G(s), \tag{6.38.16}$$

where $(\mathcal{L}f)(s) = F(s)$ and $(\mathcal{L}g)(s) = G(s)$.

6.39 MELLIN TRANSFORM

The *Mellin transform* of the real function $f(x)$ is

$$f^*(s) = \mathcal{M}[f(x); s] = \int_0^\infty f(x)x^{s-1}\, dx. \tag{6.39.1}$$

The inverse transform is

$$f(x) = \mathcal{M}^{-1}[f(s); x] = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}\, ds. \tag{6.39.2}$$

6.40 Z-TRANSFORM

The Z*-transform* of a sequence $\{f(n)\}_{-\infty}^{\infty}$ is defined by

$$\mathcal{Z}[f(n)] = F(z) = \sum_{n=-\infty}^{\infty} f(n)z^{-n}, \tag{6.40.1}$$

for all complex numbers z for which the series converges.

The series converges at least in a ring of the form $0 \leq r_1 < |z| < r_2 \leq \infty$, whose radii, r_1 and r_2, depend on the behavior of $f(n)$ at $\pm\infty$:

$$r_1 = \limsup_{n\to\infty} \sqrt[n]{|f(n)|}, \qquad r_2 = \liminf_{n\to\infty} \frac{1}{\sqrt[n]{|f(-n)|}}. \tag{6.40.2}$$

If there is more than one sequence involved, we may denote r_1 and r_2 by $r_1(f)$ and $r_2(f)$, respectively. It may happen that $r_1 > r_2$, so that the function is nowhere defined. The function $F(z)$ is analytic in this ring, but it may be possible to continue it analytically beyond the boundaries of the ring. If $f(n) = 0$ for $n < 0$, then $r_2 = \infty$, and if $f(n) = 0$ for $n \geq 0$, then $r_1 = 0$.

Let $z = re^{i\theta}$. Then the Z-transform evaluated at $r = 1$ is the Fourier transform of the sequence $\{f(n)\}_{-\infty}^{\infty}$,

$$\sum_{n=-\infty}^{\infty} f(n)e^{-in\theta}. \tag{6.40.3}$$

6.40.1 EXAMPLES

1. Let a be a complex number and define $f(n) = a^n$, for $n \geq 0$, and zero otherwise, then

$$\mathcal{Z}[f(n)] = \sum_{n=0}^{\infty} a^n z^{-n} = \frac{z}{z-a} \quad , \quad |z| > |a|. \tag{6.40.4}$$

Special case: step impulse function. If $a = 1$ then $\mathcal{Z}[u(n)] = \dfrac{z}{z-1}$ and $f(n) = u(n) = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0, \end{cases}$.

2. If $f(n) = na^n$, for $n \geq 0$, and zero otherwise, then

$$\mathcal{Z}[f(n)] = \sum_{n=0}^{\infty} na^n z^{-n} = \frac{az}{(z-a)^2} \quad , \quad |z| > |a|.$$

3. Let $\delta(n) = \begin{cases} 1, & n = 0, \\ 0, & \text{otherwise,} \end{cases}$ then $\mathcal{Z}[\delta(n-k)] = z^{-k}$ for $k = 0, \pm 1, \pm 2, \ldots$.

6.40.2 PROPERTIES

The region of convergence of the Z-transform of the sequence $\{f(n)\}$ will be denoted D_f.

1. *Linearity*:

   $\mathcal{Z}[af(n)+bg(n)] = a\mathcal{Z}[f(n)]+b\mathcal{Z}[g(n)] = aF(z)+bG(z), \quad z \in D_f \cap D_g.$

   The region $D_f \bigcap D_g$ contains the ring $r_1 < |z| < r_2$, where $r_1 = \text{maximum}\ \{r_1(f), r_1(g)\}$ and $r_2 = \text{minimum}\ \{r_2(f), r_2(g)\}$.
2. *Translation*: $\mathcal{Z}[f(n-k)] = z^{-k}F(z)$.
3. *Multiplication by exponentials*: $\mathcal{Z}[(a^n f(n))] = F(z/a)$ when $|a|r_1 < |z| < |a|r_2$.
4. *Multiplication by powers of* n: For $k = 0, 1, 2, \ldots$ and $z \in D_f$,

$$\mathcal{Z}[(n^k f(n))] = (-1)^k \left(z\frac{d}{dz}\right)^k F(z). \tag{6.40.5}$$

5. *Conjugation*: $\mathcal{Z}[\overline{f}(-n)] = \overline{F}\left(\dfrac{1}{\overline{z}}\right)$.
6. *Initial and final values*: If $f(n) = 0$ for $n < 0$, then $\lim_{z\to\infty} F(z) = f(0)$ and, conversely, if $F(z)$ is defined for $r_1 < |z|$ and for some integer m, $\lim_{z\to\infty} z^m F(z) = A$ (with $A \neq \pm\infty$), then $f(m) = A$ and $f(n) = 0$, for $n < m$.
7. *Parseval's relation*: Let $F, G \in L^2(-\pi, \pi)$, and let $F(z)$ and $G(z)$ be the Z-transforms of $\{f(n)\}$ and $\{g(n)\}$, respectively. Then

$$\sum_{n=-\infty}^{\infty} f(n)\overline{g}(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(e^{i\omega})\overline{G}(e^{i\omega})\, d\omega. \tag{6.40.6}$$

   In particular,

$$\sum_{n=-\infty}^{\infty} |f(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |F(e^{i\omega})|^2\, d\omega. \tag{6.40.7}$$

6.40.3 INVERSION FORMULA

Consider the sequences

$$f(n) = u(n) = \begin{cases} 1, & n \geq 0, \\ 0, & n < 0, \end{cases} \quad \text{and} \quad g(n) = -u(-n-1) = \begin{cases} -1, & n < 0, \\ 0, & n \geq 0. \end{cases}$$

Note that $F(z) = \dfrac{z}{z-1}$ for $|z| > 1$, and $G(z) = \dfrac{z}{z-1}$ for $|z| < 1$. Hence, the inverse Z-transform of the function $z/(z-1)$ is not unique. In general, the inverse Z-transform is not unique, unless its region of convergence is specified.

1. *Inversion by using series representation*:
   If $F(z)$ is given by its series

   $$F(z) = \sum_{n=-\infty}^{\infty} a_n z^{-n}, \quad r_1 < |z| < r_2,$$

   then its inverse Z-transform is unique and is given by $f(n) = a_n$ for all n.
2. *Inversion by using complex integration*:
   If $F(z)$ is given in a closed-form as an algebraic expression and its domain of analyticity is known, then its inverse Z-transform can be obtained by using the relationship

   $$f(n) = \frac{1}{2\pi i} \oint_{\gamma} F(z) z^{n-1} dz, \tag{6.40.8}$$

   where γ is a closed contour surrounding the origin once in the positive (counter-clockwise) direction in the domain of analyticity of $F(z)$.
3. *Inversion by using Fourier series*:
   If the domain of analyticity of F contains the unit circle, $|r| = 1$, and if F is single valued therein, then $F(e^{i\theta})$ is a periodic function with period 2π, and, consequently, it can be expanded in a Fourier series. The coefficients of the series form the inverse Z-transform of F and they are given explicitly by

   $$f(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} F(e^{i\theta}) e^{in\theta} d\theta. \tag{6.40.9}$$

   (This is a special case of (2) with $\gamma(\theta) = e^{i\theta}$.)
4. *Inversion by using partial fractions*:
   Dividing Equation (6.40.4) by z and differentiating both sides with respect to z, results in

   $$\mathcal{Z}^{-1}\left[(z-a)^{-k}\right] = \binom{n-1}{n-k} a^{n-k} u(n-k), \tag{6.40.10}$$

   for $k = 1, 2, \ldots$ and $|z| > |a| > 0$. Moreover,

   $$\mathcal{Z}^{-1}\left[z^{-k}\right] = \delta(n-k). \tag{6.40.11}$$

Let $F(z)$ be a rational function of the form

$$F(z) = \frac{P(z)}{Q(z)} = \frac{a_N z^N + \ldots + a_1 z + a_0}{b_M z^M + \ldots + b_1 z + b_0}.$$

with $a_N \neq 0$ and $b_M \neq 0$.

(a) Consider the case $N < M$. The denominator $Q(z)$ can be factored over the field of complex numbers as $Q(z) = c(z - z_1)^{k_1} \cdots (z - z_m)^{k_m}$, where c is a constant and $k_1, \ldots, k_m$ are positive integers satisfying $k_1 + \ldots + k_m = M$. Hence, F can be written in the form

$$F(z) = \sum_{i=1}^{m} \sum_{j=1}^{k_i} \frac{A_{i,j}}{(z - z_i)^j}, \tag{6.40.12}$$

where

$$A_{i,j} = \frac{1}{(k_i - j)!} \lim_{z \to z_i} \frac{d^{k_i - j}}{dz^{k_i - j}} \left((z - z_i)^{k_i} F(z) \right). \tag{6.40.13}$$

The inverse Z-transform of the fractional decomposition Equation (6.40.12) in the region that is exterior to the smallest circle containing all the zeros of $Q(z)$ can be obtained by using Equation (6.40.10).

(b) Consider the case $N \geq M$. We must divide until F can be reduced to the form

$$F(z) = H(z) + \frac{R(z)}{Q(z)},$$

where the remainder polynomial, $R(z)$, has degree less than or equal to $M - 1$, and the quotient, $H(z)$, is a polynomial of degree $N - M$. The inverse Z-transform of the quotient polynomial can be obtained by using Equation (6.40.11) and that of $R(z)/Q(z)$ can be obtained as in the case $N < M$.

EXAMPLE To find the inverse Z-transform of the function,

$$F(z) = \frac{z^4 + 5}{(z-1)^2(z-2)}, \qquad |z| > 2, \tag{6.40.14}$$

the partial fraction expansion,

$$\frac{z^4 + 5}{(z-1)^2(z-2)} = z + 4 - \frac{6}{(z-1)^2} - \frac{10}{(z-1)} + \frac{21}{(z-2)}, \tag{6.40.15}$$

is computed. With the aid of Equation (6.40.10) and Equation (6.40.11),

$$\begin{aligned} \mathcal{Z}^{-1}\left[F(z)\right] = \delta(n+1) + 4\delta(n) - 6(n-1)u(n-2) \\ - 10u(n-1) + 21 \cdot 2^{n-1} u(n-1), \end{aligned} \tag{6.40.16}$$

or $f(n) = -6n - 4 + 21 \cdot 2^{n-1}$, for $n \geq 2$, with the initial values $f(-1) = 1$, $f(0) = 4$, and $f(1) = 11$.

6.40.4 CONVOLUTION AND PRODUCT

The convolution of two sequences, $\{f(n)\}_{-\infty}^{\infty}$ and $\{g(n)\}_{-\infty}^{\infty}$, is a sequence $\{h(n)\}_{-\infty}^{\infty}$ defined by $h(n) = \sum_{k=-\infty}^{\infty} f(k)g(n-k)$. The Z-transform of the convolution of two sequences is the product of their Z-transforms,

$$\mathcal{Z}[h(n)] = \mathcal{Z}[f(n)]\mathcal{Z}[g(n)],$$

for $z \in D_f \bigcap D_g$, or $H(z) = F(z)G(z)$.

The Z-transform of the product of two sequences is given by

$$\mathcal{Z}[f(n)g(n)] = \frac{1}{2\pi i}\oint_{\gamma} F(\omega)G\left(\frac{z}{\omega}\right)\frac{d\omega}{\omega}, \tag{6.40.17}$$

where γ is a closed contour surrounding the origin in the positive direction in the domain of convergence of $F(\omega)$ and $G(z/\omega)$.

6.40.5 LINEAR CANONICAL TRANSFORMATION

The Linear Canonical Transformation (LCT) of the function $x(t)$, using the parameters $\{a, b, c, d\}$ with $ad - bc = 1$, is defined to be

$$\begin{aligned} X_{(a,b,c,d)}(u) &= \sqrt{-i}e^{\pi i d u^2/b}\int_{-\infty}^{\infty} \exp\left[\frac{\pi i}{b}(-2ut + at^2)\right] x(t)\,dt && \text{when } b \neq 0 \\ X_{(a,0,c,d)}(u) &= \sqrt{d}e^{c\pi i d u^2} x(du) && \text{when } b = 0 \end{aligned} \tag{6.40.18}$$

- Many common integral transforms can be written in terms of the LCT:
  1. Fourier Transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$
  2. Fractional Fourier Transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$
  3. Fresnel Transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 1 & \lambda z \\ 0 & 1 \end{bmatrix}$
  4. Laplace Transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix}$
  5. Fractional Laplace Transform $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} i\cos\theta & i\sin\theta \\ i\sin\theta & -i\cos\theta \end{bmatrix}$
- Composition of LCTs is equivalent to multiplying matrices. If the LCT operation is written as $X_{(a,b,c,d)}(u) = L_{(a,b,c,d)}[x(t)]$ then

$$L_{(a,b,c,d)} \circ L_{(A,B,C,D)} = L_{(\alpha,\beta,\gamma,\delta)} \tag{6.40.19}$$

  where

$$\begin{bmatrix} \alpha & \beta \\ \gamma & \delta \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}\begin{bmatrix} A & B \\ C & D \end{bmatrix} \tag{6.40.20}$$

6.41 TABLES OF TRANSFORMS

Finite sine transforms

$$f_s(n) = \int_0^\pi F(x)\,\sin nx\,dx,\ \text{for } n = 1, 2, \ldots.$$

No.	$f_s(n)$	$F(x)$
1	$(-1)^{n+1} f_s(n)$	$F(\pi - x)$
2	$1/n$	$\pi - x/\pi$
3	$(-1)^{n+1}/n$	x/π
4	$1 - (-1)^n/n$	1
5	$\frac{2}{n^2} \sin \frac{n\pi}{2}$	$\begin{cases} x & \text{when } 0 < x < \pi/2 \\ \pi - x & \text{when } \pi/2 < x < \pi \end{cases}$
6	$(-1)^{n+1}/n^3$	$x(\pi^2 - x^2)/6\pi$
7	$1 - (-1)^n/n^3$	$x(\pi - x)/2$
8	$\frac{\pi^2(-1)^{n-1}}{n} - \frac{2[1-(-1)^n]}{n^3}$	x^2
9	$\pi(-1)^n \left(\frac{6}{n^3} - \frac{\pi^2}{n}\right)$	x^3
10	$\frac{n}{n^2 + c^2} [1 - (-1)^n e^{c\pi}]$	e^{cx}
11	$\frac{n}{n^2 + c^2}$	$\frac{\sinh c(\pi - x)}{\sinh c\pi}$
12	$\frac{n}{n^2 - k^2}$ with $k \neq 0, 1, 2, \ldots$	$\frac{\sin k(\pi - x)}{\sin k\pi}$
13	$\begin{cases} \pi/2 & \text{when } n = m \\ 0 & \text{when } n \neq m,\ m = 1, 2, \ldots \end{cases}$	$\sin mx$
14	$\frac{n}{n^2 - k^2}[1 - (-1)^n \cos k\pi]$ with $k \neq 1, 2, \ldots$ (0 if $n = k$)	$\cos kx$
15	$\frac{n}{(n^2 - k^2)^2}$ with $k \neq 0, 1, 2, \ldots$	$\frac{\pi \sin kx}{2k \sin^2 k\pi} - \frac{x \cos k(\pi - x)}{2k \sin k\pi}$
16	$\frac{b^n}{n}$ with $\lvert b\rvert \leq 1$	$\frac{2}{\pi} \tan^{-1} \frac{b \sin x}{1 - b \cos x}$
17	$\frac{1 - (-1)^n}{n} b^n$ with $\lvert b\rvert \leq 1$	$\frac{2}{\pi} \tan^{-1} \frac{2b \sin x}{1 - b^2}$

Finite cosine transforms

$$f_c(n) = \int_0^\pi F(x)\,\cos nx\,dx, \text{ for } n = 0, 1, 2, \ldots.$$

No.	$f_c(n)$	$F(x)$
1	$(-1)^n f_c(n)$	$F(\pi - x)$
2	$\begin{cases} \pi & n = 0 \\ 0 & n = 1, 2, \ldots \end{cases}$	1
3	$\begin{cases} 0 & n = 0 \\ \dfrac{2}{n}\sin\dfrac{n\pi}{2} & n = 1, 2, \ldots \end{cases}$	$\begin{cases} 1 & \text{for } 0 < x < \pi/2 \\ -1 & \text{for } \pi/2 < x < \pi \end{cases}$
4	$\begin{cases} \dfrac{\pi^2}{2} & n = 0 \\ (-1)^n - 1/n^2 & n = 1, 2, \ldots \end{cases}$	x
5	$\begin{cases} \dfrac{\pi^2}{6} & n = 0 \\ (-1)^n/n^2 & n = 1, 2, \ldots \end{cases}$	$\dfrac{x^2}{2\pi}$
6	$\begin{cases} 0 & n = 0 \\ 1/n^2 & n = 1, 2, \ldots \end{cases}$	$\dfrac{(x-\pi)^2}{2\pi} - \dfrac{\pi}{6}$
7	$\begin{cases} \dfrac{\pi^4}{4} & n = 0 \\ 3\pi^2\dfrac{(-1)^n}{n^2} - 6\dfrac{1-(-1)^n}{n^4} & n = 1, 2, \ldots \end{cases}$	x^3
8	$\dfrac{(-1)^n e^c\pi - 1}{n^2 + c^2}$	$\frac{1}{c}e^{cx}$
9	$\dfrac{1}{n^2 + c^2}$	$\dfrac{\cosh c(\pi - x)}{c \sinh c\pi}$
10	$\dfrac{k}{n^2 - k^2}[(-1)^n \cos \pi k - 1]$ with $k \neq 0, 1, 2, \ldots$	$\sin kx$
11	$\begin{cases} 0 & m = 1, 2, \ldots \\ \dfrac{(-1)^{n+m} - 1}{n^2 - m^2} & m \neq 1, 2, \ldots \end{cases}$	$\dfrac{1}{m}\sin mx$
12	$\dfrac{1}{n^2 - k^2}$ with $k \neq 0, 1, 2, \ldots$	$-\dfrac{\cos k(\pi - x)}{k \sin k\pi}$
13	$\begin{cases} \pi/2 & \text{when } n = m \\ 0 & \text{when } n \neq m \end{cases}$	$\cos mx \qquad (m = 1, 2, \ldots)$

Fourier sine transforms

$$F(\omega) = F_s(f)(\omega) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x)\sin(\omega x)\,dx, \quad \omega > 0.$$

No.	$f(x)$	$F(\omega)$
1	$\begin{cases} 1 & 0 < x < a \\ 0 & x > a \end{cases}$	$\sqrt{\frac{2}{\pi}}\left(\frac{1-\cos\omega a}{\omega}\right)$
2	x^{p-1} $(0 < p < 1)$	$\sqrt{\frac{2}{\pi}}\frac{\Gamma(p)}{\omega^p}\sin\frac{p\pi}{2}$
3	$\begin{cases} \sin x & 0 < x < a \\ 0 & x > a \end{cases}$	$\frac{1}{\sqrt{2\pi}}\left(\frac{\sin[a(1-\omega)]}{1-\omega} - \frac{\sin[a(1+\omega)]}{1+\omega}\right)$
4	e^{-x}	$\sqrt{\frac{2}{\pi}}\frac{\omega}{1+\omega^2}$
5	$xe^{-x^2/2}$	$\omega e^{-\omega^2/2}$
6	$\cos\frac{x^2}{2}$	$\sqrt{2}\left[\sin\frac{\omega^2}{2}C\left(\frac{\omega^2}{2}\right) - \cos\frac{\omega^2}{2}S\left(\frac{\omega^2}{2}\right)\right]$
7	$\sin\frac{x^2}{2}$	$\sqrt{2}\left[\cos\frac{\omega^2}{2}C\left(\frac{\omega^2}{2}\right) + \sin\frac{\omega^2}{2}S\left(\frac{\omega^2}{2}\right)\right]$

Fourier cosine transforms

$$F(\omega) = F_c(f)(\omega) = \sqrt{\frac{2}{\pi}} \int_0^\infty f(x)\cos(\omega x)\,dx, \quad \omega > 0.$$

No.	$f(x)$	$F(\omega)$
1	$\begin{cases} 1 & 0 < x < a \\ 0 & x > a \end{cases}$	$\sqrt{\frac{2}{\pi}}\frac{\sin a\omega}{\omega}$
2	x^{p-1} $(0 < p < 1)$	$\sqrt{\frac{2}{\pi}}\frac{\Gamma(p)}{\omega^p}\cos\frac{p\pi}{2}$
3	$\begin{cases} \cos x & 0 < x < a \\ 0 & x > a \end{cases}$	$\frac{1}{\sqrt{2\pi}}\left(\frac{\sin[a(1-\omega)]}{1-\omega} + \frac{\sin[a(1+\omega)]}{1+\omega}\right)$
4	e^{-x}	$\sqrt{\frac{2}{\pi}}\frac{1}{1+\omega^2}$
5	$e^{-x^2/2}$	$e^{-\omega^2/2}$
6	$\cos\frac{x^2}{2}$	$\cos\left(\frac{\omega^2}{2} - \frac{\pi}{4}\right)$
7	$\sin\frac{x^2}{2}$	$\cos\left(\frac{\omega^2}{2} + \frac{\pi}{4}\right)$

Fourier transforms: functional relations

$$F(\omega) = \mathcal{F}(f)(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x}\, dx$$

No.	$f(x)$	$F(\omega)$
1	$ag(x) + bh(x)$	$aG(\omega) + bH(\omega)$
2	$f(ax)$ $\quad a \neq 0, \operatorname{Im} a = 0$	$\frac{1}{\lvert a\rvert} F\left(\frac{\omega}{a}\right)$
3	$f(-x)$	$F(-\omega)$
4	$\overline{f(x)}$	$\overline{F(-\omega)}$
5	$f(x-\tau)$ $\quad \operatorname{Im} \tau = 0$	$e^{i\omega\tau} F(\omega)$
6	$e^{i\Omega x} f(x)$ $\quad \operatorname{Im} \Omega = 0$	$F(\omega + \Omega)$
7	$F(x)$	$f(-\omega)$
8	$\frac{d^n}{dx^n} f(x)$	$(-i\omega)^n F(\omega)$
9	$(ix)^n f(x)$	$\frac{d^n}{d\omega^n} F(\omega)$
10	$\frac{\partial}{\partial a} f(x, a)$	$\frac{\partial}{\partial a} F(\omega, a)$

Fourier transforms

$$F(\omega) = \mathcal{F}(f)(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x}\, dx$$

No.	$f(x)$	$F(\omega)$
1	$\delta(x)$	$1/\sqrt{2\pi}$
2	$\delta(x-\tau)$	$e^{i\omega\tau}/\sqrt{2\pi}$
3	$\delta^{(n)}(x)$	$(-i\omega)^n/\sqrt{2\pi}$
4	$H(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 \end{cases}$	$-\frac{1}{i\omega\sqrt{2\pi}} + \sqrt{\frac{\pi}{2}}\delta(\omega)$
5	$\operatorname{sgn}(x) = \begin{cases} 1 & x > 0 \\ -1 & x < 0 \end{cases}$	$-\sqrt{\frac{2}{\pi}}\frac{1}{i\omega}$
6	$\begin{cases} 1 & \lvert x\rvert < a \\ -1 & \lvert x\rvert > a \end{cases}$	$\sqrt{\frac{2}{\pi}}\frac{\sin a\omega}{\omega}$
7	$\begin{cases} e^{i\Omega t} & \lvert x\rvert < a \\ 0 & \lvert x\rvert > a \end{cases}$	$\sqrt{\frac{2}{\pi}}\frac{\sin a(\Omega + \omega)}{\Omega + \omega}$

Fourier transforms

$$F(\omega) = \mathcal{F}(f)(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f(x) e^{i\omega x}\, dx$$

No.	$f(x)$	$F(\omega)$
8	$e^{-a\|x\|}$ $\quad a > 0$	$-\sqrt{\frac{2}{\pi}} \frac{a}{a^2+\omega^2}$
9	$\frac{\sin \Omega x}{x}$	$\sqrt{\frac{\pi}{2}} [H(\Omega - \omega) - H(-\Omega - \omega)]$
10	$\sin ax/x$	$\begin{cases} \sqrt{\frac{\pi}{2}} & \|\omega\| < a \\ 0 & \|\omega\| > a \end{cases}$
11	$\begin{cases} e^{iax} & p < x < q \\ 0 & x < p, x > q \end{cases}$	$\frac{i}{\sqrt{2\pi}} e^{ip(\omega+a)} - e^{iq(\omega+a)}/\omega + a$
12	$\begin{cases} e^{-cx+iax} & x > 0 \\ 0 & x < 0 \end{cases} \quad (c > 0)$	$\frac{i}{\sqrt{2\pi}(\omega + a + ic)}$
13	e^{-px^2} $\quad \mathrm{Re}\, p > 0$	$\frac{1}{\sqrt{2p}} e^{-\omega^2/4p}$
14	$\cos px^2$	$\frac{1}{\sqrt{2p}} \cos\left(\frac{\omega^2}{4p} - \frac{\pi}{4}\right)$
15	$\sin px^2$	$\frac{1}{\sqrt{2p}} \cos\left(\frac{\omega^2}{4p} + \frac{\pi}{4}\right)$
16	$\|x\|^{-p}$ $\quad (0 < p < 1)$	$\sqrt{\frac{2}{\pi}} \frac{\Gamma(1-p) \sin \frac{p\pi}{2}}{\|\omega\|^{1-p}}$
17	$e^{-a\|x\|}/\sqrt{\|x\|}$	$\frac{\sqrt{\sqrt{a^2+\omega^2}+a}}{\sqrt{\omega^2+a^2}}$
18	$\frac{\cosh ax}{\cosh \pi x}$ $\quad (-\pi < a < \pi)$	$\sqrt{\frac{2}{\pi}} \frac{\cos \frac{a}{2} \cosh \frac{\omega}{2}}{\cos a + \cosh \omega}$
19	$\frac{\sinh ax}{\sinh \pi x}$ $\quad (-\pi < a < \pi)$	$\frac{1}{\sqrt{2\pi}} \frac{\sin a}{\cos a + \cosh \omega}$
20	$\begin{cases} \frac{1}{\sqrt{a^2-x^2}} & \|x\| < a \\ 0 & \|x\| > a \end{cases}$	$\sqrt{\frac{\pi}{2}} J_0(a\omega)$
21	$\frac{\sin[b\sqrt{a^2+x^2}]}{\sqrt{a^2+x^2}}$	$\begin{cases} 0 & \|\omega\| > b \\ \sqrt{\frac{\pi}{2}} J_0(a\sqrt{b^2 - \omega^2}) & \|\omega\| < b \end{cases}$
22	$\begin{cases} P_n(x) & \|x\| < 1 \\ 0 & \|x\| > 1 \end{cases}$	$\frac{i^n}{\sqrt{\omega}} J_{n+1/2}(\omega)$
23	$\begin{cases} \frac{\cos[b\sqrt{a^2-x^2}]}{\sqrt{a^2-x^2}} & \|x\| < a \\ 0 & \|x\| > a \end{cases}$	$\sqrt{\frac{\pi}{2}} J_0(a\sqrt{\omega^2 + b^2})$
24	$\begin{cases} \frac{\cosh[b\sqrt{a^2-x^2}]}{\sqrt{a^2-x^2}} & \|x\| < a \\ 0 & \|x\| > a \end{cases}$	$\sqrt{\frac{\pi}{2}} J_0(a\sqrt{\omega^2 - b^2})$

Multidimensional Fourier transforms

$$F(\mathbf{u}) = (2\pi)^{-n/2} \underset{\mathbb{R}^n}{\int \cdots \int} f(\mathbf{x}) e^{i(\mathbf{x}\cdot\mathbf{u})}\, d\mathbf{x}$$

No.	$f(\mathbf{x})$	$F(\mathbf{u})$
	In n-dimensions	
1	$f(a\mathbf{x})$ Im $a = 0$	$\|a\|^{-n} F(a^{-1}\mathbf{u})$
2	$f(\mathbf{x}-\mathbf{a})$	$e^{-i\mathbf{a}\cdot\mathbf{u}} F(\mathbf{u})$
3	$e^{i\mathbf{a}\cdot\mathbf{x}} f(\mathbf{x})$	$F(\mathbf{u}+\mathbf{a})$
4	$F(\mathbf{x})$	$(2\pi)^n f(-\mathbf{u})$
	Two dimensions: let $\mathbf{x} = (x, y)$ and $\mathbf{u} = (u, v)$.	
5	$f(ax, by)$	$\frac{1}{\|ab\|} F\left(\frac{u}{a}, \frac{v}{b}\right)$
6	$f(x-a, y-b)$	$e^{i(au+bv)} F(u, v)$
7	$e^{i(ax+by)} f(x, y)$	$F(u+a, v+b)$
8	$F(x, y)$	$(2\pi)^2 F(-u, -v)$
9	$\delta(x-a)\delta(y-b)$	$\frac{1}{2\pi} e^{-i(au+bv)}$
10	$e^{-x^2/4a-y^2/4b}$ $a, b > 0$	$2\sqrt{ab}\, e^{-au^2-bv^2}$
11	$\begin{cases} 1 & \|x\| < a, \|y\| < b \\ 0 & \text{otherwise} \end{cases}$ (rectangle)	$\frac{2 \sin au \sin bv}{\pi uv}$
12	$\begin{cases} 1 & \|x\| < a \\ 0 & \text{otherwise} \end{cases}$ (strip)	$\frac{2 \sin au}{\pi uv} \delta(v)$
13	$\begin{cases} 1 & x^2+y^2 < a^2 \\ 0 & \text{otherwise} \end{cases}$ (circle)	$\frac{a J_1(a\sqrt{u^2+v^2})}{\sqrt{u^2+v^2}}$
	Three dimensions: let $\mathbf{x} = (x, y, z)$ and $\mathbf{u} = (u, v, w)$.	
14	$\delta(x-a)\delta(y-b)\delta(z-c)$	$\frac{1}{(2\pi)^{3/2}} e^{-i(au+bv+cw)}$
15	$e^{-x^2/4a-y^2/4b-z^2/4c}$ $a, b, c > 0$	$2^{3/2}\sqrt{abc}\, e^{-au^2-bv^2-cw^2}$
16	$\begin{cases} 1 & \|x\| < a, \|y\| < b, \|z\| < c \\ 0 & \text{otherwise} \end{cases}$ (box)	$\left(\frac{2}{\pi}\right)^{3/2} \frac{\sin au \sin bv \sin cw}{uvw}$
17	$\begin{cases} 1 & x^2+y^2+z^2 < a^2 \\ 0 & \text{otherwise} \end{cases}$ (ball)	$\frac{\sin a\rho - a\rho \cos a\rho}{\sqrt{2\pi}\rho^3}$ $\rho^2 = u^2+v^2+w^2$

Hankel transforms

$$\mathcal{H}_\nu(f)(y) = F_\nu(y) = \int_0^\infty f(x)\sqrt{xy}J_\nu(yx)\,dx, \quad y > 0.$$

No.	$f(x)$	$F_\nu(y)$
1	$\begin{cases} x^{\nu+1/2}, & 0 < x < 1 \\ 0, & 1 < x \end{cases}$ Re $\nu > -1$	$y^{-1/2}J_{\nu+1}(y)$
2	$\begin{cases} x^{\nu+1/2}(a^2-x^2)^\mu, & 0 < x < a \\ 0, & a < x \end{cases}$ Re ν, Re $\mu > -1$	$2^\mu\Gamma(\mu+1)a^{\nu+\mu+1}y^{-\mu-1/2}$ $\times J_{\nu+\mu+1}(ay)$
3	$x^{\nu+1/2}(x^2+a^2)^{-\nu-1/2}$, Re $a > 0$, Re $\nu > -1/2$	$\sqrt{\pi}y^{\nu-1/2}2^{-\nu}e^{-ay}$ $\times [\Gamma(\nu+1/2)]^{-1}$
4	$x^{\nu+1/2}e^{-ax}$, Re $a > 0$, Re $\nu > -1$	$a(\pi)^{-1/2}2^{\nu+1}y^{\nu+1/2}\Gamma(\nu+3/2)$ $\times (a^2+y^2)^{-\nu-3/2}$
5	$x^{\nu+1/2}e^{-ax^2}$, Re $a > 0$, Re $\nu > -1$	$y^{\nu+1/2}(2a)^{-\nu-1}\exp\left(-y^2/4a\right)$
6	$e^{-ax}/\sqrt{x}$, Re $a > 0$, Re $\nu > -1$	$y^{-\nu+1/2}\left[\sqrt{(a^2+y^2)}-a\right]^\nu$ $\times (a^2+y^2)^{-1/2}$
7	$x^{-\nu-1/2}\cos(ax)$, $a > 0$, Re $\nu > -1/2$	$\sqrt{\pi}2^{-\nu}y^{-\nu+1/2}[\Gamma(\nu+1/2)]^{-1}$ $\times (y^2-a^2)^{\nu-1/2}H(y-a)$
8	$x^{1/2-\nu}\sin(ax)$, $a > 0$, Re $\nu > 1/2$	$a2^{1-\nu}\sqrt{\pi}y^{\nu+1/2}[\Gamma(\nu-1/2)]^{-1}$ $\times (y^2-a^2)^{\nu-3/2}H(y-a)$
9	$x^{-1/2}J_{\nu-1}(ax)$, $a > 0$, Re $\nu > -1$	$a^{\nu-1}y^{-\nu+1/2}H(y-a)$
10	$x^{-1/2}J_{\nu+1}(ax)$, $a > 0$, Re $\nu > -3/2$	$\begin{cases} a^{-\nu-1}y^{\nu+1/2}, & 0 < y < a \\ 0, & a < y \end{cases}$

Hilbert transforms

$$\mathcal{H}(f)(y) = F(y) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{f(x)}{x-y}\,dx.$$

No.	$f(x)$	$F(y)$
1	1	0
2	$\begin{cases} 0, & -\infty < x < a \\ 1, & a < x < b \\ 0, & b < x < \infty \end{cases}$	$\frac{1}{\pi}\log\left\|(b-y)(a-y)^{-1}\right\|$
3	$\begin{cases} 0, & -\infty < x < a \\ x^{-1}, & a < x < \infty \end{cases}$	$(\pi y)^{-1}\log\left\|a(a-y)^{-1}\right\|,\quad 0 \neq y \neq a,\quad a > 0$
4	$(x+a)^{-1}\qquad \operatorname{Im} a > 0$	$i(y+a)^{-1}$
5	$\frac{1}{1+x^2}$	$-\frac{y}{1+y^2}$
6	$\frac{1}{1+x^4}$	$-\frac{y(1+y^2)}{\sqrt{2}(1+y^4)}$
7	$\sin(ax),\qquad a > 0$	$\cos(ay)$
8	$\frac{\sin(ax)}{x},\qquad a > 0$	$\frac{\cos(ay)-1}{y}$
9	$\frac{\sin x}{1+x^2}$	$\frac{\cos y - e^{-1}}{1+y^2}$
10	$\cos(ax),\qquad a > 0$	$-\sin(ay)$
11	$\frac{1-\cos(ax)}{x},\qquad a > 0$	$\frac{\sin(ay)}{y}$
12	$\operatorname{sgn}(x)\sin(a\|x\|^{1/2})\qquad a > 0$	$\cos(a\|y\|^{1/2}) + \exp(-a\|y\|^{1/2})$
13	$e^{iax}\qquad a > 0$	ie^{iay}

Laplace transforms: functional relations

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
1	$af(t) + bg(t)$	$aF(s) + bG(s)$
2	$f'(t)$	$sF(s) - F(0+)$
3	$f''(t)$	$s^2F(s) - sF(0+) - F'(0+)$
4	$f^{(n)}(t)$	$s^nF(s) - \sum_{k=0}^{n-1} s^{n-1-k}F^{(k)}(0+)$
5	$\int_0^t f(\tau)\,d\tau$	$\frac{1}{s}F(s)$
6	$\int_0^t \int_0^\tau f(u)\,du\,d\tau$	$\frac{1}{s^2}F(s)$
7	$\int_0^t f_1(t-\tau)f_2(\tau)\,d\tau = f_1 * f_2$	$F_1(s)F_2(s)$
8	$tf(t)$	$-F'(s)$
9	$t^nf(t)$	$(-1)^nF^{(n)}(s)$
10	$\frac{1}{t}f(t)$	$\int_s^\infty F(z)\,dz$
11	$e^{at}f(t)$	$F(s-a)$
12	$f(t-b)$ with $f(t) = 0$ for $t < 0$	$e^{-bs}F(s)$
13	$\frac{1}{c}f\left(\frac{t}{c}\right)$	$F(cs)$
14	$\frac{1}{c}e^{bt/c}f\left(\frac{t}{c}\right)$	$F(cs-b)$
15	$f(t+a) = f(t)$	$\int_0^a e^{-st}f(t)\,dt/1 - e^{-as}$
16	$f(t+a) = -f(t)$	$\int_0^a e^{-st}f(t)\,dt/1 + e^{-as}$
17	$\sum_{k=1}^n \frac{p(a_k)}{q'(a_k)}e^{a_kt}$ with $q(t) = (t-a_1)\cdots(t-a_n)$	$\frac{p(s)}{q(s)}$
18	$e^{at}\sum_{k=1}^n \frac{\phi^{(n-k)}(a)}{(n-k)!}\frac{t^{k-1}}{(k-1)!}$	$\frac{\phi(s)}{(s-a)^n}$

Laplace transforms

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
1	$\delta(t)$, delta function	1
2	$H(t)$, unit step function or Heaviside function	$1/s$
3	t	$1/s^2$
4	$\frac{t^{n-1}}{(n-1)!}$	$1/s^n$ $(n = 1, 2, \dots)$
5	$1/\sqrt{\pi t}$	$1/\sqrt{s}$
6	$2\sqrt{t/\pi}$	$s^{-3/2}$
7	$\frac{2^n t^{n-1/2}}{\sqrt{\pi}(2n-1)!!}$	$s^{-(n+1/2)}$ $(n = 1, 2, \dots)$
8	t^{k-1}	$\frac{\Gamma(k)}{s^k}$ $(k > 0)$
9	e^{at}	$\frac{1}{s-a}$
10	te^{at}	$\frac{1}{(s-a)^2}$
11	$\frac{1}{(n-1)!} t^{n-1} e^{at}$	$\frac{1}{(s-a)^n}$ $(n = 1, 2, \dots)$
12	$t^{k-1} e^{at}$	$\frac{\Gamma(k)}{(s-a)^k}$ $(k > 0)$
13	$\frac{1}{a-b}\left(e^{at} - e^{bt}\right)$	$\frac{1}{(s-a)(s-b)}$ $(a \neq b)$
14	$\frac{1}{a-b}\left(ae^{at} - be^{bt}\right)$	$\frac{s}{(s-a)(s-b)}$ $(a \neq b)$
15	$-\frac{(b-c)e^{at}+(c-a)e^{bt}+(a-b)e^{ct}}{(a-b)(b-c)(c-a)}$	$\frac{1}{(s-a)(s-b)(s-c)}$ $(a, b, c$ distinct$)$
16	$\frac{1}{a}\sin at$	$\frac{1}{s^2+a^2}$
17	$\cos at$	$\frac{s}{s^2+a^2}$
18	$\frac{1}{a}\sinh at$	$\frac{1}{s^2-a^2}$
19	$\cosh at$	$\frac{s}{s^2-a^2}$
20	$\frac{1}{a^2}(1 - \cos at)$	$\frac{1}{s(s^2+a^2)}$
21	$\frac{1}{a^3}(at - \sin at)$	$\frac{1}{s^2(s^2+a^2)}$
22	$\frac{1}{2a^3}(\sin at - at\cos at)$	$\frac{1}{(s^2+a^2)^2}$
23	$\frac{t}{2a}\sin at$	$\frac{s}{(s^2+a^2)^2}$

Laplace transforms

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
24	$\frac{1}{2a}(\sin at + at\cos at)$	$\frac{s^2}{(s^2+a^2)^2}$
25	$t\cos at$	$\frac{s^2-a^2}{(s^2+a^2)^2}$
26	$\frac{\cos at-\cos bt}{b^2-a^2}$	$\frac{s}{(s^2+a^2)(s^2+b^2)}$ $(a^2 \neq b^2)$
27	$\frac{1}{b}e^{at}\sin bt$	$\frac{1}{(s-a)^2+b^2}$
28	$e^{at}\cos bt$	$\frac{s-a}{(s-a)^2+b^2}$
29	$-\frac{e^{-at}}{4^{n-1}b^{2n}}\sum_{k=1}^{n}\binom{2n-k-1}{n-1} \times(-2t)^{k-1}\frac{d^k}{dt^k}[\cos bt]$	$\frac{1}{[(s-a)^2+b^2]^n}$
30	$\frac{e^{-at}}{4^{n-1}b^{2n}}\Big\{\sum_{k=1}^{n}\binom{2n-k-1}{n-1}\frac{(-2t)^{k-1}}{(k-1)!} \times\frac{d^k}{dt^k}[a\cos bt + b\sin at] -2b\sum_{k=1}^{n-1}\binom{2n-k-2}{n-1}\frac{(-2t)^{k-1}}{(k-1)!} \times\frac{d^k}{dt^k}[\sin bt]\Big\}$	$\frac{s}{[(s-a)^2+b^2]^n}$
31	$e^{-at} -e^{at/2}\left(\cos\frac{at\sqrt{3}}{2} - \sqrt{3}\sin\frac{at\sqrt{3}}{2}\right)$	$\frac{3a^2}{s^3+a^3}$
32	$\sin at\cosh at - \cos at\sinh at$	$\frac{4a^3}{s^4+4a^4}$
33	$\frac{1}{2a^2}\sin at\sinh at$	$\frac{s}{s^4+4a^4}$
34	$\frac{1}{2a^3}(\sinh at - \sin at)$	$\frac{1}{s^4-a^4}$
35	$\frac{1}{2a^2}(\cosh at - \cos at)$	$\frac{s}{s^4-a^4}$
36	$(1+a^2t^2)\sin at - \cos at$	$\frac{8a^3s^2}{(s^2+a^2)^3}$
37	$\frac{e^t}{n!}\frac{d^n}{dt^n}(t^ne^{-t})$	$\frac{1}{s}\left(\frac{s-1}{s}\right)^n$
38	$\frac{1}{\sqrt{\pi t}}e^{at}(1+2at)$	$\frac{s}{(s-a)^{3/2}}$
39	$\frac{1}{2\sqrt{\pi t^3}}(e^{bt}-e^{at})$	$\sqrt{s-a}-\sqrt{s-b}$
40	$\frac{1}{\sqrt{\pi t}} - ae^{a^2t}\operatorname{erfc}(a\sqrt{t})$	$\frac{1}{\sqrt{s}+a}$
41	$\frac{1}{\sqrt{\pi t}} + ae^{a^2t}\operatorname{erf}(a\sqrt{t})$	$\frac{\sqrt{s}}{s-a^2}$

Laplace transforms

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
42	$\frac{1}{\sqrt{\pi t}} - \frac{2a}{\sqrt{\pi}}e^{-a^2t}\int_0^{a\sqrt{t}} e^{\tau^2}\,d\tau$	$\frac{\sqrt{s}}{s+a^2}$
43	$\frac{1}{a}e^{a^2t}\,\mathrm{erf}(a\sqrt{t})$	$\frac{1}{\sqrt{s}(s-a^2)}$
44	$\frac{2}{a\sqrt{\pi}}e^{-a^2t}\int_0^{a\sqrt{t}} e^{\tau^2}\,d\tau$	$\frac{1}{\sqrt{s}(s+a^2)}$
45	$e^{a^2t}[b - a\,\mathrm{erf}(a\sqrt{t})]$ $-be^{b^2t}\,\mathrm{erfc}(b\sqrt{t})$	$\dfrac{b^2-a^2}{(s-a^2)(b+\sqrt{s})}$
46	$e^{a^2t}\,\mathrm{erfc}(a\sqrt{t})$	$\frac{1}{\sqrt{s}(\sqrt{s}+a)}$
47	$\frac{1}{\sqrt{b-a}}e^{-at}\,\mathrm{erf}(\sqrt{b-a}\sqrt{t})$	$\frac{1}{(s+a)\sqrt{s+b}}$
48	$e^{a^2t}\left[\frac{b}{a}\,\mathrm{erf}(a\sqrt{t}) - 1\right]$ $+e^{b^2t}\,\mathrm{erfc}(b\sqrt{t})$	$\frac{b^2-a^2}{\sqrt{s}(s-a^2)(\sqrt{s}+b)}$
49	$\frac{n!}{(2n)!\sqrt{\pi t}}H_{2n}(\sqrt{t})$	$\frac{(1-s)^n}{s^{n+1/2}}$
50	$-\frac{n!}{(2n+1)!\sqrt{\pi}}H_{2n+1}(\sqrt{t})$	$\frac{(1-s)^n}{s^{n+3/2}}$
51	$ae^{-at}\left[I_1(at) + I_0(at)\right]$	$\frac{\sqrt{s+2a}}{\sqrt{s}} - 1$
52	$e^{-(a+b)t/2}I_0\left(\frac{a-b}{2}t\right)$	$\frac{1}{\sqrt{s+a}\sqrt{s+b}}$
53	$\sqrt{\pi}\left(\frac{t}{a-b}\right)^{k-1/2} e^{-(a+b)t/2}$ $\times I_{k-1/2}\left(\frac{a-b}{2}t\right)$	$\dfrac{\Gamma(k)}{(s+a)^k(s+b)^k}$ $(k \geq 0)$
54	$te^{-(a+b)t/2}\left[I_0\left(\frac{a-b}{2}t\right)\right.$ $\left.+ I_1\left(\frac{a-b}{2}t\right)\right]$	$\dfrac{1}{\sqrt{s+a}(s+b)^{3/2}}$
55	$\frac{1}{t}e^{-at}I_1(at)$	$\frac{\sqrt{s+2a}-\sqrt{s}}{\sqrt{s+2a}+\sqrt{s}}$
56	$J_0(at)$	$\frac{1}{\sqrt{s^2+a^2}}$
57	$a^kJ_k(at)$	$\frac{(\sqrt{s^2+a^2}-s)^k}{\sqrt{s^2+a^2}}$ $(k > -1)$
58	$\frac{\sqrt{\pi}}{\Gamma(k)}\left(\frac{t}{2a}\right)^{k-1/2} J_{k-1/2}(at)$	$\frac{1}{(s^2+a^2)^k}$ $(k > 0)$
59	$\frac{ka^k}{t}J_k(at)$	$(\sqrt{s^2+a^2} - s)^k$ $(k > 0)$
60	$a^kI_k(at)$	$\frac{(s-\sqrt{s^2-a^2})^k}{\sqrt{s^2-a^2}}$ $(k > -1)$
61	$\frac{\sqrt{\pi}}{\Gamma(k)}\left(\frac{t}{2a}\right)^{k-1/2} I_{k-1/2}(at)$	$\frac{1}{(s^2-a^2)^k}$ $(k > 0)$

Laplace transforms

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
62	$\begin{cases} 0 & \text{when } 0 < t < k \\ 1 & \text{when } t > k \end{cases}$	$\dfrac{e^{-ks}}{s}$
63	$\begin{cases} 0 & \text{when } 0 < t < k \\ t-k & \text{when } t > k \end{cases}$	$\dfrac{e^{-ks}}{s^2}$
64	$\begin{cases} 0 & \text{when } 0 < t < k \\ \frac{(t-k)^{p-1}}{\Gamma(p)} & \text{when } t > k \end{cases}$	$\dfrac{e^{-ks}}{s^p}$ $(p > 0)$
65	$\begin{cases} 1 & \text{when } 0 < t < k \\ 0 & \text{when } t > k \end{cases}$	$\dfrac{1-e^{-ks}}{s}$
66	$\lvert\sin at\rvert$	$\frac{a}{s^2+a^2}\coth\frac{\pi s}{2a}$
67	$J_0(2\sqrt{at})$	$\frac{1}{s}e^{-a/s}$
68	$\frac{1}{\sqrt{\pi t}}\cos 2\sqrt{at}$	$\frac{1}{\sqrt{s}}e^{-a/s}$
69	$\frac{1}{\sqrt{\pi t}}\cosh 2\sqrt{at}$	$\frac{1}{\sqrt{s}}e^{a/s}$
70	$\frac{1}{\sqrt{\pi a}}\sin 2\sqrt{at}$	$\frac{1}{s^{3/2}}e^{-a/s}$
71	$\frac{1}{\sqrt{\pi a}}\sinh 2\sqrt{at}$	$\frac{1}{s^{3/2}}e^{a/s}$
72	$\left(\frac{t}{a}\right)^{(k-1)/2} J_{k-1}(2\sqrt{at})$	$\frac{1}{s^k}e^{-a/s}$ $(k > 0)$
73	$\left(\frac{t}{a}\right)^{(k-1)/2} I_{k-1}(2\sqrt{at})$	$\frac{1}{s^k}e^{a/s}$ $(k > 0)$
74	$\frac{a}{2\sqrt{\pi t^3}}e^{-a^2/4t}$	$e^{-a\sqrt{s}}$ $(a > 0)$
75	$\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}\right)$	$\frac{1}{s}e^{-a\sqrt{s}}$ $(a \geq 0)$
76	$\frac{1}{\sqrt{\pi t}}e^{-a^2/4t}$	$\frac{1}{\sqrt{s}}e^{-a\sqrt{s}}$ $(a \geq 0)$
77	$2\sqrt{\frac{t}{\pi}}e^{-a^2/4t} - a\operatorname{erfc}\left(\frac{a}{2\sqrt{t}}\right)$	$s^{-3/2}e^{-a\sqrt{s}}$ $(a \geq 0)$
78	$e^{ak+a^2t}\operatorname{erfc}\left(a\sqrt{t}+\frac{k}{2\sqrt{t}}\right)$	$\frac{e^{-k\sqrt{s}}}{\sqrt{s}(a+\sqrt{s})}$ $(k \geq 0)$
79	$J_0(a\sqrt{t^2+2kt})$	$\frac{e^{-k(\sqrt{s^2+a^2}-s)}}{\sqrt{s^2+a^2}}$ $(k \geq 0)$
80	$\Gamma'(1) - \log t$	$\frac{1}{s}\log s$
81	$t^{k-1}\left[\frac{\Gamma'(k)}{\lvert\Gamma(k)\rvert^2} - \frac{\log t}{\Gamma(k)}\right]$	$\frac{1}{s^k}\log s$ $(k > 0)$
82	$e^{at}[\log a - \operatorname{Ei}(-at)]$	$\frac{\log s}{s-a}$ $(a > 0)$

Laplace transforms

$$F(s) = \mathcal{L}(f)(s) = \int_0^\infty f(t)e^{-st}\,dt.$$

No.	$f(t)$	$F(s)$
83	$\cos t \operatorname{Si}(t) - \sin t \operatorname{Ci}(t)$	$\frac{\log s}{s^2+1}$
84	$-\sin t \operatorname{Si}(t) - \cos t \operatorname{Ci}(t)$	$\frac{t \log s}{s^2+1}$
85	$-\operatorname{Ei}\left(-\frac{t}{a}\right)$	$\frac{1}{s}\log(1+as)$ $(a>0)$
86	$\frac{1}{t}(e^{bt} - e^{at})$	$\log\frac{s-a}{s-b}$
87	$-2\operatorname{Ci}\left(-\frac{t}{a}\right)$	$\frac{1}{s}\log(1+a^2s^2)$
88	$2\log a - 2\operatorname{Ci}(at)$	$\frac{1}{s}\log(s^2+a^2)$ $(a>0)$
89	$\frac{2}{a}[at\log a + \sin at - at\operatorname{Ci}(at)]$	$\frac{1}{s^2}\log(s^2+a^2)$ $(a>0)$
90	$\frac{2}{t}(1-\cos at)$	$\log\frac{s^2+a^2}{s^2}$
91	$\frac{2}{t}(1-\cosh at)$	$\log\frac{s^2-a^2}{s^2}$
92	$\frac{1}{t}\sin at$	$\tan^{-1}\frac{a}{s}$
93	$\frac{1}{a\sqrt{\pi}}e^{-t^2/4a^2}$	$e^{a^2s^2}\operatorname{erfc}(as)$ $(a>0)$
94	$\operatorname{erf}\left(\frac{t}{2a}\right)$	$\frac{1}{s}e^{a^2s^2}\operatorname{erfc}(as)$ $(a>0)$
95	$\frac{\sqrt{a}}{\pi\sqrt{t}(t+a)}$	$e^{as}\operatorname{erfc}(\sqrt{as})$ $(a>0)$
96	$\frac{1}{\sqrt{\pi(t+a)}}$	$\frac{1}{\sqrt{s}}e^{as}\operatorname{erfc}(\sqrt{as})$ $(a>0)$
97	$\frac{1}{\pi t}\sin(2a\sqrt{t})$	$\operatorname{erf}\left(\frac{a}{\sqrt{s}}\right)$
98	$\frac{1}{t+a}$	$-e^{as}\operatorname{Ei}(-as)$ $(a>0)$
99	$\frac{1}{(t+a)^2}$	$\frac{1}{a} + se^{as}\operatorname{Ei}(-as)$ $(a>0)$
100	$\frac{1}{t^2+1}$	$\left[\frac{\pi}{2} - \operatorname{Si}(s)\right]\cos s + \operatorname{Ci}(s)\sin s$
101	$\begin{cases} 0 & \text{when } 0<t<a \\ (t^2-a^2)^{-1/2} & \text{when } t>a \end{cases}$	$K_0(as)$

Mellin transforms

$$f^*(s) = \mathcal{M}[f(x); s] = \int_0^\infty f(x) x^{s-1}\, dx.$$

No.	$f(x)$	$f^*(s)$
1	$ag(x) + bh(x)$	$ag^*(s) + bh^*(s)$
2	$f^{(n)}(x)$†	$(-1)^n \frac{\Gamma(s)}{\Gamma(s-n)} f^*(s-n)$
3	$x^n f^{(n)}(x)$†	$(-1)^n \frac{\Gamma(s+n)}{\Gamma(s)} f^*(s)$
4	$I_n f(x)$‡	$(-1)^n \frac{\Gamma(s)}{\Gamma(s+n)} f^*(s+n)$
5	e^{-x}	$\Gamma(s)$ Re $s > 0$
6	e^{-x^2}	$\frac{1}{2}\Gamma(\frac{1}{2}s)$ Re $s > 0$
7	$\cos x$	$\Gamma(s)\cos(\frac{1}{2}\pi s)$ $0 <$ Re $s < 1$
8	$\sin x$	$\Gamma(s)\sin(\frac{1}{2}\pi s)$ $0 <$ Re $s < 1$
9	$(1-x)^{-1}$	$\pi\cot(\pi s)$ $0 <$ Re $s < 1$
10	$(1+x)^{-1}$	$\pi\,\mathrm{cosec}(\pi s)$ $0 <$ Re $s < 1$
11	$(1+x^a)^{-b}$	$\frac{\Gamma(s/a)\Gamma(b-s/a)}{a\Gamma(b)}$ $0 <$ Re $s < ab$
12	$\log(1+ax)$ $\lvert\arg a\rvert < \pi$	$\pi s^{-1} a^{-s}\,\mathrm{cosec}(\pi s)$ $-1 <$ Re $s < 0$
13	$\tan^{-1} x$	$-\frac{1}{2}\pi s^{-1}\sec(\frac{1}{2}\pi s)$ $-1 <$ Re $s < 0$
14	$\cot^{-1} x$	$\frac{1}{2}\pi s^{-1}\sec(\frac{1}{2}\pi s)$ $0 <$ Re $s < 1$
15	$\mathrm{csch}\, ax$ Re $a > 0$	$2(1-2^{-s})a^{-s}\Gamma(s)\zeta(s)$ Re $s > 1$
16	$\mathrm{sech}^2\, ax$ Re $a > 0$	$4(2a)^{-s}\Gamma(s)\zeta(s-1)$ Re $s > 2$
17	$\mathrm{csch}^2\, ax$ Re $a > 0$	$4(2a)^{-s}\Gamma(s)\zeta(s-1)$ Re $s > 2$
18	$K_\nu(ax)$	$a^{-s}2^{s-2}\Gamma((s-\nu)/2)$ $\times\,\Gamma((s+\nu)/2)$ Re $s > \lvert$Re $\nu\rvert$

†Assuming that $\lim_{x\to 0+} x^{s-r-1} f^{(r)}(x) = 0$ for $r = 0, 1, \ldots, n-1$.
‡Where I_n denotes the n^{th} repeated integral of $f(x)$: $I_0 f(x) = f(x)$, $I_n f(x) = \int_0^x I_{n-1}(t)\, dt$.

Chapter 7

Probability and Statistics

7.1 PROBABILITY THEORY

7.1.1 INTRODUCTION

A *sample space* S associated with an experiment is a set S of elements such that any outcome of the experiment corresponds to a unique element of the set. An *event* E is a subset of a sample space S. An element in a sample space is called a *sample point* or a *simple event*.

7.1.1.1 Definition of probability

If an experiment can occur in n mutually exclusive and equally likely ways, and if exactly m of these ways correspond to an event E, then the *probability* of E is

$$P(E) = \frac{m}{n}. \tag{7.1.1}$$

If E is a subset of S, and if to each element subset of S, a non-negative number, called the probability, is assigned, and if E is the union of two or more different simple events, then the probability of E, denoted $P(E)$, is the sum of the probabilities of those simple events whose union is E.

Technically, a *probability space* consists of three parts: a sample space S (set of possible outcomes); a set of events $\{A\}$ (an event contains a set of outcomes); and a probability measure P (a function on events such that $P(A) \geq 0$, $P(S) = 1$, and $P\left(\cup_{j \in J} A_j\right) = \sum_{j \in J} P(A_j)$ if $\{A_j \mid j \in J\}$ is a countable, pairwise disjoint collection of events).

7.1.1.2 Marginal and conditional probability

Suppose a sample space S is partitioned into rs disjoint subsets where the general subset is denoted $E_i \cap F_j$ (with $i = 1, 2, \ldots, r$ and $j = 1, 2, \ldots, s$). Then the *marginal probability* of E_i is defined as

$$P(E_i) = \sum_{j=1}^{s} P(E_i \cap F_j), \tag{7.1.2}$$

and the marginal probability of F_j is defined as

$$P(F_j) = \sum_{i=1}^{r} P(E_i \cap F_j). \tag{7.1.3}$$

The *conditional probability* of E_i, given that F_j has occurred, is defined as

$$P(E_i \mid F_j) = \frac{P(E_i \cap F_j)}{P(F_j)}, \qquad \text{when } P(F_j) \neq 0 \tag{7.1.4}$$

and that of F_j, given that E_i has occurred, is defined as

$$P(F_j \mid E_i) = \frac{P(E_i \cap F_j)}{P(E_i)}, \qquad \text{when } P(E_i) \neq 0. \tag{7.1.5}$$

7.1.1.3 Probability theorems

1. If $\emptyset$ is the null set, then $P(\emptyset) = 0$.
2. If S is the sample space, then $P(S) = 1$.
3. If E and F are two events, then
$$P(E \cup F) = P(E) + P(F) - P(E \cap F). \tag{7.1.6}$$
4. If E and F are *mutually exclusive events*, then
$$P(E \cup F) = P(E) + P(F). \tag{7.1.7}$$
5. If E and E' are *complementary events*, then
$$P(E) = 1 - P(E'). \tag{7.1.8}$$
6. Two events are said to be *independent* if and only if
$$P(E \cap F) = P(E)\,P(F). \tag{7.1.9}$$
The event E is said to be *statistically independent* of the event F if $P(E \mid F) = P(E)$ and $P(F \mid E) = P(F)$.
7. The events $\{E_1, \dots, E_n\}$ are called *mutually independent* if and only if every combination of these events taken any number of times is independent.
8. *Bayes' rule*: If $\{E_1, \dots, E_n\}$ are n mutually exclusive events whose union is the sample space S, and if E is any event of S such that $P(E) \neq 0$, then
$$P(E_k \mid E) = \frac{P(E_k)\,P(E \mid E_k)}{P(E)} = \frac{P(E_k)\,P(E \mid E_k)}{\sum_{j=1}^{n} P(E_j)\,P(E \mid E_j)}. \tag{7.1.10}$$
9. For a uniform probability distribution,
$$P(A) = \frac{\text{Number of outcomes in event } A}{\text{Total number of outcomes}}. \tag{7.1.11}$$

7.1.1.4 Terminology

1. A function whose domain is a sample space S and whose range is some set of real numbers is called a *random variable*. This random variable is called *discrete* if it assumes only a finite or denumerable number of values. It is called *continuous* if it assumes a continuum of values.
2. Random variables are usually represented by capital letters (e.g., X, Y, Z).
3. "iid" or "i.i.d." is often used for the phrase "independent and identically distributed."
4. Many probability distributions have special representations:
   (a) χ^2_n: chi-square random variable with n degrees of freedom
   (b) $E(\lambda)$: exponential distribution with parameter λ
   (c) $N(\mu, \sigma)$: normal random variable with mean μ and standard deviation σ
   (d) $P(\lambda)$: Poisson distribution with parameter λ
   (e) $U[a, b)$: uniform random variable on the interval $[a, b)$

7.1.1.5 Characterizing random variables

The *density function* is defined as follows:

1. When X is a continuous random variable, let $f(x)\,dx$ denote the probability that X lies in the region $[x, x+dx]$; $f(x)$ is called the *probability density function*. (We require $f(x) \geq 0$ and $\int f(x)\,dx = 1$). Mathematically, for any event E,
$$P(E) = P(X \text{ is in } E) = \int_E f(x)\,dx. \tag{7.1.12}$$
2. When X is a discrete random variable, let p_k for $k = 0, 1, \ldots$ be the probability that $X = x_k$ (with $p_k \geq 0$ and $\sum_k p_k = 1$). Mathematically, for any event E,
$$P(E) = P(X \text{ is in } E) = \sum_{x_k \in E} p_k. \tag{7.1.13}$$

A discrete random variable can be written as a continuous density
$$f(x) = \sum_k p_k \delta(x - x_k).$$

The *cumulative distribution function*, or simply the *distribution function*, is
$$F(x) = \mathrm{P}(X \leq x) = \begin{cases} \displaystyle\sum_{x_k \leq x} p_k, & \text{in the discrete case,} \\ \displaystyle\int_{-\infty}^{x} f(t)\,dt, & \text{in the continuous case.} \end{cases} \tag{7.1.14}$$

Note that $F(-\infty) = 0$ and $F(\infty) = 1$. The probability that X is between a and b is
$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a). \tag{7.1.15}$$

Let $g(X)$ be a function of X. The *expected value* (or *expectation*) of $g(X)$, denoted by $\mathrm{E}\left[g(X)\right]$, is given by (when it exists)
$$\mathrm{E}\left[g(X)\right] = \begin{cases} \sum_k p_k g(x_k), & \text{in the discrete case,} \\ \int_{\mathbb{R}} g(t) f(t)\,dt, & \text{in the continuous case.} \end{cases} \tag{7.1.16}$$

1. $\mathrm{E}\left[aX + bY\right] = a\mathrm{E}\left[X\right] + b\mathrm{E}\left[Y\right]$.
2. $\mathrm{E}\left[XY\right] = \mathrm{E}\left[X\right]\mathrm{E}\left[Y\right]$ if X and Y are independent.

The *moments* of X are defined by $\mu'_k = \mathrm{E}\left[X^k\right]$. The first moment, μ'_1, is called the *mean* of X; it is usually denoted by $\mu = \mu'_1 = \mathrm{E}\left[X\right]$. The *centered moments* of X are defined by $\mu_k = \mathrm{E}\left[(X-\mu)^k\right]$. The second centered moment is called the *variance* and is denoted by $\sigma^2 = \mu_2 = \mathrm{E}\left[(X-\mu)^2\right]$. Here, σ is called the *standard deviation*. The *skewness* is $\gamma_1 = \mu_3/\sigma^3$, and the *excess* or *kurtosis* is $\gamma_2 = (\mu_4/\sigma^4) - 3$.

Using σ_Z^2 to denote the variance for the random variable Z, we have

1. $\sigma_{cX}^2 = c^2 \sigma_X^2$.
2. $\sigma_{c+X}^2 = \sigma_X^2$.
3. $\sigma_{aX+b}^2 = a^2 \sigma_X^2$.

7.1.1.6 Generating and characteristic functions

In the case of a discrete distribution, the *generating function* for the random variable X (when it exists) is given by $G(s) = G_X(s) = \mathrm{E}\left[s^X\right] = \sum_{k=0}^{\infty} p_k s^{x_k}$. From this function, the moments may be found from

$$\mu_n' = \left(s\frac{\partial}{\partial s}\right)^n G(s)\Big|_{s=1}. \tag{7.1.17}$$

1. If c is a constant, then the generating function of $c + X$ is $s^c G(s)$.
2. If c is a constant, then the generating function of cX is $G(s^c)$.
3. If $Z = X + Y$ where X and Y are independent discrete random variables, then $G_Z(s) = G_X(s)G_Y(s)$.
4. If $Y = \sum_{i=1}^{n} X_i$, the $\{X_i\}$ are independent, and each X_i has the common generating function $G_X(s)$, then the generating function of Y is $[G_X(s)]^n$.

In the case of a continuous distribution, the *characteristic function* for the random variable X is given by $\phi(t) = \mathrm{E}\left[e^{itX}\right] = \int_{-\infty}^{\infty} e^{itx} f(x)\,dx$; the Fourier transform of $f(x)$. From this function, the moments may be found: $\mu_n' = i^{-n}\phi^{(n)}(0)$. If $Z = X + Y$ where X and Y are independent continuous random variables, then $\phi_Z(t) = \phi_X(t)\phi_Y(t)$. The *cumulant function* is defined as the logarithm of the characteristic function. The n^{th} *cumulant*, κ_n, is defined as a certain term in the Taylor series of the cumulant function,

$$\log\phi(t) = \sum_{n=0}^{\infty} \kappa_n \frac{(it)^n}{n!}. \tag{7.1.18}$$

Note that $\kappa_1 = \mu$, $\kappa_2 = \sigma^2$, $\kappa_3 = \mu_3$, and $\kappa_4 = \mu_4 - 3\mu_2^2$. For a normal probability distribution, $\kappa_n = 0$ for $n \geq 3$. The centered moments in terms of cumulants are

$$\begin{aligned} \mu_2 &= \kappa_2, \\ \mu_3 &= \kappa_3, \\ \mu_4 &= \kappa_4 + 3\kappa_2^2, \\ \mu_5 &= \kappa_5 + 10\kappa_3\kappa_2, \end{aligned} \tag{7.1.19}$$

For both discrete and continuous distributions, the *moment generating function* (when it exists) for the random variable X is given by $G(t) = G_X(t) = \mathrm{E}\left[e^{tX}\right]$. (For a multi-dimensional random variable: $G(\mathbf{t}) = G_{\mathbf{X}}(\mathbf{t}) = \mathrm{E}\left[e^{\mathbf{t}^{\mathrm{T}}\mathbf{X}}\right]$.) The n^{th} moment is then given by $\mu_n' = \frac{d^n}{dt^n}G(t)\Big|_{t=0}$.

7.1.2 CENTRAL LIMIT THEOREM

If $\{X_i\}$ are independent and identically distributed random variables with mean μ and finite variance σ^2, then the random variable

$$Z = \frac{(X_1 + X_2 + \cdots + X_n) - n\mu}{\sqrt{n}\sigma} \tag{7.1.20}$$

tends (as $n \to \infty$) to a normal random variable with mean zero and variance one.

7.1.3 MULTIVARIATE DISTRIBUTIONS

7.1.3.1 Discrete case

The k-dimensional random variable $(X_1, \ldots, X_k)$ is a k-dimensional discrete random variable if it assumes values only at a finite or denumerable number of points $(x_1, \ldots, x_k)$. Define

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = f(x_1, x_2, \ldots, x_k) \tag{7.1.21}$$

for every value that the random variable can assume. The function $f(x_1, \ldots, x_k)$ is called the *joint density* of the k-dimensional random variable. If E is any subset of the set of values that the random variable can assume, then

$$P(E) = P\left[(X_1, \ldots, X_k) \text{ is in } E\right] = \sum_E f(x_1, \ldots, x_k) \tag{7.1.22}$$

where the sum is over all the points $(x_1, \ldots, x_k)$ in E. The cumulative distribution function is defined as

$$F(x_1, x_2, \ldots, x_k) = \sum_{z_1 \le x_1} \sum_{z_2 \le x_2} \cdots \sum_{z_k \le x_k} f(z_1, z_2, \ldots, z_k). \tag{7.1.23}$$

7.1.3.2 Continuous case

The k random variables $(X_1, \ldots, X_k)$ are said to be jointly distributed if a function f exists so that $f(x_1, \ldots, x_k) \ge 0$ for all $-\infty < x_i < \infty$ ($i = 1, \ldots, k$) and so that, for any given event E,

$$\begin{aligned} P(E) &= P[(X_1, X_2, \ldots, X_k) \ \text{ is in } E] \\ &= \int_E \cdots \int f(x_1, x_2, \ldots, x_k)\, dx_1\, dx_2 \cdots dx_k. \end{aligned} \tag{7.1.24}$$

The function $f(x_1, \ldots, x_k)$ is called the *joint density* of the random variables X_1, X_2, …, X_k. The cumulative distribution function is defined as

$$F(x_1, x_2, \ldots, x_k) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_k} f(z_1, z_2, \ldots, z_k)\, dz_k \cdots dz_2\, dz_1. \tag{7.1.25}$$

Given the cumulative distribution function, the probability density may be found from

$$f(x_1, x_2, \ldots, x_k) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \cdots \frac{\partial}{\partial x_k} F(x_1, x_2, \ldots, x_k). \tag{7.1.26}$$

7.1.3.3 Moments

The r^{th} moment of X_i is defined as

$$\mathrm{E}\left[X_i^r\right] = \begin{cases} \sum_{x_1} \cdots \sum_{x_k} x_i^r f(x_1, \ldots, x_k), & \text{in the discrete case,} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_i^r f(x_1, \ldots, x_k)\, dx_1 \cdots dx_k & \text{in the continuous case.} \end{cases}$$

Joint moments about the origin are defined as $\mathrm{E}\left[X_1^{r_1} X_2^{r_2} \cdots X_k^{r_k}\right]$ where $r_1 + r_2 + \cdots + r_k$ is the order of the moment. Joint moments about the mean are defined as $\mathrm{E}\left[(X_1 - \mu_1)^{r_1} (X_2 - \mu_2)^{r_2} \cdots (X_k - \mu_k)^{r_k}\right]$, where $\mu_k = \mathrm{E}\left[X_k\right]$.

7.1.3.4 Marginal and conditional distributions

If the random variables X_1, X_2, ..., X_k have the joint density function $f(x_1, x_2, \ldots, x_k)$, then the *marginal distribution* of the subset of the random variables, say, X_1, X_2, ..., X_p (with $p < k$), is given by

$$g(x_1, x_2, \ldots, x_p) = \begin{cases} \sum_{x_{p+1}} \sum_{x_{p+2}} \cdots \sum_{x_k} f(x_1, x_2, \ldots, x_k), & \text{in the discrete case,} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_k)\, dx_{p+1} \, \cdots \, dx_k, & \text{in the continuous case.} \end{cases} \tag{7.1.27}$$

The *conditional distribution* of a certain subset of the random variables is the joint distribution of this subset under the condition that the remaining variables are given certain values. The conditional distribution of X_1, X_2, ..., X_p, given X_{p+1}, X_{p+2}, ..., X_k, is

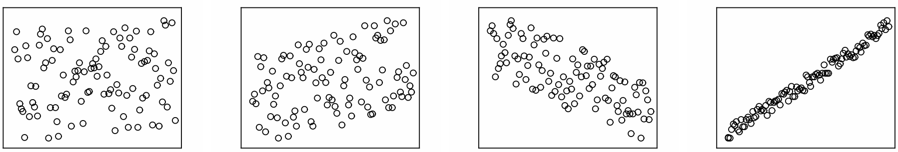
$$h(x_1, \ldots, x_p \mid x_{p+1}, \ldots, x_k) = \frac{f(x_1, x_2, \ldots, x_k)}{g(x_{p+1}, x_{p+2}, \ldots, x_k)} \tag{7.1.28}$$

if $g(x_{p+1}, x_{p+2}, \ldots, x_k) \neq 0$.

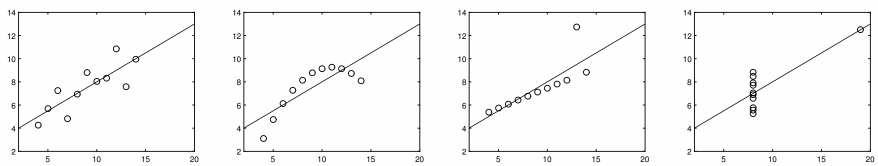
The *variance* σ_{ii} of X_i and the *covariance* σ_{ij} of X_i and X_j are given by

$$\begin{aligned} \sigma_{ii}^2 &= \sigma_i^2 = \mathrm{E}\left[(X_i - \mu_i)^2\right], \\ \sigma_{ij}^2 &= \rho_{ij}\sigma_i\sigma_j = \mathrm{E}\left[(X_i - \mu_i)(X_j - \mu_j)\right], \end{aligned} \tag{7.1.29}$$

where ρ_{ij} is the *correlation coefficient*, and σ_i and σ_j are the standard deviations of X_i and X_j. The following figures show data sets with varying correlation coefficients, from left to right: $\rho = 0.086$, $\rho = 0.28$, $\rho = -0.71$, $\rho = 0.99$.



Numerical information alone, such as correlation values, can be misleading. The following 4 data sets (Anscombe's quartet), each with 11 values, all have $\rho = 0.82$. (They also have, to many decimal places, the same mean in x, mean in y, best fit straight line, and several other parameters.)



7.1.4 AVERAGES OVER VECTORS

Let $\overline{f(\mathbf{n})}$ denote the expectation of the function f as the unit vector $\mathbf{n}$ varies uniformly in all directions in three dimensions. If $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ are constant vectors, then

$$\begin{aligned}
\overline{|\mathbf{a}\cdot\mathbf{n}|^2} &= |\mathbf{a}|^2/3,\\
\overline{(\mathbf{a}\cdot\mathbf{n})(\mathbf{b}\cdot\mathbf{n})} &= (\mathbf{a}\cdot\mathbf{b})/3,\\
\overline{(\mathbf{a}\cdot\mathbf{n})\mathbf{n}} &= \mathbf{a}/3,\\
\overline{|\mathbf{a}\times\mathbf{n}|^2} &= 2\,|\mathbf{a}|^2/3,\\
\overline{(\mathbf{a}\times\mathbf{n})\cdot(\mathbf{b}\times\mathbf{n})} &= 2\mathbf{a}\cdot\mathbf{b}/3,\\
\overline{(\mathbf{a}\cdot\mathbf{n})(\mathbf{b}\cdot\mathbf{n})(\mathbf{c}\cdot\mathbf{n})(\mathbf{d}\cdot\mathbf{n})} &= \left[(\mathbf{a}\cdot\mathbf{b})(\mathbf{c}\cdot\mathbf{d}) + (\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d}) + (\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})\right]/15.
\end{aligned} \tag{7.1.30}$$

Now let $\overline{f(\mathbf{n})}$ denote the average of the function f as the unit vector $\mathbf{n}$ varies uniformly in all directions in two dimensions. If $\mathbf{a}$ and $\mathbf{b}$ are constant vectors, then

$$\begin{aligned}
\overline{|\mathbf{a}\cdot\mathbf{n}|^2} &= |\mathbf{a}|^2/2,\\
\overline{(\mathbf{a}\cdot\mathbf{n})(\mathbf{b}\cdot\mathbf{n})} &= (\mathbf{a}\cdot\mathbf{b})/2,\\
\overline{(\mathbf{a}\cdot\mathbf{n})\mathbf{n}} &= \mathbf{a}/2.
\end{aligned} \tag{7.1.31}$$

7.1.5 TRANSFORMING VARIABLES

1. Suppose that the random variable X has the probability density function $f_X(x)$ and the random variable Y is defined by $Y = g(X)$. If g is measurable and one-to-one, then

$$f_Y(y) = f_X(h(y))\left|\frac{dh}{dy}\right| \tag{7.1.32}$$

   where $h(y) = g^{-1}(y)$.
2. If the random variables X and Y are independent and if their densities f_X and f_Y, respectively, exist almost everywhere, then the probability density of their sum, $Z = X + Y$, is given by their convolution,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,dx. \tag{7.1.33}$$

3. If the random variables X and Y are independent and if their densities f_X and f_Y, respectively, exist almost everywhere, then the probability density of their product, $Z = XY$, is given by the formula,

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y\left(\frac{z}{x}\right)\,dx. \tag{7.1.34}$$

7.1.6 INEQUALITIES

1. *Markov inequality:* If X is a random variable which takes only non-negative values, then for any $a > 0$,

$$P(X \geq a) \leq \frac{\mathrm{E}[X]}{a}. \tag{7.1.35}$$

2. *Cauchy–Schwartz inequality:* Let X and Y be random variables for which $\mathrm{E}\left[X^2\right]$ and $\mathrm{E}\left[Y^2\right]$ exist, then

$$(\mathrm{E}[XY])^2 \leq \mathrm{E}\left[X^2\right] \mathrm{E}\left[Y^2\right]. \tag{7.1.36}$$

3. *One-sided Chebyshev inequality:* Let X be a random variable with zero mean (i.e., $\mathrm{E}[X] = 0$) and variance σ^2. Then, for any positive a,

$$P(X > a) \leq \frac{\sigma^2}{\sigma^2 + a^2}. \tag{7.1.37}$$

4. *Chebyshev inequality:* Let c be any real number and let X be a random variable for which $\mathrm{E}\left[(X-c)^2\right]$ is finite. Then, for every $\epsilon > 0$ the following holds:

$$P(|X - c| \geq \epsilon) \leq \frac{1}{\epsilon^2}\mathrm{E}\left[(X-c)^2\right]. \tag{7.1.38}$$

5. *Bienaymé–Chebyshev inequality:* If $E\left[|X|^r\right] < \infty$ for all $r > 0$ (r not necessarily an integer) then, for every $a > 0$,

$$P\left(|X| \geq a\right) \leq \frac{E\left[|X|^r\right]}{a^r}. \tag{7.1.39}$$

6. *Generalized Bienaymé–Chebyshev inequality:* Let $g(x)$ be a non-decreasing non-negative function defined on $(0, \infty)$. Then, for $a \geq 0$,

$$P\left(|X| \geq a\right) \leq \frac{E\left[g(|X|)\right]}{g(a)}. \tag{7.1.40}$$

7. *Chernoff bound:* This bound is useful for sums of random variables. Let $Y_n = \sum_{i=1}^{n} X_i$ where the $\{X_i\}$ are iid. Let $M(t) = \mathrm{E}_x[e^{tX}]$ be the common moment generating function for the $\{X_i\}$, and define $g(t) = \log M(t)$. Then (the prime in this formula denotes a derivative),

$$\begin{aligned} P\left(Y_n \geq ng'(t)\right) &\leq e^{-n\left[tg'(t) - g(t)\right]}, && \text{if } t \geq 0, \\ P\left(Y_n \leq ng'(t)\right) &\leq e^{-n\left[tg'(t) - g(t)\right]}, && \text{if } t \leq 0. \end{aligned}$$

8. *Kolmogorov inequality:* Let $X_1, X_2, \ldots, X_n$ be n independent random variables such that $\mathrm{E}[X_i] = 0$ and $\mathrm{Var}(X_i) = \sigma_{X_i}^2$ is finite. Then, for all $a > 0$,

$$P\left(\max_{i=1,\ldots,n} |X_1 + X_2 + \cdots + X_i| > a\right) \leq \sum_{i=1}^{n} \frac{\sigma_i^2}{a^2}. \tag{7.1.41}$$

9. *Jensen inequality:* If $\mathrm{E}[X]$ exists, and if $f(x)$ is a convex $\cup$ ("convex cup") function, then

$$\mathrm{E}[f(X)] \geq f(\mathrm{E}[X]). \tag{7.1.42}$$

7.1.7 GEOMETRIC PROBABILITY

1. *Points in a line segment*: If A and B are uniformly and independently chosen from the interval $[0, 1)$, and X is the distance between A and B (that is, $X = |A - B|$) then the probability density of X is $f_X(x) = 2(1 - x)$.
2. *Many points in a line segment*: Uniformly and independently choose $n-1$ random values in the interval $[0, 1)$. This creates n intervals.

$$\begin{aligned} P_k(x) &= \text{Probability (exactly } k \text{ intervals have length larger than } x) \\ &= \binom{n}{k} \left\{ [1 - kx]^{n-1} - \binom{n-k}{1} [1 - (k+1)x]^{n-1} + \right. \\ &\qquad \left. \cdots + (-1)^s \binom{n-k}{s} [1 - (k+s)x]^{n-1} \right\}, \end{aligned} \tag{7.1.43}$$

where $s = \left\lfloor \frac{1}{x} - k \right\rfloor$. From this, the probability that the largest interval length exceeds x is

$$1 - P_0(x) = \binom{n}{1}(1-x)^{n-1} - \binom{n}{2}(1-2x)^{n-1} + \dots. \tag{7.1.44}$$

3. *Points in the plane*: Assume that the number of points in any region A of the plane is a Poisson variate with mean λA (λ is the "density" of the points). Given a fixed point P define R_1, R_2, $\dots$, to be the distance to the point nearest to P, second nearest to P, etc. Then

$$f_{R_s}(r) = \frac{2(\lambda\pi)^s}{(s-1)!} r^{2s-1} e^{-\lambda\pi r^2}. \tag{7.1.45}$$

4. *Points in three-dimensional space*: Assume that the number of points in any volume V is a Poisson variate with mean λV (λ is the "density" of the points). Given a fixed point P define R_1, R_2, $\dots$, to be the distance to the point nearest to P, second nearest to P, etc. Then

$$f_{R_s}(r) = \frac{3\left(\frac{4}{3}\lambda\pi\right)^s}{\Gamma(s)} r^{3s-1} e^{-\frac{4}{3}\lambda\pi r^3}. \tag{7.1.46}$$

5. *Points on a checkerboard*: Consider the unit squares on a checkerboard and select one point uniformly and independently in each square. The following results concern the average distance between points:
   (a) For adjacent squares (a black and white square with a common side) the mean distance between points is 1.088.
   (b) For diagonal squares (two white squares with a point in common) the mean between points is 1.473.
6. *Points in a cube*: Choose two points uniformly and independently within a unit cube. The distance between these points has mean 0.66171 and standard deviation 0.06214.

7. *Points in an n-dimensional cube*: Let two points be selected uniformly and independently within a unit n-dimensional cube. The expected distance between the points, $\Delta(n)$, is

- $\Delta(1) = \frac{1}{3}$
- $\Delta(2) \approx 0.54141$
- $\Delta(3) \approx 0.66171$
- $\Delta(4) \approx 0.77766$
- $\Delta(5) \approx 0.87852$
- $\Delta(6) \approx 0.96895$
- $\Delta(7) \approx 1.05159$
- $\Delta(8) \approx 1.12817$

8. *Points on a circle*: Select three points uniformly and independently on a unit circle. These points determine a triangle with area A. The mean and variance of the area are:

$$\begin{aligned} \mu_A &= \frac{3}{2\pi} \approx 0.4775 \\ \sigma_A^2 &= \frac{3\left(\pi^2 - 6\right)}{8\pi^2} \approx 0.1470 \end{aligned} \tag{7.1.47}$$

9. *Buffon needle problem*: A needle of length L is randomly placed on a plane on which has parallel lines a distance D apart. If $\frac{L}{D} < 1$ then only one intersection is possible. The probability P that the needle intersects a line is

$$P = \begin{cases} \dfrac{2L}{\pi D} & \text{if } 0 < L \leq D, \\ \dfrac{2L}{\pi D}\left(1 - \sqrt{1 - \left(\dfrac{D}{L}\right)^2}\right) + \left(1 - \dfrac{2}{\pi}\sin^{-1}\dfrac{D}{L}\right) & \text{if } 0 < D \leq L, \end{cases} \tag{7.1.48}$$

7.1.8 RANDOM SUMS OF RANDOM VARIABLES

If $T = \sum_{i=1}^{N} X_i$, and if N is an integer-valued random variable with generating function $G_N(s)$, and if the $\{X_i\}$ are discrete independent and identically distributed random variables with generating function $G_X(s)$, and the $\{X_i\}$ are independent of N, then the generating function for T is $G_T(s) = G_N(G_X(s))$. (If the $\{X_i\}$ are continuous random variables, then $\phi_T(t) = G_N(\phi_X(t))$.) Hence,

1. $\mu_T = \mu_N \mu_X$.
2. $\sigma_T^2 = \mu_N \sigma_X^2 + \mu_X^2 \sigma_N^2$.

EXAMPLE A game is played as follows. A fair coin is flipped until a tail shows up and then the game stops. When a head appears, a die is rolled and the number shown is the number of dollars paid. What is the amount expected to be paid playing this game?

1. The number of coin flips, N, has a geometric distribution so that the expected number of flips is $\mu_N = \mu_{\text{flips}} = 2$.
2. The average return for the die is $\mu_X = \mu_{\text{die}} = \frac{1}{6}(1 + 2 + \cdots + 6) = \frac{7}{2}$.
3. Hence the expected return is $\mu_T = \mu_{\text{flips}}\mu_{\text{die}} = 7$; or \$7 per game.

7.2 CLASSICAL PROBABILITY PROBLEMS

7.2.1 BIRTHDAY PROBLEM

The probability that n people have different birthdays (neglecting February 29[th]) is

$$q_n = \left(\frac{364}{365}\right) \cdot \left(\frac{363}{365}\right) \cdots \left(\frac{366-n}{365}\right) \tag{7.2.1}$$

Let $p_n = 1 - q_n$. For 23 independent people the probability of at least two people having the same birthday is more than half ($p_{23} = 1 - q_{23} > 1/2$).

n	10	20	23	30	40	50
p_n	0.117	0.411	0.507	0.706	0.891	0.970

That is, the number of people needed to have a 50% chance of two people having the same birthday is 23. The number of people needed to have a 50% chance of three people having the same birthday is 88. For four, five, and six people having the same birthday the number of people necessary is 187, 313, and 460.

The number of people needed so that there is a 50% chance that two people have a birthday within one day of each other is 14. In general, in an n-day year the probability that p people all have birthdays at least k days apart (so that $k = 1$ and $p = 2$ is the original birthday problem) is

$$\text{probability} = \binom{n - p(k-1) - 1}{p - 1} \frac{(p-1)!}{n^{p-1}}. \tag{7.2.2}$$

7.2.2 COUPON COLLECTORS PROBLEM

There are n coupons that can be collected. Random coupons are selected, with replacement. How long must one wait until they have a specified collection of coupons?

Let $W_{n,j}$ be the number of steps until j different coupons are seen; then

$$\mathrm{E}\left[W_{n,j}\right] = n \sum_{i=1}^{j} \frac{1}{n-i+1} \qquad \mathrm{Var}\left[W_{n,j}\right] = n \sum_{i=1}^{j} \frac{i-1}{(n-i+1)^2} \tag{7.2.3}$$

When $j = n$, then all coupons are being collected and $\mathrm{E}\left[W_{n,n}\right] = nH_n$ with $H_n = 1 + 1/2 + 1/3 + \cdots + 1/n$. As $n \to \infty$, $\mathrm{E}\left[W_{n,n}\right] \sim n \log n + \gamma n + \frac{1}{2}$.

n	2	5	10	50	100	200
$\mathrm{E}\left[W_{n,n}\right]$	3	11.4	29.3	225	519	1,176
$\sigma_{n,n} = \sqrt{\mathrm{Var}\left[W_{n,n}\right]}$	1.4	5.0	11.2	62	126	254

7.2.3 CARD GAMES

7.2.3.1 Poker hands

The number of distinct 5-card poker hands is $\binom{52}{5} = 2{,}598{,}960$.

	Given 5 cards		Best 5 of 7 cards	
Hand	Probability	Odds	Probability	Odds
royal flush	1.54×10^{-6}	649,739:1	0.00031	3,226:1
straight flush	1.39×10^{-5}	72,192:1	0.0017	588:1
four of a kind	2.40×10^{-4}	4,164:1	0.026	38:1
full house	1.44×10^{-3}	693:1	0.030	33:1
flush	1.97×10^{-3}	508:1	0.046	22:1
straight	3.92×10^{-3}	254:1	0.048	21:1
three of a kind	0.0211	46:1	0.235	4.2:1
two pair	0.0475	20:1	0.438	2.3:1
one pair	0.423	1.37:1	0.174	5.7:1

There are 7,462 distinct 5-card poker hand equivalence classes.[1] For the equivalence classes a royal flush is counted once, not 4 times. Similarly, when choosing the best 5 cards in 7-card poker, there are 4,824 equivalence classes. There are fewer than 7,462 equivalence classes because some hands cannot be the best 5 of 7 cards; for example, a high card hand where the highest card is a 7.

Hand	**Unique**	**Distinct**	**Best 5 of 7**
royal flush	4	1	1
straight flush	40	10	10
four of a kind	624	156	156
full house	3,744	156	156
flush	5,108	1,277	1,277
straight	10,200	10	10
three of a kind	4,912	858	575
two pair	123,552	858	763
one pair	1,098,240	2,860	1,470
high card	1,302,540	1,277	407
TOTAL	2,598,960	7,462	4,824

For multiple players, each with their own cards, the probability of the winning hand (the one that is the best among all players) is shown below. An ϵ represents a probability less than 0.0001, or 1 in ten thousand. For example when 5 people are playing poker the winning hand will be one pair 63% of the time.

[1] Two poker hands are in the same equivalence class if, when playing poker, the hands would be tied.

5 cards dealt	**number of players**							
	2	3	4	5	6	7	8	9
straight flush	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
four of a kind	0.0005	0.001	0.001	0.001	0.001	0.002	0.001	0.002
full house	0.003	0.005	0.007	0.008	0.007	0.009	0.01	0.02
flush	0.005	0.006	0.009	0.01	0.01	0.01	0.02	0.02
straight	0.007	0.01	0.02	0.02	0.02	0.02	0.03	0.03
three of a kind	0.04	0.06	0.08	0.10	0.12	0.14	0.14	0.16
two pair	0.09	0.13	0.16	0.19	0.22	0.24	0.26	0.28
one pair	0.60	0.66	0.67	0.63	0.60	0.57	0.54	0.49
high card	0.26	0.12	0.06	0.04	0.02	0.009	0.005	0.002

7 cards dealt	**number of players**					
	2	3	4	5	6	7
straight flush	0.0005	0.0008	0.001	0.001	0.002	0.002
four of a kind	0.003	0.005	0.007	0.009	0.01	0.01
full house	0.05	0.07	0.10	0.12	0.14	0.16
flush	0.06	0.08	0.11	0.13	0.15	0.16
straight	0.08	0.11	0.13	0.15	0.17	0.18
three of a kind	0.09	0.11	0.13	0.14	0.15	0.15
two pair	0.32	0.36	0.36	0.34	0.31	0.28
one pair	0.37	0.25	0.17	0.11	0.07	0.05
high card	0.04	0.008	0.002	0.0003	ϵ	ϵ

7.2.3.2 Poker with wild cards

Using wild cards in poker leads to a paradox. The usual hierarchy of poker hands (straight flush, four of a kind, full house, flush, straight, three of a kind, two pair, one pair, high card) corresponds to the likelihood of those hands appearing.

Assume that 1 or 2 jokers are added to a deck and that they are wild (so the deck has either 53 or 54 cards). The number of possible 5-card hands is now $\binom{53}{5} = 2,869,685$ or $\binom{54}{5} = 3,162,510$. The distribution of these hands when the wild card is used to make the best possible hand in the usual hierarchy is

Hand	**one joker**	**two jokers**
five of a kind	13	78
straight flush	204	624
four of a kind	3,120	9,360
full house	6,552	9,360
flush	7,804	11,388
straight	20,532	34,708
three of a kind	137,280	232,968
two pair	123,552	123,552
one pair	1,268,088	1,437,936
high card	1,302,540	1,302,540
TOTAL	2,869,685	3,162,510

In each case, three of a kind is more likely than two pair. Hence, two pair is more rare and should be a rated higher than three of a kind; the hierarchy could be changed to reflect this. Then, the hand $9973W$ (i.e., 2 nines, a seven, a three, and a wild card) would be considered to be two pair (i.e., wild card is a seven) rather than three of kind (i.e., wild card is a nine).

Unfortunately, when assigning all the hands in this revised hierarchy, the number of hands with three of kind is now fewer than the number of hands with two pair. So this revised hierarchy is also inconsistent.

7.2.3.3 Bridge hands

The number of distinct 13-card bridge hands is $\binom{52}{13} = 635{,}013{,}559{,}600$.

In bridge, the *honors* are the ten, jack, queen, king, or ace. Obtaining the three top cards (ace, king, and queen) of three suits and the ace, king, queen, and jack of the remaining suit is called *13 top honors*. Obtaining all cards of the same suit is called a *13-card suit*. Obtaining 12 cards of the same suit with ace high and the 13th card not an ace is called a *12-card suit, ace high*. Obtaining no honors is called a *Yarborough*.

Hand	Probability	Odds
13 top honors	6.30×10^{-12}	158,753,389,899:1
13-card suit	6.30×10^{-12}	158,753,389,899:1
12-card suit, ace high	2.72×10^{-9}	367,484,698:1
Yarborough	5.47×10^{-4}	1,827:1
four aces	2.64×10^{-3}	378:1
nine honors	9.51×10^{-3}	104:1

Assign card points as follows: ace=4, king=3, queen=3, and jack=2. Then each hand has an aggregate number of points. For a random hand let P be the probability of obtaining N points and let C be the probability of obtaining at least N points.

The likelihood of obtaining a hand with a specified number of points is below. For example, the probability of getting 15 points is 4.4% and the probability of getting 15at least points is 14.2%.

N	P	C
27	0.0000	0.0000
26	0.0001	0.0002
25	0.0003	0.0005
24	0.0006	0.0010
23	0.0011	0.0021
22	0.0021	0.0042
21	0.0038	0.0080
20	0.0064	0.0145
19	0.0104	0.0248

N	P	C
18	0.0161	0.0409
17	0.0236	0.0645
16	0.0331	0.0976
15	0.0442	0.1418
14	0.0569	0.1988
13	0.0691	0.2679
12	0.0803	0.3482
11	0.0894	0.4376
10	0.0941	0.5317

N	P	C
9	0.0936	0.6252
8	0.0889	0.7142
7	0.0803	0.7944
6	0.0655	0.8600
5	0.0519	0.9118
4	0.0385	0.9503
3	0.0246	0.9749
2	0.0136	0.9885
1	0.0079	0.9964
0	0.0036	1

Each bridge hand may have 1 to 4 different suits. The distribution of suits can be represented as A-B-C-D where A is the number of cards in the longest suit and D is the number of cards in the shortest suit. For example, 4-3-3-3 means that the longest suit has 4 cards, and there are 3 cards in each of the other 3 suits.

The probability of each possible suit distribution is shown below. The most common distribution is 4-4-3-2 which occurs 22% of the time.

distribution	probability
13-0-0-0	6.30×10^{-12}
12-1-0-0	0.000000003
11-2-0-0	0.0000001
11-1-1-0	0.0000002
10-3-0-0	0.0000015
10-2-1-0	0.000011
10-1-1-1	0.000004
9-4-0-0	0.000010
9-3-1-0	0.00010
9-2-2-0	0.000082
9-2-1-1	0.00018
8-5-0-0	0.000031
8-4-1-0	0.00045
8-3-2-0	0.0011
8-3-1-1	0.0012
8-2-2-1	0.0019

distribution	probability
7-6-0-0	0.000056
7-5-1-0	0.0011
7-4-2-0	0.0036
7-4-1-1	0.0039
7-3-3-0	0.0027
7-3-2-1	0.0188
7-2-2-2	0.0051
6-6-1-0	0.00072
6-5-2-0	0.0065
6-5-1-1	0.0071
6-4-3-0	0.0133
6-4-2-1	0.0470
6-3-3-1	0.0345
6-3-2-2	0.0564

distribution	probability
5-5-3-0	0.0090
5-5-2-1	0.0317
5-4-4-0	0.0124
5-4-3-1	0.1293
5-4-2-2	0.1058
5-3-3-2	0.1552
4-4-4-1	0.0299
4-3-3-3	0.1054
4-4-3-2	0.2155

7.2.4 DISTRIBUTION OF DICE SUMS

When rolling two dice, the probability distribution of the sum is

$$\text{Prob}\,(\text{sum of } s) = \frac{6 - |s - 7|}{36} \qquad \text{for } 2 \leq s \leq 12. \tag{7.2.4}$$

When rolling three dice, the probability distribution of the sum is

$$\text{Prob}\,(\text{sum of } s) = \frac{1}{216} \begin{cases} \frac{1}{2}(s-1)(s-2) & \text{for } 3 \leq s \leq 8 \\ -s^2 + 21s - 83 & \text{for } 9 \leq s \leq 14 \\ \frac{1}{2}(19-s)(20-s) & \text{for } 15 \leq s \leq 18 \end{cases} \tag{7.2.5}$$

For 2 dice, the most common roll is a 7 (probability $\frac{1}{6}$). For 3 dice, the most common rolls are 10 and 11 (probability $\frac{1}{8}$ each). For 4 dice, the most common roll is a 14 (probability $\frac{73}{648}$).

Sicherman dice have sides $\{1, 2, 2, 3, 3, 4\}$ and $\{1, 3, 4, 5, 6, 8\}$. The distribution of values obtained from rolling these two dice and adding the values is the same as the distribution of the sum of two "regular" dice ($\{1, 2, 3, 4, 5, 6\}$).

7.2.5 GAMBLER'S RUIN PROBLEM

A gambler starts with z dollars. At each iteration the gambler wins one dollar with probability p or loses one dollar with probability q (with $p+q=1$). Gambling stops when the gambler has either A dollars or zero dollars.

If q_z denotes the probability of eventually stopping with A dollars ("gambler's success") then

$$q_z = \begin{cases} \dfrac{(q/p)^A - (q/p)^z}{(q/p)^A - 1} & \text{if } p \neq q, \\ \\ 1 - \dfrac{z}{A} & \text{if } p = q = \frac{1}{2}. \end{cases} \tag{7.2.6}$$

For example:

	p	q	z	A	q_z
	0.5	0.5	9	10	.900
fair	0.5	0.5	90	100	.900
games	0.5	0.5	900	1000	.900
	0.5	0.5	9000	10000	.900
biased	0.4	0.6	90	100	.017
games	0.4	0.6	90	99	.667

7.2.6 GENDER DISTRIBUTIONS

For these problems, assume there is a 50/50 chance of male or female on each birth.

1. Hospital deliveries

   Every day a large hospital delivers 1,000 babies and a small hospital delivers 100 babies. Which hospital has a better chance of having the same number of boys as girls?

   Answer: The small one. If $2n$ babies are born, then the probability of an even split is $\binom{2n}{n}2^{-2n}$. This is a decreasing function of n.

2. Family planning

   Suppose that every family continues to have children until they have a girl, then they stop having children. After many generations of families, what is the ratio of males to females?

   Answer: The ratio will be 50-50; half of all conceptions will be male, half female.

3. If a person has two children and the older one is a girl, then the probability that both children are girls is $\frac{1}{2}$.

4. If a person has two children and at least one is a girl, then the probability that both children are girls is $\frac{1}{3}$.

7.2.7 MONTE HALL PROBLEM

Consider a game in which there are three doors: one door has a prize, two doors have no prize. A player selects one of the three doors. One of the two unselected doors is opened and shown that it does not contain the prize. The player is then allowed to exchange their originally selected door for the remaining unopened door. To increase the chance of winning, the player should switch doors: without switching the probability of winning is $1/3$; by switching the probability of winning is $2/3$.

7.2.8 NON-TRANSITIVE GAMES

A non-transitive game has strategies which produce "loops" of preferences. For example, "rock-paper-scissors" is a non-transitive game.

1. Bingo cards

   Consider the 4 bingo cards shown below (labeled $A - D$). Two players each select a bingo card. Numbers from 1 to 6 are randomly drawn without replacement. If a selected number is on a card, it is marked. The first player to complete a horizontal row wins. Probabilistically, card A beats card B, card B beats card C, card C beats card D, and card D beats card A.

   A

1	2
3	4

   B

2	4
5	6

   C

1	3
4	5

   D

1	5
2	6

2. Efron's Dice For 3 dice with the sides

   - $A = \{3, 3, 5, 5, 7, 7\}$
   - $B = \{2, 2, 4, 4, 9, 9\}$
   - $C = \{1, 1, 6, 6, 8, 8\}$

   it turns out that die A beats die B, die B beats die C, and die C beats die A. ("Beats" means "has a greater chance of winning".)

3. Schwenk's Dice For 3 dice with the sides

   - $A = \{1, 1, 1, 13, 13, 13\}$
   - $B = \{0, 3, 3, 12, 12, 12\}$
   - $C = \{2, 2, 2, 11, 11, 14\}$

   it turns out that die A beats die B, die B beats die C, and die C beats die A (all with probability $\frac{7}{12}$). However, if the dice are rolled twice and the values added, then 2 of die A beats 2 of die C, 2 of die C beats 2 of die B, and 2 of die B beats 2 of die A.

7.2.9 ODDS OF WINNING THE LOTTERY

Consider a lottery in which a player chooses N values, without repetition, from the numbers 1, 2, ..., M. Then the winning N values are chosen, without repetition, from among the numbers 1, 2, ..., M. The number of ways to obtain k matching values is $\binom{N}{k}\binom{M-N}{N-k}$. The number of possible player choices is $\binom{M}{N}$. The probability of having exactly k values match is the ratio of these two numbers.

EXAMPLE In a "6/49" lottery, $N = 6$ number are chosen from $M = 49$ possibilities. The probabilities are as follows:

1. no numbers match ($k = 0$); probability $= \binom{6}{0}\binom{43}{6}/\binom{49}{6} \approx 0.4360$
2. exactly $k = 1$ number matches; probability $= \binom{6}{1}\binom{43}{5}/\binom{49}{6} \approx 0.4130$
3. exactly $k = 2$ numbers match; probability $= \binom{6}{2}\binom{43}{4}/\binom{49}{6} \approx 0.1324$
4. exactly $k = 3$ numbers match; probability $= \binom{6}{3}\binom{43}{3}/\binom{49}{6} \approx 0.0177$
5. exactly $k = 4$ numbers match; probability $= \binom{6}{4}\binom{43}{2}/\binom{49}{6} \approx 0.00097$
6. exactly $k = 5$ numbers match; probability $= \binom{6}{5}\binom{43}{1}/\binom{49}{6} \approx 0.000018$
7. all $k = 6$ numbers match; probability $= \binom{6}{6}\binom{43}{0}/\binom{49}{6} \approx 7.1 \times 10^{-8}$

7.2.10 RAISIN COOKIE PROBLEM

A baker creates enough cookie dough for $C = 1000$ raisin cookies. The number of raisins to be added to the dough, R, is to be determined.

1. If you want to be 99% certain that the *first* cookie will have at least one raisin, then $1 - \left(\frac{C-1}{C}\right)^R = 1 - \left(\frac{999}{1000}\right)^R \geq 0.99$, or $R \geq 4,603$.
2. If you want to be 99% certain that *every* cookie will have at least one raisin, then $C^{-R} \sum_{i=0}^{C} \binom{C}{i}(-1)^i (C-i)^R \geq 0.99$. Hence $R \geq 11,508$.

7.2.11 WAITING PROBLEM

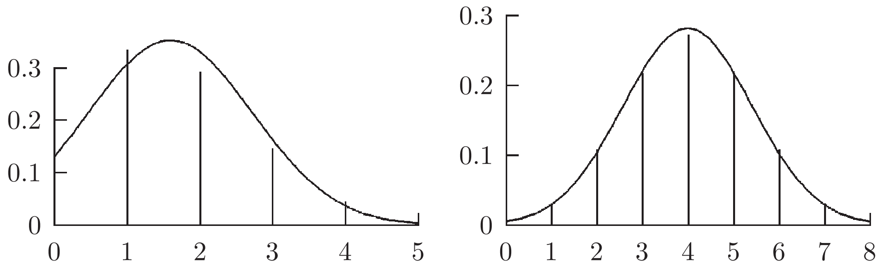
Assume two people arrive at a meeting place randomly and uniformly between 1 and 2 P.M. If person i waits w_i (as a fraction of an hour) before they leave, then the probability that the two people will be at the same place at the same time is: $P_2 = (w_1 + w_2) - \frac{w_1^2 + w_2^2}{2}$. For three people the probability of them all meeting is

$$P_3 = (w_1 w_2 + w_1 w_3 + w_2 w_3) - \frac{1}{2}\left(w_1 w_2^2 + w_1 w_3^2 + w_2 w_3^2\right) - \frac{1}{3} w_1^3 - \frac{1}{6} w_2^3$$

For n people all waiting w, the probability is $P_n = n w^{n-1} - (n-1) w^n$.

7.3 PROBABILITY DISTRIBUTIONS

FIGURE 7.1
Comparison of a binomial distribution and the approximating normal distribution. Left figure is for $(n = 8, \theta = 0.2)$, right figure is for $(n = 8, \theta = 0.5)$; horizontal axis is x. Only use the normal approximation for large n.



7.3.1 DISCRETE DISTRIBUTIONS

1. *Bernoulli distribution*: This is a special case of the binomial distribution with $n = 1$. Example: flipping a coin that has a probability p of coming up heads.
2. *Binomial distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \binom{n}{x}\theta^x(1-\theta)^{n-x}, \qquad \text{for } x = 0, 1, \ldots, n, \tag{7.3.1}$$

then the variable X has a binomial distribution. Note that $f(x)$ is the general term in the expansion of $[\theta + (1-\theta)]^n$.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = n\theta, \\ \text{Variance} &= \sigma^2 = n\theta(1-\theta), \\ \text{Standard deviation} &= \sigma = \sqrt{n\theta(1-\theta)}, \\ \text{Moment generating function} &= G(t) = [\theta e^t + (1-\theta)]^n. \end{aligned} \tag{7.3.2}$$

As $n \to \infty$ the binomial distribution approximates a normal distribution with a mean of $n\theta$ and variance of $n\theta(1-\theta)$; see Figure 7.1.
3. *Discrete uniform distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \frac{1}{n}, \qquad \text{for } x = x_1, x_2, \ldots, x_n, \tag{7.3.3}$$

then the variable X has a discrete uniform probability distribution.

Properties: When $x_i = i$ for $i = 1, 2, \ldots, n$ then

$$\begin{aligned}
\text{Mean} = \mu &= \frac{n+1}{2}, \\
\text{Variance} = \sigma^2 &= \frac{n^2-1}{12}, \\
\text{Standard deviation} = \sigma &= \sqrt{\frac{n^2-1}{12}}, \\
\text{Moment generating function} = G(t) &= \frac{e^t(1-e^{nt})}{n(1-e^t)}.
\end{aligned} \tag{7.3.4}$$

4. *Geometric distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \theta(1-\theta)^{x-1} \qquad \text{for } x = 1, 2, 3, \ldots, \tag{7.3.5}$$

then the variable X has a geometric distribution.
Properties:

$$\begin{aligned}
\text{Mean} = \mu &= \frac{1}{\theta}, \\
\text{Variance} = \sigma^2 &= \frac{1-\theta}{\theta^2}, \\
\text{Standard deviation} = \sigma &= \sqrt{\frac{1-\theta}{\theta^2}}, \\
\text{Moment generating function} = G(t) &= \frac{\theta e^t}{1-e^t(1-\theta)}.
\end{aligned} \tag{7.3.6}$$

5. *Hypergeometric distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}} \quad \text{for } x = 1, 2, 3, \ldots, \min(n, k) \tag{7.3.7}$$

then the variable X has a hypergeometric distribution.
Properties:

$$\begin{aligned}
\text{Mean} = \mu &= \frac{kn}{N}, \\
\text{Variance} = \sigma^2 &= \frac{k(N-k)n(N-n)}{N^2(N-1)}, \\
\text{Standard deviation} = \sigma &= \sqrt{\frac{k(N-k)n(N-n)}{N^2(N-1)}}.
\end{aligned} \tag{7.3.8}$$

6. *Multinomial distribution*: If a set of random variables $X_1, X_2, \ldots, X_n$ has a probability function given by

$$\begin{aligned}
P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) &= f(x_1, x_2, \ldots, x_n) \\
&= N! \prod_{i=1}^{n} \frac{\theta_i^{x_i}}{x_i!}
\end{aligned} \tag{7.3.9}$$

where the $\{x_i\}$ are positive integers, each $\theta_i > 0$, and

$$\sum_{i=1}^{n} \theta_i = 1 \quad \text{and} \quad \sum_{i=1}^{n} x_i = N, \tag{7.3.10}$$

then the joint distribution of X_1, X_2, ..., X_n is called the multinomial distribution. Note that $f(x_1, x_2, \dots, x_n)$ is a term in the expansion of $(\theta_1 + \theta_2 + \cdots + \theta_n)^N$.
Properties:

$$\begin{aligned} \text{Mean of } X_i &= \mu_i = N\theta_i, \\ \text{Variance of } X_i &= \sigma_i^2 = N\theta_i(1-\theta_i), \\ \text{Covariance of } X_i \text{ and } X_j &= \sigma_{ij}^2 = -N\theta_i\theta_j, \\ \text{Joint moment generating function} &= (\theta_1 e^{t_1} + \cdots + \theta_n e^{t_n})^N. \end{aligned} \tag{7.3.11}$$

7. *Negative binomial distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \binom{x+r-1}{r-1} \theta^r (1-\theta)^x \qquad \text{for } x = 0, 1, 2, \dots, \tag{7.3.12}$$

then the variable X has a negative binomial distribution (also known as a *Pascal* or *Polya distribution*).
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \frac{r}{\theta} - r, \\ \text{Variance} &= \sigma^2 = \frac{r}{\theta}\left(\frac{1}{\theta} - 1\right) = \frac{r(1-\theta)}{\theta^2}, \\ \text{Standard deviation} &= \sqrt{\frac{r}{\theta}\left(\frac{1}{\theta} - 1\right)} = \sqrt{\frac{r(1-\theta)}{\theta^2}}, \\ \text{Moment generating function} &= G(t) = \theta^r [1 - (1-\theta)e^t]^{-r}. \end{aligned} \tag{7.3.13}$$

8. *Poisson distribution*: If the random variable X has a probability density function given by

$$P(X = x) = f(x) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \dots, \tag{7.3.14}$$

with $\lambda > 0$, then the variable X has a Poisson distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \lambda, \\ \text{Variance} &= \sigma^2 = \lambda, \\ \text{Standard deviation} &= \sigma = \sqrt{\lambda}, \\ \text{Moment generating function} &= G(t) = e^{\lambda(e^t - 1)}. \end{aligned} \tag{7.3.15}$$

7.3.2 CONTINUOUS DISTRIBUTIONS

1. *Normal distribution*: If the random variable X has the density function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \qquad \text{for } -\infty < x < \infty, \tag{7.3.16}$$

then the variable X has a normal distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu, \\ \text{Variance} &= \sigma^2, \\ \text{Standard deviation} &= \sigma, \\ \text{Moment generating function} &= G(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right). \end{aligned} \tag{7.3.17}$$

(a) Set $y = \frac{x-\mu}{\sigma}$ to obtain a standard normal distribution.
(b) The cumulative distribution function is

$$F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{x} \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt.$$

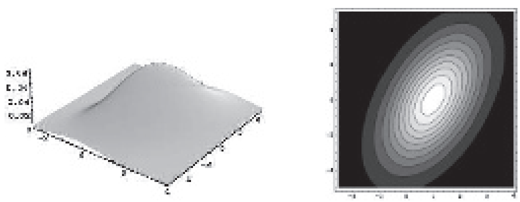
2. *Multi-dimensional normal distribution*:
The random vector $\mathbf{X}$ is a *multivariate normal* (or a multi-dimensional normal) if and only if the linear combination $\mathbf{a}^{\mathrm{T}}\mathbf{X}$ is normal for all vectors $\mathbf{a}$. If the mean of $\mathbf{X}$ is μ, and if the second moment matrix $R = \mathrm{E}\left[(\mathbf{X}-\mu)(\mathbf{X}-\mu)^{\mathrm{T}}\right]$ is non-singular, the density function of $\mathbf{X}$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}\sqrt{\det R}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)^{\mathrm{T}} R^{-1} (\mathbf{x}-\mu)\right]. \tag{7.3.18}$$

Sometimes integrals of the form $I_k = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \left(\mathbf{x}^{\mathrm{T}} M \mathbf{x}\right)^k f(\mathbf{x})\, d\mathbf{x}$ are desired. Defining $a_k = \mathrm{tr}(MR)^k$, we find:

$$\begin{aligned} &I_0 = 1, \qquad I_1 = a_1, \qquad I_2 = a_1^2 + 2a_2, \\ &I_3 = a_1^3 + 6a_1 a_2 + 8a_3, \\ &I_4 = a_1^4 + 12a_1^2 a_2 + 32 a_1 a_3 + 12 a_2^2 + 48 a_4. \end{aligned} \tag{7.3.19}$$

Shown below is Equation (7.3.18) with $\mu = \begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathrm{T}}$ and $R^{-1} = \begin{bmatrix} 1 & 2 \\ 0 & 4 \end{bmatrix}$.



3. *Beta distribution*: If the random variable X has the density function

$$f(x) = B(1+\alpha, 1+\beta)x^{\alpha}(1-x)^{\beta} = \frac{\Gamma(\alpha+\beta+2)}{\Gamma(1+\alpha)\Gamma(1+\beta)}x^{\alpha}(1-x)^{\beta}, \tag{7.3.20}$$

for $0 < x < 1$, where $\alpha > -1$ and $\beta > -1$, then the variable X has a beta distribution.
Properties:

$$\text{Mean} = \mu = \frac{1+\alpha}{2+\alpha+\beta}, \tag{7.3.21}$$
$$\text{Variance} = \sigma^2 = \frac{(1+\alpha)(1+\beta)}{(2+\alpha+\beta)^2(3+\alpha+\beta)},$$
$$r^{\text{th}} \text{ moment about the origin} = \nu_r = \frac{\Gamma(2+\alpha+\beta)\Gamma(1+\alpha+r)}{\Gamma(2+\alpha+\beta+r)\Gamma(1+\alpha)}.$$

4. *Chi-square distribution*: If the random variable X has the density function

$$f(x) = \frac{x^{(n-2)/2}e^{-x/2}}{2^{n/2}\Gamma(n/2)} \qquad \text{for } 0 < x < \infty \tag{7.3.22}$$

then the variable Xhas a chi-square (χ^2) distribution with n degrees of freedom. This is a special case of the gamma distribution.
Properties:

$$\text{Mean} = \mu = n,$$
$$\text{Variance} = \sigma^2 = 2n, \tag{7.3.23}$$
$$\text{Standard deviation} = \sigma = \sqrt{2n}.$$

(a) If $Y_1, Y_2, \ldots, Y_n$ are independent and identically distributed normal random variables with a mean of 0 and a variance of 1, then $\chi^2 = \sum_{i=1}^{n} Y_i^2$ is distributed as chi-square with n degrees of freedom.

(b) If $\chi_1^2, \chi_2^2, \ldots, \chi_k^2$, are independent random variables and have chi-square distributions with $n_1, n_2, \ldots, n_k$ degrees of freedom, then $\sum_{i=1}^{k} \chi_i^2$ has a chi-squared distribution with $n = \sum_{i=1}^{k} n_i$ degrees of freedom.

5. *Exponential distribution*: If the random variable X has the density function

$$f(x) = \frac{e^{-x/\theta}}{\theta}, \qquad \text{for } 0 < x < \infty, \tag{7.3.24}$$

where $\theta > 0$, then the variable X has an exponential distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \theta, \\ \text{Variance} &= \sigma^2 = \theta^2, \\ \text{Standard deviation} &= \sigma = \theta, \\ \text{Moment generating function} &= G(t) = (1-\theta t)^{-1}. \end{aligned} \tag{7.3.25}$$

6. *Gamma distribution*: If the random variable X has the density function

$$f(x) = \frac{1}{\Gamma(1+\alpha)\beta^{1+\alpha}} x^{\alpha} e^{-x/\beta}, \qquad \text{for } 0 < x < \infty, \tag{7.3.26}$$

with $\alpha > -1$ and $\beta > 0$, then the variable X has a gamma distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \beta(1+\alpha), \\ \text{Variance} &= \sigma^2 = \beta^2(1+\alpha), \\ \text{Standard deviation} &= \sigma = \beta\sqrt{1+\alpha} \end{aligned} \tag{7.3.27}$$

$$\text{Moment generating function} = G(t) = (1-\beta t)^{-1-\alpha}, \quad \text{for } t < \beta^{-1}.$$

7. *Rayleigh distribution*: If the random variable X has the density function

$$f(x) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2}{2\sigma^2}\right), \quad \text{for } x \geq 0 \tag{7.3.28}$$

for $\sigma > 0$, then the variable X has a Rayleigh distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \sigma\sqrt{\pi/2} \\ \text{Variance} &= \sigma^2\left(2 - \frac{\pi}{2}\right) \\ \text{Skewness} &= \frac{(\pi-3)\sqrt{\pi/2}}{\left(2-\frac{\pi}{2}\right)^{3/2}} \end{aligned} \tag{7.3.29}$$

8. *Snedecor's F-distribution*: If the random variable X has the density function

$$f(x) = \frac{\Gamma\left(\frac{n+m}{2}\right)\left(\frac{m}{n}\right)^{m/2} x^{(m-2)/2}}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\left(1+\frac{m}{n}x\right)^{(n+m)/2}}, \quad \text{for } 0 < x < \infty, \tag{7.3.30}$$

then the variable X has a F-distribution with m and n degrees of freedom.
Properties:

$$\begin{aligned} \text{Mean} = \mu &= \frac{n}{n-2}, && \text{for } n > 2, \\ \text{Variance} &= \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, && \text{for } n > 4. \end{aligned} \tag{7.3.31}$$

(a) The transformation $w = \dfrac{mx/n}{1+\frac{mx}{n}}$ transforms the F-density to the beta density.

(b) If the random variable X has a χ^2-distribution with m degrees of freedom, the random variable Y has a χ^2-distribution with n degrees of freedom, and X and Y are independent, then $F = \dfrac{X/m}{Y/n}$ is distributed as an F-distribution with m and n degrees of freedom.

9. *Student's t-distribution*: If the random variable X has the density function

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left(1+\frac{x^2}{n}\right)^{(n+1)/2}}, \quad \text{for } -\infty < x < \infty. \tag{7.3.32}$$

then the variable X has a t-distribution with n degrees of freedom.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = 0, \\ \text{Variance} &= \sigma^2 = \frac{n}{n-2}, \qquad \text{for } n > 2. \end{aligned} \tag{7.3.33}$$

(a) If the random variable X is normally distributed with mean 0 and variance σ^2, and if Y^2/σ^2 has a χ^2 distribution with n degrees of freedom, and if X and Y are independent, then $t = \dfrac{X\sqrt{n}}{Y}$ is distributed as a t-distribution with n degrees of freedom.

10. *Uniform distribution*: If the random variable X has the density function

$$f(x) = \frac{1}{\beta - \alpha}, \qquad \text{for } \alpha < x < \beta, \tag{7.3.34}$$

then the variable X has a uniform distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \frac{\alpha+\beta}{2}, \\ \text{Variance} &= \sigma^2 = \frac{(\beta-\alpha)^2}{12}, \\ \text{Standard deviation} &= \sigma = \sqrt{\frac{(\beta-\alpha)^2}{12}}, \\ \text{Moment generating function} &= G(t) = \frac{e^{\beta t} - e^{\alpha t}}{(\beta-\alpha)t} \\ &= \frac{2}{(\beta-\alpha)t} \sinh\left[\frac{(\beta-\alpha)t}{2}\right] e^{(\alpha+\beta)t/2}. \end{aligned} \tag{7.3.35}$$

11. *Weibull distribution*: If the random variable X has the density function

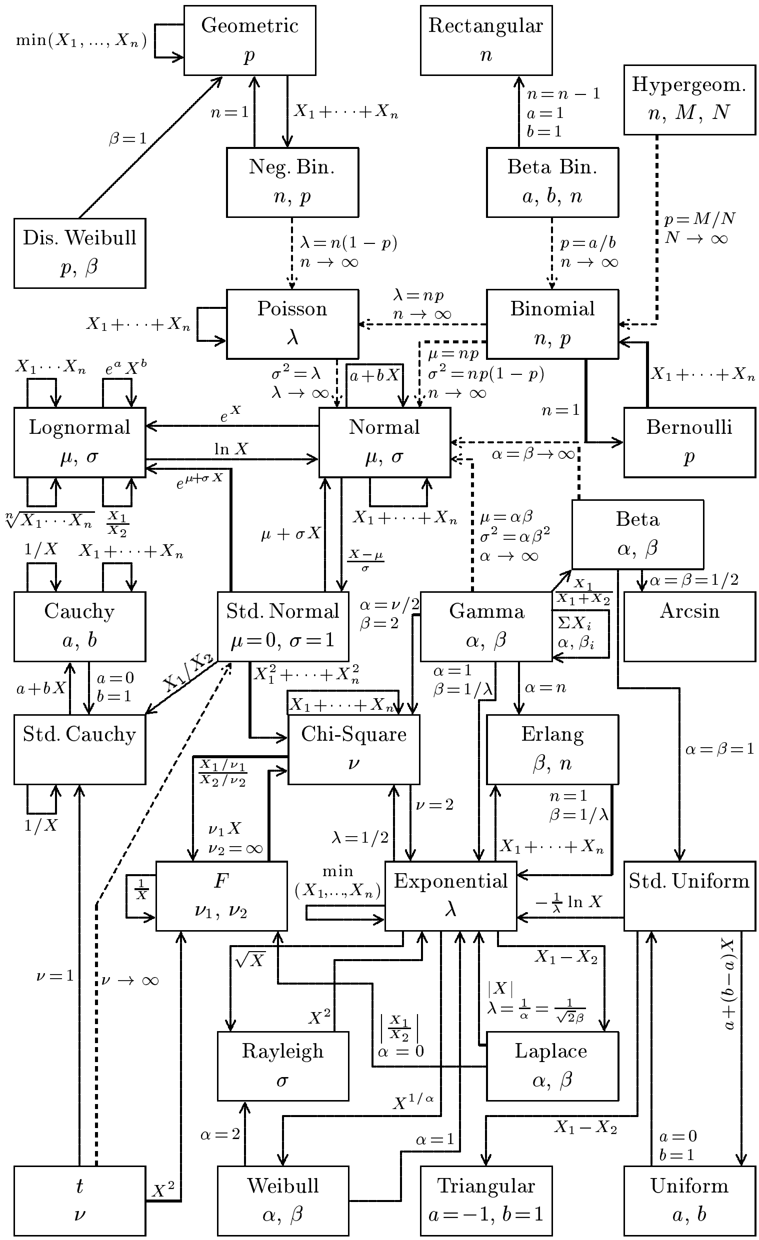
$$f(x) = \frac{\alpha}{\beta^\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad \text{for } x \geq 0 \tag{7.3.36}$$

for $\alpha > 0$ and $\beta > 0$, then the variable X has a Weibull distribution.
Properties:

$$\begin{aligned} \text{Mean} &= \mu = \beta\,\Gamma\left(1 + \frac{1}{\alpha}\right) \\ \text{Variance} &= \sigma^2 = \beta^2\left[\Gamma\left(1+\frac{2}{\alpha}\right) - \Gamma^2\left(1+\frac{1}{\alpha}\right)\right] \end{aligned} \tag{7.3.37}$$

7.3.3 RELATIONSHIPS AMONG DISTRIBUTIONS

Dotted lines represent limiting cases of a distribution. (From D. Zwillinger and S. Kokoska, *Standard Probability and Statistics Tables and Formulae*, Chapman & Hall/CRC, Boca Raton, Florida, 2000. With permission.)



7.3.4 CIRCULAR NORMAL PROBABILITIES

The joint probability density of two independent and normally distributed random variables X and Y (each of mean zero and variance σ^2) is

$$f(x,y) = \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2\sigma^2}\left(x^2+y^2\right)\right] \tag{7.3.38}$$

The probability that a sample value of X and Y is obtained inside a circle C of radius R at a distance d from the origin is $F\left(\frac{d}{\sigma}, \frac{R}{\sigma}\right) = \iint_C f(x,y)\,dx\,dy$.

d/σ	$R/\sigma = 0.2$	0.4	0.6	0.8	1.0	1.5	2.0	2.5	3.0
0.2	0.020	0.077	0.164	0.273	0.392	0.674	0.863	0.955	0.989
0.4	0.019	0.074	0.158	0.264	0.380	0.659	0.852	0.950	0.987
0.6	0.017	0.068	0.147	0.246	0.357	0.631	0.831	0.938	0.982
0.8	0.016	0.061	0.132	0.222	0.326	0.591	0.799	0.920	0.975
1.0	0.013	0.052	0.114	0.194	0.288	0.540	0.756	0.895	0.964
2.0	0.0040	0.016	0.037	0.069	0.111	0.264	0.469	0.676	0.837
3.0	0.0004	0.0018	0.0045	0.0094	0.017	0.060	0.154	0.311	0.510
4.0		0.0001	0.0002	0.0005	0.0011	0.006	0.024	0.072	0.171
5.0						0.0001	0.0008	0.0041	0.017

7.3.5 CIRCULAR ERROR PROBABILITIES

The joint probability density of two independent and normally distributed random variables X and Y, each of mean zero and having variances σ_x^2 and σ_y^2, is

$$f(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left\{\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right\}\right] \tag{7.3.39}$$

The probability that a sample value of X and Y will lie within a circle with center at the origin and radius $K\sigma_x$ is given by $P(K,\sigma_x,\sigma_y) = \iint_R f(x,y)\,dx\,dy$ where R is the region $\sqrt{x^2+y^2} < K\sigma_x$. The following table gives the value of P for various values of K and $c = \sigma_x/\sigma_y$ (for convenience we assume that $\sigma_y \leq \sigma_x$).

K	$c = 0.0$	0.2	0.4	0.6	0.8	1.0
0.2	0.1585	0.0885	0.0482	0.0327	0.0247	0.0198
0.4	0.3108	0.2635	0.1742	0.1238	0.0951	0.0769
0.6	0.4515	0.4256	0.3357	0.2548	0.2010	0.1647
0.8	0.5763	0.5604	0.4942	0.4026	0.3283	0.2739
1.0	0.6827	0.6724	0.6291	0.5461	0.4621	0.3935
1.5	0.8664	0.8628	0.8493	0.8129	0.7490	0.6753
2.0	0.9545	0.9534	0.9494	0.9388	0.9116	0.8647
2.5	0.9876	0.9873	0.9863	0.9838	0.9761	0.9561
3.0	0.9973	0.9972	0.9970	0.9965	0.9949	0.9889
3.5	0.9995	0.9995	0.9995	0.9994	0.9992	0.9978
4.0	0.9999	0.9999	0.9999	0.9999	0.9999	0.9997

7.4 QUEUING THEORY

A queue is represented as $A/B/c/K/m/Z$ where (see Figure 7.2):

1. A and B represent the interarrival times and service times:

   GI general independent interarrival time,
   G general service time distribution,
   H_k k-stage hyperexponential interarrival or service time distribution,
   E_k Erlang-k interarrival or service time distribution,
   M exponential interarrival or service time distribution,
   D deterministic (constant) interarrival or service time distribution.

2. c is the number of identical servers.

3. K is the system capacity.

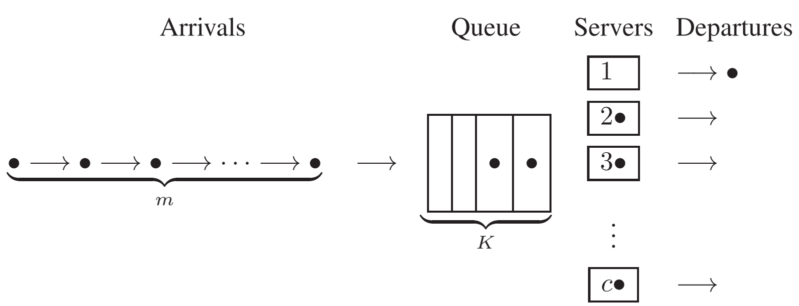
4. m is the number in the source.

5. Z is the queue discipline:

   FCFS first come, first served (also known as FIFO: "first in, first out"),
   LIFO last in, first out,
   RSS random,
   PRI priority service.

When not all variables are present, the trailing ones have the default values, $K = \infty$, $m = \infty$, and Z is FIFO. Note that the m and Z are rarely used.

FIGURE 7.2
Conceptual layout of a queue.



7.4.1 VARIABLES

1. Proportions
   (a) a_n: proportion of customers that find n customers already in the system when they arrive.
   (b) d_n: proportion of customers leaving behind n customers in the system.
   (c) p_n: proportion of time the system contains n customers.
2. Intrinsic queue parameters
   (a) λ: average arrival rate of customers to the system (number per unit time)
   (b) μ: average service rate per server (number per unit time), $\mu = 1/\mathrm{E}\left[T_s\right]$.
   (c) u: traffic intensity, $u = \lambda/\mu$.
   (d) ρ: server utilization, the probability that any particular server is busy, $\rho = u/c = (\lambda/\mu)/c$.
3. Derived queue parameters
   (a) L: average number of customers in the system.
   (b) L_Q: average number of customers in the queue.
   (c) N: number in system.
   (d) W: average time for customer in system.
   (e) W_Q: average time for customer in the queue.
   (f) T_s: service time.
4. Probability functions
   (a) $f_s(x)$: probability density function of customer's service time.
   (b) $f_w(x)$: probability density function of customer's time in system.
   (c) $\pi(z)$: probability generating function of p_n: $\pi_n = \sum_{n=0}^{\infty} p_n z^n$.
   (d) $\pi_Q(z)$: probability generating function of the number in the queue.
   (e) $\alpha(s)$: Laplace transform of $f_w(x)$: $\alpha(s) = \int_0^\infty f_w(x) e^{-xs}\, dx$.
   (f) $\alpha_T(s)$: Laplace transform of the service time.

7.4.2 THEOREMS

1. *Little's law*: $L = \lambda W$ and $L_Q = \lambda W_Q$.
2. If the arrivals have a Poisson distribution: $p_n = a_n$.
3. If customers arrive one at a time and are served one at a time: $a_n = d_n$.
4. For an $M/M/1$ queue with $\rho < 1$,
   (a) $p_n = (1-\rho)\rho^n$,
   (b) $L = \rho/(1-\rho)$,
   (c) $L_Q = \rho^2/(1-\rho)$,
   (d) $W = 1/(\mu - \lambda)$,
   (e) $W_Q = \rho/(\mu - \lambda)$,
   (f) $\pi(z) = (1-\rho)/(1-z\rho)$,
   (g) $\alpha(s) = (\lambda - \mu)/(\lambda - \mu - s)$.

5. For an $M/M/c$ queue with $\rho < 1$ (so that $\mu_n = n\mu$ for $n = 1, 2, \dots, c$ and $\mu_n = c\mu$ for $n > c$),
   (a) $p_0 = \left[\dfrac{u^c}{c!(1-\rho)} + \displaystyle\sum_{n=0}^{c-1} \frac{u^n}{n!} \right]^{-1}$,
   (b) $p_n = \begin{cases} p_0 u^n/n! & \text{for } n = 0, 1, \dots, c, \\ p_0 u^n/c! c^{n-c} & \text{for } n > c, \end{cases}$
   (c) $L_Q = p_0 u^c \rho / c!(1-\rho)^2$,
   (d) $W_Q = L_Q/\lambda$,
   (e) $W = W_Q + 1/\mu$,
   (f) $L = \lambda W$.

6. *Pollaczek–Khintchine formula*: For an $M/G/1$ queue with $\rho < 1$ and $\mathrm{E}\left[T_s^2\right] < \infty$
   (a) $L = L_Q + \rho$,
   (b) $L_Q = \dfrac{\lambda^2 \mathrm{E}\left[T_s^2\right]}{2(1-\rho)}$,
   (c) $W = \lambda L$,
   (d) $W_Q = \lambda L_Q$,
   (e) $\pi(z) = \pi_Q(z)\alpha_T(\lambda - \lambda z)$,
   (f) $\pi_Q(z) = (1-\rho)(1-z)/(\alpha(\lambda - \lambda z) - z)$.

7. For an $M/G/\infty$ queue
   (a) $p_n = e^{-u} u^n / n!$,
   (b) $\pi(z) = \exp\left(-(1-z)u\right)$.

8. *Erlang B formula*: For an $M/G/c/c$ queue, $p_c = \dfrac{u^c}{c!} \left(\displaystyle\sum_{k=0}^{c} \frac{u^k}{k!} \right)^{-1}$.

9. *Distributional form of Little's law*: For any single server system for which: (i) Arrivals are Poisson at rate λ, (ii) all arriving customers enter the system and remain in the system until served (i.e., there is no balking or reneging), (iii) the customers leave the system one at a time in order of arrival, (iv) for any time t, the arrival process after time t and the time in the system for any customer arriving before t are independent, then
   (a) $\pi(z) = \alpha(\lambda(1-z))$,
   (b) $\mathrm{E}\left[L^n\right] = \sum_{k=1}^{n} S(n,k)\mathrm{E}\left[(\lambda W)^k\right]$
   where $S(n,k)$ is Stirling number of the second kind. For example:
      i. $\mathrm{E}[L] = \lambda \mathrm{E}[W]$ (*Little's law*),
      ii. $\mathrm{E}\left[L^2\right] = \mathrm{E}\left[(\lambda W)^2\right] + \mathrm{E}[\lambda W]$.

7.5 MARKOV CHAINS

A *discrete parameter stochastic process* is a collection of random variables $\{X(t), t = 0,1,2,\dots\}$. (The values of t usually represent points in time.) The values of $X(t)$ are the *states* of the process. The collection of states is the *state space*. (The number of states is either finite or countably infinite.) A discrete parameter stochastic process is called a *Markov chain* if, for any set of n time points $t_1 < t_2 < \cdots < t_n$, the conditional distribution of $X(t_n)$ given values for $X(t_1)$, $X(t_2)$, $\dots$, $X(t_{n-1})$ depends only on $X(t_{n-1})$. This can be written

$$P\left[X(t_n) \le x_n \mid X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}\right] = P\left[X(t_n) \le x_n \mid X(t_{n-1}) = x_{n-1}\right]. \quad (7.5.1)$$

A Markov chain is said to be *stationary* if the value of the conditional probability $P\left[X(t_{n+1}) = x_{n+1} | X(t_n) = x_n\right]$ is independent of n. The presentation here is restricted to stationary Markov chains.

7.5.1 TRANSITION FUNCTION AND MATRIX

7.5.1.1 Transition function

Let x and y be states and let $\{t_n\}$ be time points in $T = \{0, 1, 2, \dots\}$. The *transition function*, $P(x, y)$, is defined by

$$P(x, y) = P_{n,n+1}(x, y) = P\left[X(t_{n+1}) = y \mid X(t_n) = x\right], \qquad t_n, t_{n+1} \in T. \quad (7.5.2)$$

$P(x, y)$ is the probability that a Markov chain in state x at time t_n will be in state y at time t_{n+1}. Note that $P(x, y) \ge 0$ and $\sum_y P(x, y) = 1$. The values of $P(x, y)$ are commonly called the *one-step transition probabilities*.

The function $\pi_0(x) = P(X(0) = x)$, with $\pi_0(x) \ge 0$ and $\sum_x \pi_0(x) = 1$ is called the *initial distribution* of the Markov chain. It is the probability distribution when the chain is started. Thus,

$$P\left[X(0) = x_0, X(1) = x_1, \dots, X(n) = x_n\right] = \pi_0(x_0) P_{0,1}(x_0, x_1) P_{1,2}(x_1, x_2) \cdots P_{n-1,n}(x_{n-1}, x_n). \quad (7.5.3)$$

7.5.1.2 Transition matrix

A convenient way to summarize the transition function of a Markov chain is by using the *one-step transition matrix*. It is defined as

$$\mathbf{P} = \begin{bmatrix} P(0,0) & P(0,1) & \dots & P(0,n) & \dots \\ P(1,0) & P(1,1) & \dots & P(1,n) & \dots \\ \vdots & \vdots & \ddots & \vdots & \\ P(n,0) & P(n,1) & \dots & P(n,n) & \dots \\ \vdots & \vdots & & \vdots & \end{bmatrix}. \quad (7.5.4)$$

Define the *n–step transition matrix* by $\mathbf{P}^{(n)}$ as the matrix with entries

$$P^{(n)}(x,y) = P\left[X(t_{m+n}) = y \mid X(t_m) = x\right]. \tag{7.5.5}$$

This can be written in terms of the one-step transition matrix as $\mathbf{P}^{(n)} = \mathbf{P}^n$.

Suppose the state space is finite. The one-step transition matrix is said to be *regular* if, for some positive power m, all of the elements of $\mathbf{P}^m$ are strictly positive.

THEOREM 7.5.1 *(Chapman–Kolmogorov equation)*
Let $P(x,y)$ be the one-step transition function of a Markov chain and define $P^{(0)}(x,y)$ = 1, if x = y, and 0, otherwise. Then, for any pair of non-negative integers, s and t, such that $s+t=n$,

$$P^{(n)}(x,y) = \sum_z P^{(s)}(x,z)P^{(t)}(z,y). \tag{7.5.6}$$

7.5.2 RECURRENCE

Define the probability that a Markov chain starting in state x returns to state x for the first time after n steps by

$$f^n(x,x) = P\left[X(t_n) = x, X(t_{n-1}) \neq x, \dots, X(t_1) \neq x \mid X(t_0) = x\right]. \tag{7.5.7}$$

It follows that $P^n(x,x) = \sum_{k=1}^{n} f^k(x,x)P^{n-k}(x,x)$. A state x is said to be *recurrent* if $\sum_{n=1}^{\infty} f^n(x,x) = 1$. This means that a state x is recurrent if, after starting in x, the probability of returning to it after some finite length of time is one. A state which is not recurrent is said to be *transient*.

THEOREM 7.5.2
A state x of a Markov chain is recurrent if and only if $\sum_{n=1}^{\infty} P^n(x,x) = \infty$.

Two states, x and y, are said to *communicate* if, for some $n > 0$, $P^n(x,y) > 0$. This theorem implies that, if x is a recurrent state and x communicates with y, y is also a recurrent state. A Markov chain is said to be *irreducible* if every state communicates with every other state and with itself.

Let x be a recurrent state and define T_x the (*return time*) as the number of stages for a Markov chain to return to state x, having begun there. A recurrent state x is said to be *null recurrent* if $\mathrm{E}\left[T_x\right] = \infty$. A recurrent state that is not null recurrent is said to be *positive recurrent*.

7.5.3 STATIONARY DISTRIBUTIONS

Let $\{X(t), t = 0, 1, 2, \dots\}$ be a Markov chain having a one-step transition function $P(x,y)$. A function $\pi(x)$ where each $\pi(x)$ is non-negative, $\sum_x \pi(x)P(x,y) = \pi(y)$, and $\sum_y \pi(y) = 1$, is called a *stationary distribution*. If a Markov chain has a stationary distribution and $\lim_{n\to\infty} P^n(x,y) = \pi(y)$ for every x, then, regardless of the initial distribution, $\pi_0(x)$, the distribution of $X(t_n)$ approaches $\pi(x)$ as n tends to infinity. When this happens, $\pi(x)$ is often referred to as the *steady state distribution*. The following categorizes those Markov chains with stationary distributions.

THEOREM 7.5.3

Let X_P denote the set of positive recurrent states of a Markov chain.

1. *If X_P is empty, the chain has no stationary distribution.*
2. *If X_P is a non-empty irreducible set, the chain has a unique stationary distribution.*
3. *If X_P is non-empty but not irreducible, the chain has an infinite number of distinct stationary distributions.*

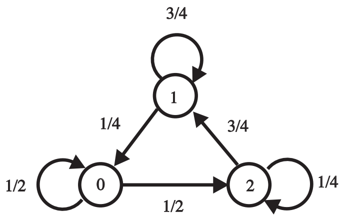
The *period* of a state x is denoted by $d(x)$ and is defined as the greatest common divisor of all integers, $n \geq 1$, for which $P^n(x, x) > 0$. If $P^n(x, x) = 0$ for all $n \geq 1$, then define $d(x) = 0$. If each state of a Markov chain has $d(x) = 1$, the chain is said to be *aperiodic*. If each state has period $d > 1$, the chain is said to be *periodic* with period d. The vast majority of Markov chains encountered in practice are aperiodic. An irreducible, positive recurrent, aperiodic Markov chain always possesses a steady-state distribution. An important special case occurs when the state space is finite. Suppose that $X = \{1, 2, \ldots, K\}$. Let $\pi_0 = \{\pi_0(1), \pi_0(2), \ldots, \pi_0(K)\}$.

THEOREM 7.5.4

Let $\boldsymbol{P}$ be a regular one-step transition matrix and π_0 be an arbitrary vector of initial probabilities. Then $\lim_{n\to\infty} \pi_0(x)\boldsymbol{P}^n = \boldsymbol{y}$, where $\boldsymbol{yP} = \boldsymbol{y}$, and $\sum_{i=1}^{K} \pi_0(y_i) = 1$.

7.5.3.1 Example: A simple three-state Markov chain

A Markov chain having three states $\{0, 1, 2\}$ is shown



This Markov chain has the one-step and two-step transition matrices:

$$P = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 1/4 & 3/4 & 0 \\ 0 & 3/4 & 1/4 \end{bmatrix} \qquad P^{(2)} = P^2 = \begin{bmatrix} 1/4 & 3/8 & 3/8 \\ 5/16 & 9/16 & 1/8 \\ 3/16 & 3/4 & 1/16 \end{bmatrix}.$$

The one-step transition matrix is regular. This Markov chain is irreducible, and all three states are recurrent. In addition, all three states are positive recurrent. Since all states have period 1, the chain is aperiodic. The unique steady state distribution is $\pi(0) = 3/11$, $\pi(1) = 6/11$, and $\pi(2) = 2/11$.

7.5.4 RANDOM WALKS

Let $\eta(t_1), \eta(t_2), \ldots$ be independent random variables having a common density $f(x)$, and let $t_1, t_2, \ldots$ be integers. Let X_0 be an integer–valued random variable that is independent of $\eta(t_1)$, $\eta(t_2)$, $\ldots$, and $X(t_n) = X_0 + \sum_{i=1}^{n} \eta(t_i)$. The sequence $\{X(t_i), i = 0, 1, \ldots\}$ is called a *random walk*. An important special case is a *simple random walk*. It is defined by

$$P(x,y) = \begin{cases} p, & \text{if } y = x-1, \\ r, & \text{if } y = x, \\ q, & \text{if } y = x+1, \end{cases} \qquad \text{where } p+q+r=1, \text{ and } P(0,0)=p+r. \tag{7.5.8}$$

Here, an object begins at a certain point in a lattice and at each step either stays at that point or moves to a neighboring lattice point.

This one-dimensional random walk can be extended to higher-dimensional lattices. A common case is that an object can only transition to an adjacent lattice point, and all such transitions are equally likely. In this case

1. In a one- or two-dimensional lattice, a random walk will return to its starting point with probability 1.
2. In a three-dimensional lattice, the probability that a random walk will return to its starting point is $P = 0.3405$.
3. In a d-dimensional lattice (with $d \geq 3$), the probability that a random walk will return to its starting point is $P = 1 - \dfrac{1}{u_d}$ with $u_d = \displaystyle\int_0^\infty e^{-t}\left[I_0\left(\frac{t}{d}\right)\right]^d\,dt.$

7.6 RANDOM NUMBER GENERATION

7.6.1 METHODS OF PSEUDORANDOM NUMBER GENERATION

In Monte Carlo applications, and other computational situations where randomness is required, random numbers must be used. While numbers measured from a physical process known to be random have been used, it is much more practical to use recursions that produce numbers that behave as random in applications and with respect to statistical tests of randomness. These are called *pseudorandom numbers* and are produced by a pseudorandom number generator (PRNG). Depending on the application, either integers in some range or floating point numbers in $[0,1)$ are the output from a PRNG. Since most PRNGs use integer recursions, a conversion into integers in a desired range or into a floating point number in $[0,1)$ is required. If x_n is an integer produced by some PRNG in the range $0 \leq x_n \leq M-1$, then an integer in the range $0 \leq x_n \leq N-1$, with $N \leq M$, is given by $y_n = \left\lfloor \frac{N}{M} x_n \right\rfloor$. If $N \ll M$, then $y_n = x_n \pmod N$ may be used. Alternately, if a floating point value in $[0,1)$ is desired, let $y_n = x_n/M$.

7.6.1.1 Linear congruential generators

Perhaps the oldest generator still in use is the *linear congruential generator* (LCG). The underlying integer recursion for LCGs is

$$x_n = ax_{n-1} + b \pmod M. \tag{7.6.1}$$

Equation (7.6.1) defines a periodic sequence of integers modulo M starting with x_0, the initial seed. The constants of the recursion are referred to as the *modulus* M, *multiplier* a, and *additive constant* b. If $M = 2^m$, a very efficient implementation is possible. Alternately, there are theoretical reasons why choosing M prime is optimal. Hence, the only moduli that are used in practical implementations are $M = 2^m$ or the prime $M = 2^p - 1$ (i.e., M is a Mersenne prime). With a Mersenne prime or any modulus "close to" 2^p, modular multiplication can be implemented at about twice the computational cost of multiplication modulo 2^p.

Equation (7.6.1) yields a sequence $\{x_n\}$ whose period, denoted $\text{Per}(x_n)$, depends on M, a, and b. The values of the maximal period for the three most common cases used and the conditions required to obtain them are

a	b	M	$\text{Per}(x_n)$
Primitive root of M	Anything	Prime	$M - 1$
3 or 5 $\pmod 8$	0	2^m	2^{m-2}
$1 \pmod 4$	$1 \pmod 2$	2^m	2^m

A major shortcoming of LCGs modulo a power-of-two compared with prime modulus LCGs derives from the following theorem for LCGs:

THEOREM 7.6.1

Define the following LCG sequence: $x_n = ax_{n-1} + b \pmod{M_1}$. *If* M_2 *divides* M_1 *then* $y_n = x_n \pmod{M_2}$ *satisfies* $y_n = ay_{n-1} + b \pmod{M_2}$.

Theorem 7.6.1 implies that the k least-significant bits of any power-of-two modulus LCG with $\text{Per}(x_n) = 2^m = M$ has $\text{Per}(y_n) = 2^k$, $0 < k \leq m$. Since a long period is crucial in PRNGs, when these types of LCGs are employed in a manner that makes use of only a few least-significant-bits, their quality may be compromised. When M is prime, no such problem arises.

Since LCGs are in such common usage, below is a list of parameter values in the literature. The Park–Miller LCG is widely considered a minimally acceptable PRNG. Using any values other than those in the following table may result in a "weaker" LCG.

a	b	M	Source
7^5	0	$2^{31} - 1$	Park–Miller
131	0	2^{35}	Neave
16333	25887	2^{15}	Oakenfull

7.6.1.2 Shift-register generators

Another popular method of generating pseudorandom numbers is using binary shift-register sequences to produce pseudorandom bits. A *binary shift-register sequence* (SRS) is defined by a binary recursion of the type,

$$x_n = x_{n-j_1} \oplus x_{n-j_2} \oplus \cdots \oplus x_{n-j_k}, \qquad j_1 < j_2 < \cdots < j_k = \ell, \qquad (7.6.2)$$

where $\oplus$ is the exclusive "or" operation. Note that $x \oplus y \equiv x + y \pmod 2$. Thus the new bit, x_n, is produced by adding k previously computed bits together modulo 2. The implementation of this recurrence requires keeping the last ℓ bits from the sequence in a shift register, hence the name. The longest possible period is equal to the number of non-zero ℓ-dimensional binary vectors, namely $2^\ell - 1$.

A sufficient condition for achieving $\mathrm{Per}(x_n) = 2^\ell - 1$ is that the characteristic polynomial, corresponding to Equation (7.6.2), be primitive modulo 2. Since primitive trinomials of nearly all degrees of interest have been found, SRSs are usually implemented using two-term recursions of the form,

$$x_n = x_{n-k} \oplus x_{n-\ell}, \qquad 0 < k < \ell. \tag{7.6.3}$$

In these two-term recursions, k is the *lag* and ℓ is the register length. Proper choice of the pair (ℓ, k) leads to SRSs with $\mathrm{Per}(x_n) = 2^\ell - 1$. Here is a list with suitable (ℓ, k) pairs:

Primitive trinomial exponents					
(5,2)	(7,1)	(7,3)	(17,3)	(17,5)	(17,6)
(31,3)	(31,6)	(31,7)	(31,13)	(127,1)	(521,32)

7.6.1.3 Lagged-Fibonacci generators

Another way of producing pseudorandom numbers uses lagged-Fibonacci generators. The term "lagged-Fibonacci" refers to two-term recurrences of the form,

$$x_n = x_{n-k} \diamond x_{n-\ell}, \qquad 0 < k < \ell, \tag{7.6.4}$$

where $\diamond$ refers to one of the three common methods of combination: (1) addition modulo 2^m, (2) multiplication modulo 2^m, or (3) bitwise exclusive 'OR'ing of m-long bit vectors. Combination method (3) can be thought of as a special implementation of a two-term shift-register sequence.

Using combination method (1) leads to *additive lagged-Fibonacci sequences* (ALFSs). If x_n is given by

$$x_n = x_{n-k} + x_{n-\ell} \pmod{2^m}, \qquad 0 < k < \ell, \tag{7.6.5}$$

then the maximal period is $\mathrm{Per}(x_n) = (2^\ell - 1)2^{m-1}$.

ALFSs are especially suitable for producing floating point deviates using the real-valued recursion $y_n = y_{n-k} + y_{n-\ell} \pmod 1$. This circumvents the need to convert from integers to floating point values and allows floating point hardware to be used. One caution with ALFSs is that Theorem 7.6.1 holds, and so the low-order bits have periods that are shorter than the maximal period. However, this is not nearly the problem as in the LCG case. With ALFSs, the j least-significant bits will have period $(2^\ell - 1)2^{j-1}$, so, if ℓ is large, there really is no problem. Note that one can use the table of primitive trinomial exponents to find (ℓ, k) pairs that give maximal period ALFSs.

7.6.1.4 Non-linear generators

Also possible for PRNGs are non-linear integer recurrences. For example, if in equation (7.6.4) "$\diamond$" referred to multiplication modulo 2^m, then this recurrence would be a *multiplicative lagged-Fibonacci generator* (MLFG); a non-linear generator. The mathematical structure of non-linear generators is qualitatively different than that of linear generators. Thus, their defects and deficiencies are thought to be complementary to their linear counterparts.

The maximal period of a MLFG is $\text{Per}(x_n) = (2^\ell - 1)2^{m-3}$, a factor of 4 shorter than the corresponding ALFS. However, there are benefits to using multiplication as the combining function due to the bit mixing achieved. Because of this, the perceived quality of the MLFG is considered superior to an ALFS with the same lag, ℓ.

We now illustrate two non-linear generators, the *inversive congruential generators* (ICGs), which were designed as non-linear analogs of the LCG.

1. The *implicit ICG* is defined by the following recurrence that is almost that of an LCG (assume that M is a prime)

$$x_n = a\overline{x_{n-1}} + b \pmod{M}. \tag{7.6.6}$$

   The difference is that we must also take the multiplicative inverse of x_{n-1}, which is defined by $\overline{x_{n-1}}\, x_{n-1} \equiv 1 \pmod{M}$, and $\overline{0} = 0$. This recurrence is indeed non-linear, and avoids some of the problems inherent in linear recurrences, such as the fact that linear tuples must lie on hyperplanes.

2. The *explicit ICG* is

$$x_n = \overline{an + b} \pmod{M}. \tag{7.6.7}$$

One drawback of ICGs is the cost of inversion, which is $O(\log_2 M)$ times the cost of multiplication modulo M.

7.6.2 GENERATING NON-UNIFORM RANDOM VARIABLES

Suppose we want deviates from a distribution with probability density function $f(x)$ and distribution function $F(x) = \int_{-\infty}^{x} f(u)\, du$. In the following "$y$ is $U[0,1)$" means y is uniformly distributed on $[0,1)$.

Two general techniques for converting uniform random variables into those from other distributions are as follows:

1. The *inverse transform method*:
   If y is $U[0,1)$, then the random variable $F^{-1}(y)$ will have its density equal to $f(x)$. (Note that $F^{-1}(y)$ exists since $0 \leq F(x) \leq 1$.)
2. The *acceptance-rejection method*:
   Suppose the density can be written as $f(x) = C\,h(x)g(x)$ where $h(x)$ is the density of a computable random variable, the function g satisfies $0 < g(x) \leq 1$, and $C^{-1} = \int_{-\infty}^{\infty} h(u)g(u)\, du$ is a normalization constant. If x is $U[0,1)$, y has density $h(x)$, and if $x < g(y)$, then x has density $f(x)$. Thus one generates $\{x, y\}$ pairs, rejecting both if $x \geq g(y)$ and returning x if $x < g(y)$.

EXAMPLES

1. Examples of the inverse transform method:

   (a) *Exponential distribution*: The exponential distribution with rate λ has $f(x) = \lambda e^{-\lambda x}$ (for $x \geq 0$) and $F(x) = 1 - e^{-\lambda x}$. Thus $u = F(x)$ can be solved to give $x = F^{-1}(u) = -\lambda^{-1}\ln(1-u)$. If u is $U[0,1)$, then so is $1-u$. Hence $x = -\lambda^{-1}\ln u$ is exponentially distributed with rate λ.

(b) *Normal distribution*: Suppose the z_i's are normally distributed with density function $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$. The polar transformation then gives random variables $r = \sqrt{z_1^2 + z_2^2}$ (exponentially distributed with $\lambda = 2$) and $\theta = \tan^{-1}(z_2/z_1)$ (uniformly distributed on $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$). Inverting these relationships results in $z_1 = \sqrt{-2\ln x_1}\cos 2\pi x_2$ and $z_2 = \sqrt{-2\ln x_1}\sin 2\pi x_2$; each is normally distributed when x_1 and x_2 are $U[0, 1)$. (This is the *Box–Muller technique*.)

2. Examples of the rejection method:

(a) *Exponential distribution with* $\lambda = 1$:

i. Generate random numbers $\{U_i\}_{i=1}^{N}$ uniformly in $[0, 1]$), stopping at $N = \min\{n \mid U_1 \geq \cdots \geq U_{n-1} < U_n\}$.
ii. If N is even, accept that run, and go to step iii. If N is odd reject the run, and return to step i.
iii. Set X equal to the number of failed runs plus U_1 (the first random number in the successful run).

(b) *Normal distribution* - using uniform random variables:

i. Select two random variables (V_1, V_2) from $U[0, 1)$. Form $R = V_1^2 + V_2^2$.
ii. If $R > 1$, then reject the (V_1, V_2) pair, and select another pair.
iii. If $R < 1$, then $x = V_1\sqrt{-2\frac{\ln R}{R}}$ has a $N(0, 1)$ distribution.

(c) *Normal distribution* - using exponential random variables:

i. Select two exponentially distributed random variables with rate 1: (V_1, V_2).
ii. If $V_2 \geq (V_1 - 1)^2/2$, then reject the (V_1, V_2) pair, and select another pair.
iii. Otherwise, V_1 has a $N(0, 1)$ distribution.

(d) *Cauchy distribution*:
To generate values of X from $f(x) = \frac{1}{\pi(1+x^2)}$ on $-\infty < x < \infty$,

i. Generate random numbers U_1, U_2 (uniform on $[0, 1)$), and set $Y_1 = U_1 - \frac{1}{2}$, $Y_2 = U_2 - \frac{1}{2}$.
ii. If $Y_1^2 + Y_2^2 \leq \frac{1}{4}$, then return $X = Y_1/Y_2$. Otherwise return to step (a).

To generate values of X from a Cauchy distribution with parameters β and θ, $\left(f(x) = \frac{\beta}{\pi\left[\beta^2 + (x-\theta)^2\right]} \text{ for } -\infty < x < \infty \right)$, construct X as above, and then use $\beta X + \theta$.

7.6.2.1 Discrete random variables

The density function of a discrete random variable that attains finitely many values can be represented as a vector $\mathbf{p} = (p_0, p_1, \ldots, p_{n-1}, p_n)$ by defining the probabilities $P(x = j) = p_j$ (for $j = 0, \ldots, n$). The distribution function can be defined by the vector $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}, 1)$, where $c_j = \sum_{i=0}^{j} p_i$. Given this representation of $F(x)$, we can apply the inverse transform by computing x to be $U[0, 1)$, and then finding the index j so that $c_j \leq x < c_{j+1}$. In this case event j will have occurred. Examples:

1. (Binomial distribution) The binomial distribution with n trials of mean p has $p_j = \binom{n}{j} p^j (1-p)^{n-j}$, for $j = 0, \ldots, n$.

   (a) As an example, consider the result of flipping a fair coin. In 2 flips, the probability of obtaining $(0, 1, 2)$ heads is $\mathbf{p} = (\frac{1}{4}, \frac{1}{2}, \frac{1}{4})$. Hence $\mathbf{c} = (\frac{1}{4}, \frac{3}{4}, 1)$. If x (chosen from $U[0, 1)$) turns out to be say, 0.4, then "1 head" is returned (since $\frac{1}{4} \leq 0.4 < \frac{3}{4}$).

   (b) Note that, when n is large, it is costly to compute the density and distribution vectors. When n is large and relatively few binomially distributed pseudorandom numbers are desired, an alternative is to use the normal approximation to the binomial.

   (c) Alternately, one can form the sum $\sum_{i=1}^{n} \lfloor u_i + p \rfloor$, where each u_i is $U[0, 1)$.

2. (Geometric distribution) To simulate a value from $P(X = i) = p(1-p)^{i-1}$ for $i \geq 1$, use $X = 1 + \left\lceil \dfrac{\log U}{\log(1-p)} \right\rceil$.

3. (Poisson distribution) The Poisson distribution with mean λ has $p_j = \lambda^j e^{-\lambda}/j!$ for $j \geq 0$. The Poisson distribution counts the number of events in a unit time interval if the times are exponentially distributed with rate λ. Thus, if the times t_i are exponentially distributed with rate λ, then j will be Poisson distributed with mean λ when $\sum_{i=0}^{j} t_i \leq 1 < \sum_{i=0}^{j+1} t_i$. Since $t_i = -\lambda^{-1} \ln u_i$, where u_i is $U[0, 1)$, the previous equation may be written as $\prod_{i=0}^{j} u_i \geq e^{-\lambda} > \prod_{i=0}^{j+1} u_i$. This allows us to compute Poisson random variables by iteratively computing $P_j = \prod_{i=0}^{j} u_i$ until $P_j < e^{-\lambda}$. The first such j that makes this inequality true will have the desired distribution.

Random variables can be simulated using the following table (each U and U_i is uniform on the interval $[0, 1)$):

Distribution	Density	Formula for deviate
Binomial	$p_j = \binom{n}{j} p^j (1-p)^{n-j}$	$\sum_{i=1}^{n} \lfloor U_i + p \rfloor$
Cauchy	$f(x) = \dfrac{\sigma}{\pi(x^2 + \sigma^2)}$	$\sigma \tan(\pi U)$
Exponential	$f(x) = \lambda e^{-\lambda x}$	$-\lambda^{-1} \ln U$
Pareto	$f(x) = ab^a / x^{a+1}$	$b/U^{1/a}$
Rayleigh	$f(x) = x/\sigma e^{-x^2/2\sigma^2}$	$\sigma\sqrt{-\ln U}$

7.7 RANDOM MATRICES

A random matrix is a matrix whose entries are from a specific distribution.

Let M_n be an $n \times n$ matrix where each (m_{ij}) independently comes from a standard normal distribution (i.e., every element has mean zero and variance one).

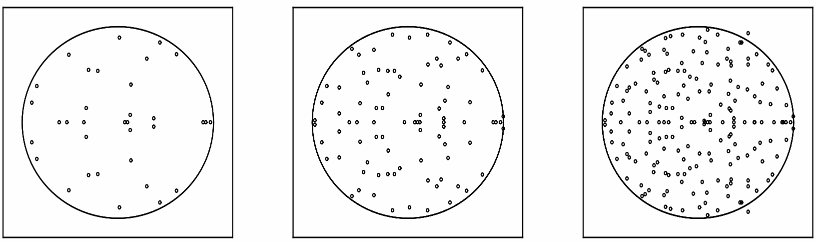
1. The expected number of real eigenvalues (E_n) of the matrix M_n is:

$$E_n = \begin{cases} \sqrt{2} \displaystyle\sum_{k=0}^{n/2-1} \frac{(4k-1)!!}{(4k)!!} & \text{when } n \text{ is even} \\ 1 + \sqrt{2} \displaystyle\sum_{k=0}^{(n-1)/2} \frac{(4k-3)!!}{(4k-2)!!} & \text{when } n \text{ is odd} \end{cases} \tag{7.7.1}$$
$$\sim \sqrt{\frac{2n}{\pi}} \qquad \text{as } n \to \infty$$

2. The expected number of complex eigenvalues (c_n) of the matrix M_n is $c_n = n - E_n$ with $c_1 = 0$, $c_2 = 2 - \sqrt{2}$, $c_3 = 2 - \frac{1}{2}\sqrt{2}$, and $c_4 = 4 - \frac{11}{8}\sqrt{2}$.
3. Let $p_{n,k}$ be the probability that M_n has exactly k real eigenvalues. Then $p_{n,n} = 2^{-n(n-1)/4}$. Formulae are also available for the other values of $\{p_{n,k}\}$.
4. Let $\rho_n(x, y)$ be the probability density that M_n has $x \pm iy$ as eigenvalues. Then

$$\rho_n(x,y) = \sqrt{\frac{2}{\pi}}\, e^{2y^2} y \operatorname{erfc}\left(\sqrt{2}y\right) \frac{\Gamma(n-1, x^2+y^2)}{\Gamma(n-1)} \tag{7.7.2}$$

5. Consider the eigenvalues $\{\lambda\}$ of M_n. *Girko's circular law* states that $\frac{\lambda}{n}$ is uniformly distributed in the unit circle as $n \to \infty$. Below are images of $\frac{\lambda}{n}$ for samples with n equal to 20, 50, and 100.



Wigner's Semicircle law: Let R_n be an n-by-n real symmetric matrix with random elements $\{r_{ij} \mid i \leq j \leq n\}$. Assume the random elements of R_n are identically distributed, have second moment equal to m^2, and all the moments are bounded independent of n. Let $S_{\alpha,\beta}(n)$ be the number of eigenvalues of R_n that are in the real interval $(\alpha\sqrt{n}, \beta\sqrt{n})$, with $\alpha < \beta$. Then the expected value of $S_{\alpha,\beta}(n)$ satisfies

$$\lim_{n\to\infty} \frac{\mathrm{E}\left[S_{\alpha,\beta}(n)\right]}{n} = \frac{1}{2\pi m^2} \int_\alpha^\beta \sqrt{4m^2 - x^2}\, dx \tag{7.7.3}$$

7.8 CONTROL CHARTS AND RELIABILITY

7.8.1 CONTROL CHARTS

Control charts are graphical tools used to assess and maintain the stability of a process. They are used to separate random variation from specific causes. Data measurements are plotted versus time along with upper and lower control limits (UCL and LCL) and a center line. If the process is in control and the underlying distribution is normal, then the control limits represent three standard deviations from the center line (mean).

If all of the data points are contained within the control limits, the process is considered stable and the mean and standard deviations can be reliably calculated. The variations between data points occur from random causes. Data outside the control limits or forming abnormal patterns point to unstable, out-of-control processes.

There are two types of control charts:

- Control charts for attributes: the data are from a count. If every item is either "good" or "bad," then "defectives" are counted. If each item may have several flaws, then "defects" are counted.
  - p chart (defectives, sample size varies, uses Binomial statistics)
  - np chart (defectives, sample size fixed, uses Binomial statistics)
  - u chart (defects, sample size varies, uses Poisson statistics)
  - c chart (defects, sample size varies, uses Poisson statistics)

- Control charts for variables: the data are from measurements on a variable or continuous scale. Statistics of the measurements are used, Gaussian statistics used.
  - $\overline{x} - R$ chart (sample average versus sample range)
  - $\tilde{x} - R$ chart (sample median versus sample range)
  - $\overline{x} - s$ chart (sample average versus sample standard deviation)
  - $x - Rs$ chart (samples versus moving range, $Rs = |x_i - x_{i-1}|$). Similar to a $\overline{x} - R$ chart but single measurements are made. Used when measurements are expensive or dispersion of measured values is small.

In the tables, k denotes the number of samples taken, i is an index for the samples ($i = 1 \ldots k$), n is the sample size (number of elements in each sample), and R is the range of the values in a sample (minimum element value subtracted from the maximum element value). The mean is μ and the standard deviation is σ.

Types of control charts and limits ("P" stands for parameters)

Chart	(μ, σ) known?	P	Centerline	UCL	LCL
$\overline{x} - R$	No	$\overline{x}$	$\overline{\overline{x}} = \frac{\sum \overline{x}}{k}$	$\overline{\overline{x}} + A_2\overline{R}$	$\overline{\overline{x}} - A_2\overline{R}$
$\overline{x} - R$	No	R	$\overline{R} = \frac{\sum R}{k}$	$D_4\overline{R}$	$D_3\overline{R}$
$\overline{x} - R$	Yes	$\overline{x}$	$\overline{\overline{x}} = \mu$	$\mu + \frac{3\sigma}{\sqrt{n}}$	$\mu - \frac{3\sigma}{\sqrt{n}}$
$\overline{x} - R$	Yes	R	$\overline{R} = d_2\sigma$	$D_2\sigma$	$D_1\sigma$
$\tilde{x} - R$	No	$\tilde{x}$	$\overline{\tilde{x}} = \frac{\sum \tilde{x}}{k}$	$\tilde{x} + m_3A_2$	$\tilde{x} - m_3A_2$
$\tilde{x} - R$	No	R	$\overline{R} = \frac{\sum R}{k}$	$D_4\overline{R}$	$D_3\overline{R}$
$x - Rs$	No	x	$\overline{x} = \frac{\sum x}{k}$	$\overline{x} + 2.66\overline{R}s$	$\overline{x} - 2.66\overline{R}s$
$x - Rs$	No	Rs	$\overline{R}s = \frac{\sum \overline{R}s}{k}$	$3.27\overline{R}s$	—
pn	No	pn	$\overline{p}n = \frac{\sum pn}{k}$	$\overline{p}n + \sqrt{\overline{p}n(1-\overline{p})}$	$\overline{p}n - \sqrt{\overline{p}n(1-\overline{p})}$
p	No	p	$\overline{p} = \frac{\sum pn}{\sum n}$	$\overline{p}n + 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$	$\overline{p}n - 3\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$
c	No	c	$\overline{c} = \frac{\sum c}{k}$	$\overline{c} + 3\sqrt{\overline{c}}$	$\overline{c} - 3\sqrt{\overline{c}}$
u	No	u	$\overline{u} = \frac{\sum c}{\sum n}$	$\overline{u} + 3\sqrt{\frac{\overline{u}}{n}}$	$\overline{u} - 3\sqrt{\frac{\overline{u}}{n}}$

Sample size n	A_2	d_2	D_1	D_2	D_3	D_4	m_3	m_3A_2
2	1.880	1.128	0	3.686	–	3.267	1.000	1.880
3	1.023	1.693	0	4.358	–	2.575	1.160	1.187
4	0.729	2.059	0	4.698	–	2.282	1.092	0.796
5	0.577	2.326	0	4.918	–	2.115	1.198	0.691
6	0.483	2.534	0	5.078	–	2.004	1.135	0.549
7	0.419	2.704	0.205	5.203	0.076	1.924	1.214	0.509
8	0.373	2.847	0.387	5.307	0.136	1.864	1.160	0.432
9	0.337	2.970	0.546	5.394	0.184	1.816	1.223	0.412
10	0.308	3.078	0.687	5.469	0.223	1.777	1.176	0.363
11	0.285	3.173	0.812	5.534	0.256	1.744		
12	0.266	3.258	0.924	5.592	0.284	1.716		
13	0.249	3.336	1.026	5.646	0.308	1.692		
14	0.235	3.407	1.121	5.693	0.329	1.671		
15	0.223	3.472	1.207	5.737	0.348	1.652		
16	0.212	3.532	1.285	5.779	0.364	1.636		
17	0.203	3.588	1.359	5.817	0.379	1.621		
18	0.194	3.640	1.426	5.854	0.392	1.608		
19	0.187	3.689	1.490	5.888	0.404	1.596		
20	0.180	3.735	1.548	5.922	0.414	1.586		
25	0.153	3.931	1.806	6.056	0.459	1.541		

Abnormal distributions of points in control charts

Abnormality	Description
Sequence	Seven or more consecutive points on one side of the center line. Denotes the average value has shifted.
Bias	Fewer than seven consecutive points on one side of the center line, but most of the points are on that side. • 10 of 11 consecutive points • 12 or more of 14 consecutive points • 14 or more of 17 consecutive points • 16 or more of 20 consecutive points
Trend	Seven or more consecutive rising or falling points.
Approaching the limit	Two out of three or three or more out of seven consecutive points are more than two-thirds the distance from the center line to a control limit.
Periodicity	The data points vary in a regular periodic pattern.

7.8.2 SIGMA CONVERSION TABLE

Defects per million opportunities	Sigma level (with 1.5 sigma shift)	C_{pk} (Sigma level/3) with 1.5 sigma shift
933,200	0.000	0.000
500,000	1.500	0.500
66,800	3.000	1.000
6,210	4.000	1.333
233	5.000	1.667
3.4	6.000	2.000

7.8.3 RELIABILITY

1. The *reliability* of a product is the probability that the product will function within specified limits for at least a specified period of time.
2. A *series system* is one in which the entire system will fail if any of its components fail.
3. A *parallel system* is one in which the entire system will fail only if all of its components fail.
4. Let R_i denote the reliability of the i^{th} component.
5. Let R_s denote the reliability of a series system.
6. Let R_p denote the reliability of a parallel system.

The *product law of reliabilities* states $R_s = \prod_{i=1}^{n} R_i$.

The *product law of unreliabilities* states $R_p = 1 - \prod_{i=1}^{n}(1 - R_i)$.

7.8.4 FAILURE TIME DISTRIBUTIONS

1. Let the probability of an item failing between times t and $t + \Delta t$ be $f(t)\Delta t + o(\Delta t)$ as $\Delta t \to 0$.
2. The probability that an item will fail in the interval from 0 to t is

$$F(t) = \int_0^t f(x)\, dx. \tag{7.8.1}$$

3. The *reliability function* is the probability that an item survives to time t

$$R(t) = 1 - F(t). \tag{7.8.2}$$

4. The *instantaneous hazard rate*, $Z(t)$, is approximately the probability of failure in the interval from t to $t + \Delta t$, given that the item survived to time t

$$Z(t) = \frac{f(t)}{R(t)} = \frac{f(t)}{1 - F(t)}. \tag{7.8.3}$$

Note the relationships:

$$R(t) = e^{-\int_0^t Z(x)\, dx} \qquad\qquad f(t) = Z(t)e^{-\int_0^t Z(x)\, dx} \tag{7.8.4}$$

EXAMPLE If $f(t) = \alpha\beta t^{\beta-1}e^{\alpha t^\beta}$ with $\alpha > 0$ and $\beta > 0$, the probability distribution function for a Weibull random variable, then the failure rate is $Z(t) = \alpha\beta t^{\beta-1}$ and $R(t) = e^{-\alpha t^\beta}$. Note that failure rate decreases with time if $\beta < 1$ and increases with time if $\beta > 1$.

7.8.4.1 Use of the exponential distribution

If the hazard rate is a constant $Z(t) = \alpha$ (with $\alpha > 0$) then $f(t) = \alpha e^{-\alpha t}$ (for $t > 0$) which is the probability density function for an exponential random variable. If a failed item is replaced with another having the same constant hazard rate α, then the sequence of occurrence of failures is a Poisson process. The constant $1/\alpha$ is called the *mean time between failures* (MTBF). The reliability function is $R(t) = e^{-\alpha t}$.

If a series system has n components, each with constant hazard rate $\{\alpha_i\}$, then

$$R_s(t) = \exp\left(-\sum_{i=1}^{n} \alpha_i\right). \tag{7.8.5}$$

The MTBF for the series system is μ_s

$$\mu_s = \frac{1}{\frac{1}{\mu_1} + \frac{1}{\mu_2} + \cdots + \frac{1}{\mu_n}}. \tag{7.8.6}$$

If a parallel system has n components, each with identical constant hazard rate α, then the MTBF for the parallel system is μ_p

$$\mu_p = \frac{1}{\alpha}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n}\right). \tag{7.8.7}$$

7.8.5 PROCESS CAPABILITY

Notation

- C_p process capability
- C_{pk} process capability index
- C_{pl} process capability (one-sided)
- C_{pm} process capability index
- C_{ul} process capability (one-sided)
- LSL lower specification limit
- T target value
- USL upper specification limit
- μ process mean
- σ process standard deviation

Definitions

$$\begin{aligned} C_{pu} &= \frac{\text{USL} - \mu}{3\sigma} \qquad\qquad C_{pl} = \frac{\mu - \text{LSL}}{3\sigma} \\ C_p &= \frac{\text{USL} - \text{LSL}}{6\sigma} = \frac{C_{pu} + C_{pl}}{2} \\ C_{pk} &= \min\left(\frac{\text{USL} - \mu}{3\sigma}, \frac{\mu - \text{LSL}}{3\sigma}\right) = \min\left(C_{pu}, C_{pl}\right) \\ C_{pm} &= \frac{\text{USL} - \text{LSL}}{6\sqrt{\sigma^2 + (\mu - T)^2}} \end{aligned} \tag{7.8.8}$$

Estimates (here $\{\widehat{\mu}, \widehat{\sigma}\}$ are estimates of $\{\mu, \sigma\}$)

$$\begin{aligned} \widehat{C}_p &= \frac{\text{USL} - \text{LSL}}{6\widehat{\sigma}} \\ \widehat{C}_{pk} &= \min\left(\frac{\text{USL} - \widehat{\mu}}{3\widehat{\sigma}}, \frac{\widehat{\mu} - \text{LSL}}{3\widehat{\sigma}}\right) \\ \widehat{C}_m &= \frac{\text{USL} - \text{LSL}}{6\sqrt{\widehat{\sigma}^2 + (\widehat{\mu} - T)^2}} \end{aligned} \tag{7.8.9}$$

Approximate $100(1-\alpha)\%$ confidence interval assuming normal data and $n > 25$:

$$C_{pk} = \widehat{C}_{pk} \pm z_{1-\alpha/2}\sqrt{\frac{1}{9n} + \frac{\widehat{C}_{pk}^2}{2(n-1)}}$$

Comments

1. C_{pk} adjusts C_p to account for the effect of a non-centered distribution.
2. C_{pm} is used when a target value other than the center of the specification spread has been designated as desirable.
3. Usually $\widehat{\mu} = \bar{x}$. When there are subgroups (page 575 has d_2 and d_4 values):
   (a) $\widehat{\sigma} = \bar{R}/d_2$ where $\bar{R}$ is the average of the ranges; or
   (b) $\widehat{\sigma} = \bar{s}/d_4$ where $\bar{s}$ is the average of the sample standard deviations
4. The process must be in control before the process capability is computed; an out-of-control process will not yield a valid $\widehat{\sigma}$ estimate.

EXAMPLE The specification limits of a process are LSL $= 8$ and USL $= 20$. Process measurements yield $\bar{x} = 16$ and $\widehat{\sigma} = 2$. Hence, $\widehat{C}_P = 1.0$ which means that this process is capable. However, $\widehat{C}_{pk} = 0.67$ which means this process is not good (a value of $\widehat{C}_{pk} \geq 1$ is desired).

7.9 STATISTICS

Probability and statistics are related, but different concepts; statistics uses probabilistic thinking and analysis. The probability of an event is the likelihood of that event occurring; it is the relative frequency over a very large sample. A statistic is a value computed from a finite sample population.

7.9.1 DESCRIPTIVE STATISTICS

1. Sample distribution and density functions
   (a) Sample distribution function:
   $$\widehat{F}(x) = \frac{1}{n}\sum_{i=1}^{n} u(x - x_i) \tag{7.9.1}$$
   where $u(x)$ be the unit step function (or Heaviside function) defined by $u(x) = 0$ for $x \leq 0$ and $u(x) = 1$ for $x > 0$.
   (b) Sample density function or histogram:
   $$\hat{f}(x) = \frac{\widehat{F}(x_0 + (i+1)w) - \widehat{F}(x_0 + iw)}{w} \tag{7.9.2}$$
   for $x \in [x_0 + iw, x_0 + (i+1)w)$. The interval $[x_0 + iw, x_0 + (i+1)w)$ is called the i^{th} bin, w is the bin width, and $f_i = \widehat{F}(x_0 + (i+1)w) - \widehat{F}(x_0 + iw)$ is the *bin frequency*.
2. Order statistics and quantiles:
   (a) Order statistics are obtained by arranging the sample values $\{x_1, x_2, \ldots, x_n\}$ in increasing order, denoted by
   $$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}. \tag{7.9.3}$$
      i. $x_{(1)}$ and $x_{(n)}$ are the minimum and maximum data values
      ii. For $i = 1, \ldots, n$, $x_{(i)}$ is called the i^{th} *order statistic*.
   (b) *Quantiles*: If $0 < p < 1$, then the *quantile of order* p, ξ_p, is defined as the $p(n+1)^{\text{th}}$ order statistic. It may be necessary to interpolate between successive values.
      i. If $p = j/4$ for $j = 1, 2$, or 3, then $\xi_{\frac{j}{4}}$ is called the j^{th} *quartile*.
      ii. If $p = j/10$ for $j = 1, 2, \ldots, 9$, then $\xi_{\frac{j}{10}}$ is called the j^{th} *decile*.
      iii. If $p = j/100$ for $j = 1, 2, \ldots, 99$, then $\xi_{\frac{j}{100}}$ is called the j^{th} *percentile*.
3. Measures of central tendency:
   (a) *Arithmetic mean*:
   $$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + \cdots + x_n}{n}. \tag{7.9.4}$$

(b) α-*trimmed mean*:

$$\overline{x}_\alpha = \frac{1}{n(1-2\alpha)} \left((1-r)\left(x_{(k+1)} + x_{(n-k)}\right) + \sum_{i=k+2}^{n-k-1} x_{(i)} \right), \tag{7.9.5}$$

where $k = \lfloor \alpha n \rfloor$ is the greatest integer less than or equal to αn, and $r = \alpha n - k$. If $\alpha = 0$ then $\bar{x}_\alpha = \overline{x}$.

(c) *Weighted mean*: If to each x_i is associated a weight $w_i \geq 0$ so that

$$\sum_{i=1}^{n} w_i = 1, \text{ then } \overline{x}_{\mathrm{w}} = \sum_{i=1}^{n} w_i x_i. \tag{7.9.6}$$

(d) *Geometric mean*: (for $x_i \geq 0$)

$$\mathrm{GM} = \left(\prod_{i=1}^{n} x_i \right)^{1/n} = (x_1 x_2 \cdots x_n)^{1/n}. \tag{7.9.7}$$

(e) *Harmonic mean*:

$$\mathrm{HM} = \frac{n}{\sum_{i=1}^{n} \frac{1}{x_i}} = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}. \tag{7.9.8}$$

(f) Relationship between arithmetic mean ($\overline{x}$), geometric mean (GM), and harmonic mean (HM), when $x_i \geq 0$:

$$\mathrm{HM} \leq \mathrm{GM} \leq \overline{x} \tag{7.9.9}$$

with equality holding only when all sample values are equal.

(g) The *mode* is the data value that occurs with the greatest frequency. Note that the mode may not be unique.

(h) *Median*:

i. If n is odd and $n = 2k+1$, then $M = x_{(k+1)}$.

ii. If n is even and $n = 2k$, then $M = (x_{(k)} + x_{(k+1)})/2$.

(i) *Midrange*:

$$\mathrm{mid} = \frac{x_{(1)} + x_{(n)}}{2}. \tag{7.9.10}$$

4. Measures of dispersion

(a) *Mean deviation* or *absolute deviation*:

$$\mathrm{M.D.} = \frac{1}{n} \sum_{i=1}^{n} |x_i - \overline{x}|, \qquad \text{or} \qquad \mathrm{A.D.} = \frac{1}{n} \sum_{i=1}^{n} |x_i - M|. \tag{7.9.11}$$

(b) *Sample standard deviation*:

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}{n-1}}. \tag{7.9.12}$$

(c) The *sample variance* is the square of the sample standard deviation.

(d) *Root mean square*: $\text{R.M.S.} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2}$.

(e) *Sample range*: $x_{(n)} - x_{(1)}$.

(f) *Interquartile range*: $\xi_{\frac{3}{4}} - \xi_{\frac{1}{4}}$.

(g) The quartile deviation or semi-interquartile range is one-half the interquartile range.

5. Higher-order statistics

   (a) Sample moments: $m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$.

   (b) Sample central moments, or sample moments about the mean:

$$\mu_k = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^k .$$

7.9.2 STATISTICAL ESTIMATORS

7.9.2.1 Definitions

1. A function of a set of random variables is a *statistic*. It is a function of observable random variables that does not contain any unknown parameters. A statistic is itself an observable random variable.
2. Let θ be a parameter appearing in the density function for the random variable X. Suppose that we know a formula for computing an approximate value $\widehat{\theta}$ of θ from a given sample $\{x_1, \ldots, x_n\}$ (call such a function g). Then $\widehat{\theta} = g(x_1, x_2, \ldots, x_n)$ can be considered as a single observation of the random variable $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$. The random variable $\widehat{\Theta}$ is an *estimator* for the parameter θ.
3. A *hypothesis* is an assumption about the distribution of a random variable X. This may usually be cast into the form $\theta \in \Theta_0$. We use H_0 to denote the *null hypothesis* and H_1 to denote an *alternative hypothesis*. The null hypothesis is a statistical hypothesis tested for possible rejection (usually by finding that a very unlikely event occurs when the null hypothesis is assumed to be true).
4. In *significance testing*, a *test statistic* $T = T(X_1, \ldots, X_n)$ is used to *reject* H_0, or to *not reject* H_0. Generally, if $T \in C$, where C is a *critical region*, then H_0 is rejected.
5. A *type I error*, denoted α, is to reject H_0 when it should not be rejected. A *type II error*, denoted β, is to not reject H_0 when it should be rejected.
6. The power of a test is $\eta = 1 - \beta$.

	Unknown truth	
	H_0	H_1
Do not reject H_0	True decision. Probability is $1 - \alpha$	Type II error. Probability is β
Reject H_0	Type I error. Probability is α	True decision. Probability is $\eta = 1 - \beta$

7.9.2.2 Consistent estimators

Let $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$ be an estimator for the parameter θ, and suppose that g is defined for arbitrarily large values of n. If the estimator has the property, $\mathrm{E}[(\widehat{\Theta} - \theta)^2] \to 0$, as $n \to \infty$, then the estimator is called a *consistent estimator*.

1. A consistent estimator is not unique.
2. A consistent estimator need not be unbiased.

7.9.2.3 Efficient estimators

An unbiased estimator $\widehat{\Theta} = g(X_1, X_2, \ldots, X_n)$ for a parameter θ is said to be *efficient* if it has finite variance ($\mathrm{E}\left[(\widehat{\Theta} - \Theta)^2\right] < \infty$) and if there does not exist another estimator $\widehat{\Theta}^* = g^*(X_1, X_2, \ldots, X_n)$ for θ, whose variance is smaller than that of $\widehat{\Theta}$. The *efficiency* of an unbiased estimator is the ratio,

$$\frac{\text{Cramer–Rao lower bound}}{\text{Actual variance}}.$$

The *relative efficiency* of two unbiased estimators is the ratio of their variances.

7.9.2.4 Maximum likelihood estimators (MLE)

Suppose X is a random variable whose density function is $f(x;\theta)$, where $\theta = (\theta_1, \ldots, \theta_r)$. If the independent sample values $x_1, \ldots, x_n$ are obtained, then define the likelihood function as $L = \prod_{i=1}^n f(x_i;\theta)$. The MLE estimate for θ is the solution of the simultaneous equations, $\frac{\partial L}{\partial \theta_i} = 0$, for $i = 1, \ldots, r$.

1. A MLE need not be consistent.
2. A MLE may not be unbiased.
3. A MLE need not be unique.
4. If a single sufficient statistic T exists for the parameter θ, the MLE of θ must be a function of T.
5. Let $\widehat{\Theta}$ be a MLE of θ. If $\tau(\cdot)$ is a function with a single-valued inverse, then a MLE of $\tau(\theta)$ is $\tau(\widehat{\Theta})$.

Define $\overline{x} = \sum_{i=1}^n X_i/n$ and $S^2 = \sum_{i=1}^n (X_i - \overline{x})^2/n$ (note that $S \neq s$). Then:

Distribution	Estimated parameter	MLE estimate of parameter
Exponential $E(\lambda)$	$1/\lambda$	$1/\overline{x}$
Exponential $E(\lambda)$	$\lambda^2 = \sigma^2$	$\overline{x}^2$
Normal $N(\mu, \sigma)$	μ	$\overline{x}$
Normal $N(\mu, \sigma)$	σ^2	S^2
Poisson $P(\lambda)$	λ	$\overline{x}$
Uniform $U(0, \theta)$	θ	$X_{\max}$

7.9.2.5 Method of moments (MOM)

Let $\{X_i\}$ be independent and identically distributed random variables with density $f(x;\theta)$. Let $\mu'_r(\theta) = E[X^r]$ be the r^{th} moment (if it exists). Let $m'_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ be the r^{th} sample moment. Form the k equations, $\mu'_r = m'_r$, and solve to obtain an estimate of θ.

1. MOM estimators need not be unique.
2. MOM estimators may not be functions of sufficient or complete statistics.

7.9.2.6 Sufficient statistics

A statistic $G = g(X_1, \dots, X_n)$ is called as a *sufficient statistic* if and only if the conditional distribution of H, given G, does not depend on θ for any statistic $H = h(X_1, \dots, X_n)$.

Let $\{X_i\}$ be independent and identically distributed random variables, with density $f(x;\theta)$. The statistics $\{G_1, \dots, G_r\}$ are said to be *jointly sufficient statistics* if and only if the conditional distribution of X_1, X_2, …, X_n given $G_1 = g_1$, $G_1 = g_2$, …, $G_r = g_r$ does not depend on θ.

Note that a single sufficient statistic may not exist.

7.9.2.7 UMVU estimators

A *uniformly minimum variance unbiased* estimator, called a UMVU estimator, is unbiased and has the minimum variance among all unbiased estimators.

Define, as usual, $\overline{x} = \sum_{i=1}^{n} X_i/n$ and $s^2 = \sum_{i=1}^{n}(X_i - \overline{x})^2/(n-1)$. Then:

Distribution	Estimated parameter	UMVU estimate of parameter	Variance of estimator
Exponential $E(\lambda)$	λ	$\dfrac{n-1}{s}$	$\dfrac{\lambda^2}{n-2}$
Exponential $E(\lambda)$	$\dfrac{1}{\lambda}$	$\overline{x}$	$\dfrac{1}{n\lambda^2}$
Normal $N(\mu,\sigma)$	μ	$\overline{x}$	$\dfrac{\sigma^2}{n}$
Normal $N(\mu,\sigma)$	σ^2	s^2	$\dfrac{2\sigma^4}{n-1}$
Poisson $P(\lambda)$	λ	$\overline{x}$	$\dfrac{\lambda^2}{n}$
Uniform $U(0,\theta)$	θ	$\dfrac{n+1}{n}X_{\text{max}}$	$\dfrac{\theta^2}{n(n+2)}$

7.9.2.8 Unbiased estimators

An estimator $g(X_1, X_2, \dots, X_n)$ for a parameter θ is said to be *unbiased* if

$$E[g(X_1, X_2, \dots, X_n)] = \theta. \tag{7.9.13}$$

1. An unbiased estimator may not exist.
2. An unbiased estimator is not unique.
3. An unbiased estimator need not be consistent.

7.9.2.9 Estimators for mean and variance

The estimates below do not use all of the data points; except in the $n = 2$ case. Hence, their efficiencies are all less than 1.

The variance of the estimates below must be multiplied by the true variance of the sample, σ^2.

- Small samples
  1. Estimating standard deviation σ from the sample range w

n	Estimator	Variance	Efficiency
2	$0.886w$	.571	1.000
3	$0.591w$	.275	.992
4	$0.486w$	.183	.975
5	$0.430w$	.138	.955
10	$0.325w$	.067	.850
20	$0.268w$	.038	.700

  2. Best linear estimate of the standard deviation σ

n	Estimator	Efficiency
2	$0.8862(x_2 - x_1)$	1.000
3	$0.5908(x_3 - x_1)$	.992
4	$0.4539(x_4 - x_1) + 0.1102(x_3 - x_2)$	.989
5	$0.3724(x_5 - x_1) + 0.1352(x_3 - x_2)$	.988
6	$0.3175(x_6 - x_1) + 0.1386(x_5 - x_2) + 0.0432(x_4 - x_3)$	.988
7	$0.2778(x_7 - x_1) + 0.1351(x_6 - x_2) + 0.0625(x_5 - x_3)$	.989

- Large samples
  Use percentile estimates to estimate the mean and standard deviation.
  1. Estimators for the mean

Number of terms	Estimator using percentiles	Efficiency
1	P_{50}	.64
2	$1/2\,(P_{25} + P_{75})$	.81
3	$1/3\,(P_{17} + P_{50} + P_{83})$	.88
4	$1/4\,(P_{12.5} + P_{37.5} + P_{62.5} + P_{87.5})$	.91
5	$1/5\,(P_{10} + P_{30} + P_{50} + P_{70} + P_{90})$	.93

  2. Estimators for the standard deviation

Number of terms	Estimator using percentiles	Efficiency
2	$0.3388\,(P_{93} - P_7)$	.65
4	$0.1714\,(P_{97} + P_{85} - P_{15} - P_3)$	.80
6	$0.1180\,(P_{98} + P_{91} + P_{80} - P_{20} - P_9 - P_2)$	.87

7.9.3 CRAMER–RAO BOUND

The *Cramer–Rao bound* gives a lower bound on the variance of an unknown unbiased statistical parameter, when n samples are taken. When the single unknown parameter is θ,

$$\sigma^2(\theta) \geq \frac{1}{-n\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x;\theta)\right]} = \frac{1}{n\mathrm{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x;\theta)\right)^2\right]}. \tag{7.9.14}$$

EXAMPLES

1. For a normal random variable with unknown mean θ and known variance σ^2, the density is $f(x;\theta) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\theta)^2}{2\sigma^2}\right)$. Hence, $\frac{\partial}{\partial\theta}\log f(x;\theta) = (x-\theta)/\sigma^2$. The computation
$$\mathrm{E}\left[\frac{(x-\theta)^2}{\sigma^4}\right] = \int_{-\infty}^{\infty}\frac{(x-\theta)^2}{\sigma^4}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\theta)^2/2\sigma^2}\,dx = \frac{1}{\sigma^2}$$
results in $\sigma^2(\theta) \geq \frac{\sigma^2}{n}$.
2. For a normal random variable with known mean μ and unknown variance $\theta = \sigma^2$, the density is $f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}}\exp\left(-\frac{(x-\mu)^2}{2\theta^2}\right)$. Hence, $\frac{\partial}{\partial\theta}\log f(x;\theta) = ((x-\mu)^2 - 2\theta)/(2\theta)^2$. The computation $\mathrm{E}\left[\frac{(x-\mu)^2-2\theta}{(2\theta)^2}\right] = \frac{1}{2\theta^2} = \frac{1}{2\sigma^4}$ results in $\sigma^2(\theta) \geq 2\sigma^4/n$.
3. For a Poisson random variable with unknown mean θ, the density is $f(x;\theta) = \theta^x e^{-\theta}/x!$. Hence, $\frac{\partial}{\partial\theta}\log f(x;\theta) = x/\theta - 1$. The computation $\mathrm{E}\left[\left(\frac{x}{\theta}-1\right)^2\right] = \sum_{x=0}^{\infty}\left(\frac{x}{\theta}-1\right)^2\frac{\theta^x e^{-\theta}}{x!} = \frac{1}{\theta}$ results in $\sigma^2(\theta) \geq \theta/n$.

7.9.4 CLASSIC STATISTICS PROBLEMS

7.9.4.1 Sample size problem

Suppose that a Bernoulli random variable (page 621) is to be estimated from a sample. What sample size n is required so that, with 99% certainty, the error is no more than $e = 5$ percentage points (i.e., $\mathrm{Prob}(|\hat{p} - p| < 0.05) > 0.99$)?

If an a priori estimate of p is available, then the minimum sample size is $n_p = z^2_{\alpha/2}p(1-p)/e^2$. If no a priori estimate is available (i.e., no estimate is available before the experiment), then $n_n = z^2_{\alpha/2}/4e^2 \geq n_p\,|_{p=1/2}$. For the numbers above, $n \geq n_n = 664$. Section 7.14 on page 620 has more information on sample sizes.

7.9.4.2 Large scale testing with infrequent success

Suppose that a disease occurs in one person out of every 1,000. Suppose that a test for this disease has a type I and a type II error of 1% (that is, $\alpha = \beta = 0.01$). Imagine that 100,000 people are tested. Of the 100 people who have the disease, 99 will be diagnosed as having it. Of the 99,900 people who do not have the disease, 999 will be diagnosed as having it. Hence, only $\frac{99}{1098} \approx 9\%$ of the people who test positive for the disease actually have it.

7.9.5 ORDER STATISTICS

When $\{X_i\}$ are n independent and identically distributed random variables with the common distribution function $F_X(x)$, let Z_m be the m^{th} largest of the values ($m = 1, 2, \ldots, n$). Hence Z_1 is the maximum of the n values and Z_n is the minimum of the n values. Then

$$F_{Z_m}(x) = \sum_{i=m}^{n} \binom{n}{i} [F_X(x)]^i [1 - F_X(x)]^{n-i} \tag{7.9.15}$$

Hence

$$F_{\text{max}}(z) = [F_X(z)]^n, \qquad f_{\text{max}}(z) = n [F_X(z)]^{n-1} f_X(z), \tag{7.9.16}$$

$$F_{\text{min}}(z) = 1 - [1 - F_X(z)]^n, \qquad f_{\text{min}}(z) = n [1 - F_X(z)]^{n-1} f_X(z). \tag{7.9.17}$$

The expected value of the i^{th} order statistic is given by

$$\mathrm{E}\left[x_{(i)}\right] = \frac{n!}{(i-1)!(n-i)!} \int_{-\infty}^{\infty} x f(x) F^{i-1}(x) [1 - F(x)]^{n-i} \, dx. \tag{7.9.18}$$

7.9.5.1 Uniform distribution

If X is uniformly distributed on the interval $[0, 1)$ then

$$\mathrm{E}\left[x_{(i)}\right] = \frac{n!}{(i-1)!(n-i)!} \int_0^1 x^i (1-x)^{n-i} \, dx. \tag{7.9.19}$$

The expected value of the largest of n samples is $\frac{n}{n+1}$; the expected value of the least of n samples is $\frac{1}{n+1}$.

7.9.5.2 Normal distribution

The following table gives values of $\mathrm{E}\left[x_{(i)}\right]$ for a standard normal distribution. Missing values (indicated by a dash) may be obtained from $\mathrm{E}\left[x_{(i)}\right] = -\mathrm{E}\left[x_{(n-i+1)}\right]$.

i	$n=2$	3	4	5	6	7	8	10
1	0.5642	0.8463	1.0294	1.1630	1.2672	1.3522	1.4236	1.5388
2	—	0.0000	0.2970	0.4950	0.6418	0.7574	0.8522	1.0014
3		—	—	0.0000	0.2016	0.3527	0.4728	0.6561
4			—	—	—	0.0000	0.1522	0.3756
5				—	—	—	—	0.1226
6					—	—	—	—

EXAMPLE

If a person of average intelligence takes five intelligence tests (each test having a normal distribution with a mean of 100 and a standard deviation of 20), then the expected value of the largest score is $100 + (1.1630)(20) \approx 123$.

7.10 CONFIDENCE INTERVALS

A probability distribution may have one or more unknown parameters. A *confidence interval* is an assertion that an unknown parameter lies in a computed range, with a specified probability. Before constructing a confidence interval, first select a confidence coefficient, denoted $1-\alpha$. Typically, $1-\alpha = 0.95, 0.99$, or the like. The definitions of z_α, t_α, and χ^2_α are in Section 7.17.1 on page 628.

7.10.1 SAMPLE FROM ONE POPULATION

The following confidence intervals assume a random sample of size n, given by $\{x_1, x_2, \ldots, x_n\}$.

#	parameter	assumptions	$100(1-\alpha)\%$ confidence interval
1	μ	normal distribution, σ^2 known	$\overline{x} \pm z_{\alpha/2}\sigma/\sqrt{n}$
2	μ	normal distribution, σ^2 unknown	$\overline{x} \pm t_{\alpha/2}s/\sqrt{n}$
3	p	Bernoulli trials	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
4	σ^2	normal distribution	$\left[\frac{(n-1)s^2}{\chi^2_{\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}\right]$
5	quantile ξ_p	large sample	$\left[x_{(k_1)}, x_{(k_2)}\right]$
6	median	large sample	$\left[w_{(k_1)}, w_{(k_2)}\right]$

1. Find mean μ of the normal distribution with known variance σ^2.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1-\alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the mean $\overline{x}$ of the sample.
   (c) Compute $k = z_{\alpha/2}\sigma/\sqrt{n}o$.
   (d) The $100(1-\alpha)$ percent confidence interval for μ is given by $[\overline{x}-k, \overline{x}+k]$.

2. Find mean μ of the normal distribution with unknown variance σ^2.
   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1-\alpha/2$, where $F(\cdot)$ is the t-distribution with $n-1$ degrees of freedom.
   (b) Compute the mean $\overline{x}$ and standard deviation s of the sample.
   (c) Compute $k = t_{\alpha/2}s/\sqrt{n}$.
   (d) The $100(1-\alpha)$ percent confidence interval for μ is given by $[\overline{x}-k, \overline{x}+k]$.

3. Find the probability of success p for Bernoulli trials with large sample sizes.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the proportion $\hat{p}$ of "successes" out of n trials.
   (c) Compute $k = z_{\alpha/2}\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}$.
   (d) The $100(1-\alpha)$ percent confidence interval for p is given by $[\hat{p} - k, \hat{p} + k]$.

4. Find variance σ^2 of the normal distribution.
   (a) Determine the critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ such that $F\left(\chi^2_{\alpha/2}\right) = 1 - \alpha/2$ and $F\left(\chi^2_{1-\alpha/2}\right) = \alpha/2$, where $F(z)$ is the chi-square distribution function with $n-1$ degrees of freedom.
   (b) Compute the sample standard deviation s.
   (c) Compute $k_1 = \dfrac{(n-1)s^2}{\chi^2_{\alpha/2}}$ and $k_2 = \dfrac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$.
   (d) The $100(1-\alpha)$ percent confidence interval for σ^2 is given by $[k_1, k_2]$.
   (e) The $100(1-\alpha)$ percent confidence interval for the standard deviation σ is given by $\left[\sqrt{k_1}, \sqrt{k_2}\right]$.

5. Find quantile ξ_p of order p for large sample sizes.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the order statistics $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.
   (c) Compute $k_1 = \left\lfloor np - z_{\alpha/2}\sqrt{np(1-p)} \right\rfloor$ and $k_2 = \left\lceil np + z_{\alpha/2}\sqrt{np(1-p)} \right\rceil$.
   (d) The $100(1-\alpha)$ percent confidence interval for ξ_p is given by $\left[x_{(k_1)}, x_{(k_2)}\right]$.

6. Find median M based on the Wilcoxon one-sample statistic for a large sample.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the order statistics $w_{(1)}, w_{(2)}, \ldots, w_{(N)}$ of the $N = n(n-1)/2$ averages $\left(x_i + x_j\right)/2$, for $1 \leq i < j \leq n$.
   (c) Compute $k_1 = \left\lfloor \dfrac{N}{2} - \dfrac{z_{\alpha/2}N}{\sqrt{3n}} \right\rfloor$ and $k_2 = \left\lceil \dfrac{N}{2} + \dfrac{z_{\alpha/2}N}{\sqrt{3n}} \right\rceil$.
   (d) The $100(1-\alpha)$ percent confidence interval for M is given by $\left[w_{(k_1)}, w_{(k_2)}\right]$.

7.10.2 SAMPLES FROM TWO POPULATIONS

The following confidence intervals assume random samples from two large populations: one sample of size n, given by $\{x_1, x_2, \dots, x_n\}$, and one sample of size m, given by $\{y_1, y_2, \dots, y_m\}$.

#	parameter	assumptions	$100(1-\alpha)\%$ confidence interval
1	$\mu_x - \mu_y$	independent samples, known variances	$(\overline{x}-\overline{y}) \pm z_{\alpha/2}\sqrt{\frac{\sigma_x^2}{n}+\frac{\sigma_y^2}{m}}$
2	$\mu_x - \mu_y$	independent samples, unknown variances	$(\overline{x}-\overline{y}) \pm z_{\alpha/2}\sqrt{\frac{s_x^2}{n}+\frac{s_y^2}{m}}$
3	$\mu_x - \mu_y$	independent samples, unknown but equal variances	$(\overline{x}-\overline{y}) \pm t_{\alpha/2}\, s\sqrt{\frac{1}{n}+\frac{1}{m}}$
4	$\mu_x - \mu_y$	paired samples, unknown but equal variances	$\bar{\mu}_d \pm t_{\alpha/2}\, s_d/\sqrt{n}$
5	$p_x - p_y$	independent samples, Bernoulli trials, large samples	$(\hat{p}_x-\hat{p}_y) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n}+\frac{\hat{p}_y(1-\hat{p}_y)}{m}}$
6	$M_x - M_y$	independent samples, large samples	$\left[w_{(k_1)}, w_{(k_2)}\right]$
7	σ_x^2/σ_y^2	independent samples, large samples	$\left[\frac{s_x^2}{s_y^2}F_{1-\alpha/2}, \frac{s_x^2}{s_y^2}F_{\alpha/2}\right]$

1. Find the difference in population means μ_x and μ_y from independent samples with known variances σ_x^2 and σ_y^2.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1-\alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the means $\overline{x}$ and $\overline{y}$.
   (c) Compute $k = z_{\alpha/2}\sqrt{\dfrac{\sigma_x^2}{n}+\dfrac{\sigma_y^2}{m}}$.
   (d) The $100(1-\alpha)$ percent confidence interval for $\mu_x-\mu_y$ is given by $[(\overline{x}-\overline{y})-k, (\bar{x}-\overline{y})+k]$.
2. Find the difference in population means μ_x and μ_y from independent samples with unknown variances σ_x^2 and σ_y^2.
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1-\alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the means $\overline{x}$ and $\overline{y}$, and the standard deviations s_x and s_y.

(c) Compute $k = z_{\alpha/2}\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}$.

(d) The $100(1-\alpha)$ percent confidence interval for $\mu_x - \mu_y$ is given by $[(\overline{x} - \overline{y}) - k, (\bar{x} - \overline{y}) + k]$.

3. Find the difference in population means μ_x and μ_y from independent samples with unknown but equal variances $\sigma_x^2 = \sigma_y^2$.

   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the t-distribution with $n + m - 2$ degrees of freedom.

   (b) Compute the means $\overline{x}$ and $\overline{y}$, the standard deviations s_x and s_y, and the pooled standard deviation estimate,

$$s = \sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}. \tag{7.10.1}$$

   (c) Compute $k = t_{\alpha/2}\, s\sqrt{\frac{1}{n} + \frac{1}{m}}$.

   (d) The $100(1-\alpha)$ percent confidence interval for $\mu_x - \mu_y$ is given by $[(\overline{x} - \overline{y}) - k, (\overline{x} - \bar{y}) + k]$.

4. Find the difference in population means μ_x and μ_y for paired samples with unknown but equal variances $\sigma_x^2 = \sigma_y^2$.

   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the t-distribution with $n - 1$ degrees of freedom.

   (b) Compute the mean $\bar{\mu}_d$ and standard deviation s_d of the paired differences $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$.

   (c) Compute $k = t_{\alpha/2}\, s_d/\sqrt{n}$.

   (d) The $100(1-\alpha)$ percent confidence interval for $\mu_d = \mu_x - \mu_y$ is given by $[\bar{\mu}_d - k, \bar{\mu}_d + k]$.

5. Find the difference in Bernoulli trial success rates, $p_x - p_y$, for large, independent samples.

   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.

   (b) Compute the proportions $\hat{p}_x$ and $\hat{p}_y$ of "successes" for the samples.

   (c) Compute $k = z_{\alpha/2}\sqrt{\frac{\hat{p}_x(1-\hat{p}_x)}{n} + \frac{\hat{p}_y(1-\hat{p}_y)}{m}}$.

   (d) The $100(1-\alpha)$ percent confidence interval for $p_x - p_y$ is given by $[(\hat{p}_x - \hat{p}_y) - k, (\hat{p}_x - \hat{p}_y) + k]$.

6. Find the difference in medians $M_x - M_y$ based on the Mann–Whitney–Wilcoxon procedure.

   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.

(b) Compute the order statistics $w_{(1)}, w_{(2)}, \dots, w_{(N)}$ of the $N = nm$ differences $x_i - y_j$, for $1 \le i \le n$ and $1 \le j \le m$.

(c) Compute

$$k_1 = \frac{nm}{2} + \left\lfloor 0.5 - z_{\alpha/2}\sqrt{\frac{nm\,(n+m+1)}{12}} \right\rfloor$$

and

$$k_2 = \left\lceil \frac{nm}{2} - 0.5 + z_{\alpha/2}\sqrt{\frac{nm\,(n+m+1)}{12}} \right\rceil .$$

(d) The $100(1-\alpha)$ percent confidence interval for $M_x - M_y$ is given by $\left[w_{(k_1)}, w_{(k_2)}\right]$.

7. Find the ratio of variances σ_x^2/σ_y^2, for independent samples.

   (a) Determine the critical values $F_{\alpha/2}$ and $F_{1-\alpha/2}$ such that $F\left(F_{\alpha/2}\right) = 1-\alpha/2$ and $F\left(F_{1-\alpha/2}\right) = \alpha/2$, where $F(\cdot)$ is the F-distribution with $m-1$ and $n-1$ degrees of freedom.

   (b) Compute the standard deviations s_x and s_y of the samples.

   (c) Compute $k_1 = F_{1-\alpha/2}$ and $k_2 = F_{\alpha/2}$.

   (d) The $100(1-\alpha)$ percent confidence interval for σ_x^2/σ_y^2 is given by $\left[\frac{s_x^2}{s_y^2}k_1, \frac{s_x^2}{s_y^2}k_2\right]$.

7.10.3 CONFIDENCE INTERVAL FOR BINOMIAL PARAMETER

The probability distribution function of a binomial random variable is

$$f(x;n,p) = \binom{n}{x} p^x (1-p)^{n-x} \qquad \text{for } x = 0,1,2,\dots,n \tag{7.10.2}$$

For known n, any value of x' less than n, and $\alpha < 1$, lower and upper values of p (p_{lower} and p_{upper}) may be determined such that $p_{\text{lower}} < p_{\text{upper}}$ and

$$\sum_{x=0}^{x'} f(x;n,p_{\text{lower}}) = \frac{\alpha}{2} \quad \text{and} \quad \sum_{x=x'}^{n} f(x;n,p_{\text{upper}}) = \frac{\alpha}{2} \tag{7.10.3}$$

The following tables show p_{lower} and p_{upper} for $\alpha = 0.01$ and $\alpha = 0.05$.

EXAMPLE In a binomial experiment having $n = 12$ trials, a total of $x = 3$ successes were observed. Determine a 95% and a 99% confidence interval for the probability of a success p.

From the following tables with $n - x = 9$ and $x = 3$

- 95% confidence: $p_{\text{lower}} = 0.055$ and $p_{\text{upper}} = 0.572$. Hence, the probability is .95 that the interval $(0.055, 0.572)$ contains the true value of p.
- 99% confidence: $p_{\text{lower}} = 0.030$ and $p_{\text{upper}} = 0.655$. Hence, the probability is .99 that the interval $(0.030, 0.655)$ contains the true value of p.

Confidence limits of proportions (confidence coefficient 0.95)

	Denominator minus numerator: $n - x$ (Lower limit in italic type, upper limit in roman type)								
x	1	2	3	4	5	6	7	8	9
0	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	0.975	0.842	0.708	0.602	0.522	0.459	0.410	0.369	0.336
1	*0.013*	*0.008*	*0.006*	*0.005*	*0.004*	*0.004*	*0.003*	*0.003*	*0.003*
	0.987	0.906	0.806	0.716	0.641	0.579	0.527	0.482	0.445
2	*0.094*	*0.068*	*0.053*	*0.043*	*0.037*	*0.032*	*0.028*	*0.025*	*0.023*
	0.992	0.932	0.853	0.777	0.710	0.651	0.600	0.556	0.518
3	*0.194*	*0.147*	*0.118*	*0.099*	*0.085*	*0.075*	*0.067*	*0.060*	*0.055*
	0.994	0.947	0.882	0.816	0.755	0.701	0.652	0.610	0.572
4	*0.284*	*0.223*	*0.184*	*0.157*	*0.137*	*0.122*	*0.109*	*0.099*	*0.091*
	0.995	0.957	0.901	0.843	0.788	0.738	0.692	0.651	0.614
5	*0.359*	*0.290*	*0.245*	*0.212*	*0.187*	*0.168*	*0.152*	*0.139*	*0.128*
	0.996	0.963	0.915	0.863	0.813	0.766	0.723	0.684	0.649
10	*0.587*	*0.516*	*0.462*	*0.419*	*0.384*	*0.354*	*0.329*	*0.308*	*0.289*
	0.999	0.989	0.972	0.953	0.932	0.910	0.889	0.868	0.847
50	*0.896*	*0.868*	*0.843*	*0.821*	*0.800*	*0.781*	*0.763*	*0.746*	*0.730*
	0.999	0.995	0.988	0.979	0.970	0.960	0.949	0.939	0.928
100	*0.946*	*0.931*	*0.917*	*0.904*	*0.892*	*0.881*	*0.870*	*0.859*	*0.849*
	1.000	0.998	0.994	0.989	0.984	0.979	0.973	0.967	0.962

Confidence limits of proportions (confidence coefficient 0.99)

	Denominator minus numerator: $n - x$ (Lower limit in italic type, upper limit in roman type)								
x	1	2	3	4	5	6	7	8	9
0	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*	*0.000*
	0.995	0.929	0.829	0.734	0.653	0.586	0.531	0.484	0.445
1	*0.003*	*0.002*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*	*0.001*
	0.997	0.959	0.889	0.815	0.746	0.685	0.632	0.585	0.544
2	*0.041*	*0.029*	*0.023*	*0.019*	*0.016*	*0.014*	*0.012*	*0.011*	*0.010*
	0.998	0.971	0.917	0.856	0.797	0.742	0.693	0.648	0.608
3	*0.111*	*0.083*	*0.066*	*0.055*	*0.047*	*0.042*	*0.037*	*0.033*	*0.030*
	0.999	0.977	0.934	0.882	0.830	0.781	0.735	0.693	0.655
4	*0.185*	*0.144*	*0.118*	*0.100*	*0.087*	*0.077*	*0.069*	*0.062*	*0.057*
	0.999	0.981	0.945	0.900	0.854	0.809	0.767	0.727	0.691
5	*0.254*	*0.203*	*0.170*	*0.146*	*0.128*	*0.114*	*0.103*	*0.094*	*0.087*
	0.999	0.984	0.953	0.913	0.872	0.831	0.791	0.755	0.720
10	*0.491*	*0.427*	*0.379*	*0.342*	*0.312*	*0.287*	*0.266*	*0.247*	*0.232*
	1.000	0.991	0.972	0.947	0.920	0.891	0.863	0.835	0.808
50	*0.863*	*0.834*	*0.808*	*0.785*	*0.763*	*0.744*	*0.725*	*0.708*	*0.692*
	1.000	0.998	0.993	0.987	0.980	0.972	0.963	0.954	0.945
100	*0.929*	*0.912*	*0.898*	*0.884*	*0.871*	*0.859*	*0.847*	*0.836*	*0.826*
	1.000	0.999	0.997	0.993	0.990	0.985	0.981	0.976	0.971

7.10.4 CONFIDENCE INTERVAL FOR POISSON PARAMETER

The probability density function for a Poisson random variable is

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} \qquad \text{for } x = 0, 1, 2, \dots \tag{7.10.4}$$

For any value of x' and $\alpha < 1$, lower and upper values of λ (λ_{lower} and λ_{upper}) may be determined such that $\lambda_{\text{lower}} < \lambda_{\text{upper}}$ and

$$\sum_{x=0}^{x'} \frac{e^{-\lambda_{\text{lower}}}\lambda_{\text{lower}}^x}{x!} = \frac{\alpha}{2} \quad \text{and} \quad \sum_{x=x'}^{\infty} \frac{e^{-\lambda_{\text{upper}}}\lambda_{\text{upper}}^x}{x!} = \frac{\alpha}{2} \tag{7.10.5}$$

The table below has λ_{lower} and λ_{upper} for $\alpha = 0.01$ and $\alpha = 0.05$. For $x' > 50$, λ_{upper} and λ_{lower} may be approximated by

$$\begin{aligned}\lambda_{\text{upper}} &= \frac{\chi^2_{1-\alpha,n}}{2} \text{ where } 1 - F(\chi^2_{1-\alpha,n}) = \alpha, \text{ and } n = 2(x'+1) \\ \lambda_{\text{lower}} &= \frac{\chi^2_{\alpha,n}}{2} \text{ where } F(\chi^2_{\alpha,n}) = \alpha, \text{ and } n = 2x'\end{aligned} \tag{7.10.6}$$

where $F(\chi^2)$ is the cumulative distribution function for a chi-square random variable with n degrees of freedom.

EXAMPLE In a Poisson process, 5 outcomes are observed during a time interval. Find a 95% and a 99% confidence interval for the parameter λ in this Poisson process.

For an observed count of 5 and a 95% significance level ($\alpha = 0.05$), the confidence interval bounds are $\lambda_{\text{lower}} = 1.6$ and $\lambda_{\text{upper}} = 11.7$. Hence, the probability is .95 that the interval $(1.6, 11.7)$ contains the true value of λ.

For an observed count of 5 and a 99% significance level ($\alpha = 0.01$), the confidence interval bounds are $\lambda_{\text{lower}} = 1.1$ and $\lambda_{\text{upper}} = 14.1$. Hence, the probability is .99 that the interval $(1.1, 14.1)$ contains the true value of λ.

Confidence limits for the parameter in a Poisson distribution.

Observed	Significance level			
	$\alpha = 0.01$		$\alpha = 0.05$	
count	λ_{lower}	λ_{higher}	λ_{lower}	λ_{higher}
0	0.0	5.3	0.0	3.7
1	0.0	7.4	0.0	5.6
2	0.1	9.3	0.2	7.2
3	0.3	11.0	0.6	8.8
4	0.7	12.6	1.1	10.2
5	1.1	14.1	1.6	11.7
6	1.5	15.7	2.2	13.1
7	2.0	17.1	2.8	14.4
8	2.6	18.6	3.5	15.8
9	3.1	20.0	4.1	17.1
10	3.7	21.4	4.8	18.4

Observed	Significance level			
	$\alpha = 0.01$		$\alpha = 0.05$	
count	λ_{lower}	λ_{higher}	λ_{lower}	λ_{higher}
11	4.3	22.8	5.5	19.7
12	4.9	24.1	6.2	21.0
13	5.6	25.5	6.9	22.2
14	6.2	26.8	7.7	23.5
15	6.9	28.2	8.4	24.7
20	10.4	34.7	12.2	30.9
25	14.0	41.0	16.2	36.9
30	17.8	47.2	20.2	42.8
35	21.6	53.3	24.4	48.7
40	25.6	59.4	28.6	54.5
45	29.6	65.3	32.8	60.2
50	33.7	71.3	37.1	65.9

7.11 TESTS OF HYPOTHESES

A statistical hypothesis is a statement about the distribution of a random variable. A statistical test of a hypothesis is a procedure in which a sample is used to determine whether we should "reject" or "not reject" the hypothesis. Before employing a hypothesis test, first select a *significance level* α. The significance level of a test is the probability of mistakenly rejecting the null hypothesis. Typically, α is 0.05 or 0.01 (i.e, 5% or 1%) or similar.

7.11.1 HYPOTHESIS TESTS: PARAMETER FROM ONE POPULATION

The following hypothesis tests assume a random sample of size n, given by $\{x_1, x_2, \ldots, x_n\}$.

1. Test of the hypothesis $\mu = \mu_0$ against the alternative $\mu \neq \mu_0$ of the mean of a normal distribution with known variance σ^2:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the mean $\overline{x}$ of the sample.
   (c) Compute the test statistic $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$.
   (d) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

2. Test of the hypothesis $\mu = \mu_0$ against the alternative $\mu > \mu_0$ (or $\mu < \mu_0$) of the mean of a normal distribution with known variance σ^2:
   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the mean $\overline{x}$ of the sample.
   (c) Compute the test statistic $z = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{\sigma}$. (For the alternative $\mu < \mu_0$, multiply z by -1.)
   (d) If $z > z_\alpha$, then reject the hypothesis. If $z \leq z_\alpha$, then do not reject the hypothesis.

3. Test of the hypothesis $\mu = \mu_0$ against the alternative $\mu \neq \mu_0$ of the mean of a normal distribution with unknown variance σ^2:
   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the t-distribution with $n-1$ degrees of freedom.
   (b) Compute the mean $\overline{x}$ and standard deviation s of the sample.
   (c) Compute the test statistic $t = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{s}$.
   (d) If $|t| > t_{\alpha/2}$, then reject the hypothesis. If $|t| \leq t_{\alpha/2}$, then do not reject the hypothesis.

4. Test of the hypothesis $\mu = \mu_0$ against the alternative $\mu > \mu_0$ (or $\mu < \mu_0$) of the mean of a normal distribution with unknown variance σ^2:
   (a) Determine the critical value t_α such that $F(t_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the t-distribution with $n - 1$ degrees of freedom.
   (b) Compute the mean $\overline{x}$ and standard deviation s of the sample.
   (c) Compute the test statistic $t = \dfrac{(\overline{x} - \mu_0)\sqrt{n}}{s}$. (For the alternative $\mu < \mu_0$, multiply t by -1.)
   (d) If $t > t_\alpha$, then reject the hypothesis. If $t \leq t_\alpha$, then do not reject the hypothesis.

5. Test of the hypothesis $p = p_0$ against the alternative $p \neq p_0$ of the probability of success for a binomial distribution, large sample:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the proportion $\hat{p}$ of "successes" for the sample.
   (c) Compute the test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.
   (d) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

6. Test of the hypothesis $p = p_0$ against the alternative $p > p_0$ (or $p < p_0$) of the probability of success for a binomial distribution, large sample:
   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the proportion $\hat{p}$ of "successes" for the sample.
   (c) Compute the test statistic $z = \dfrac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$. (For the alternative $p < p_0$, multiply z by -1.)
   (d) If $z > z_\alpha$, then reject the hypothesis. If $z \leq z_\alpha$, then do not reject the hypothesis.

7. Wilcoxon signed rank test of the hypothesis $M = M_0$ against the alternative $M \neq M_0$ of the median of a population, large sample:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution.
   (b) Compute the quantities $|x_i - M_0|$, and keep track of the sign of $x_i - M_0$. If $|x_i - M_0| = 0$, then remove it from the list and reduce n by one.
   (c) Order the $|x_i - M_0|$ from smallest to largest, assigning rank 1 to the smallest and rank n to the largest; $|x_i - M_0|$ has rank r_i if it is the r_i^{th} entry in the ordered list. In case of ties (i.e., $|x_i - M_0| = |x_j - M_0|$ for 2 or more values) assign each the average of their ranks.
   (d) Compute the sum of the signed ranks $R = \displaystyle\sum_{i=1}^{n} \text{sign}(x_i - M_0)\, r_i$.

(e) Compute the test statistic $z = \dfrac{R}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$.

(f) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \le z_{\alpha/2}$, then do not reject the hypothesis.

8. Wilcoxon signed rank test of the hypothesis $M = M_0$ against the alternative $M > M_0$ (or $M < M_0$) of the median of a population, large sample:

   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution.

   (b) Compute the quantities $|x_i - M_0|$, and keep track of the sign of $x_i - M_0$. If $|x_i - M_0| = 0$, then remove it from the list and reduce n by one.

   (c) Order the $|x_i - M_0|$ from smallest to largest, assigning rank 1 to the smallest and rank n to the largest; $|x_i - M_0|$ has rank r_i if it is the r_i^{th} entry in the ordered list. If $|x_i - M_0| = |x_j - M_0|$, then assign each the average of their ranks.

   (d) Compute the sum of the signed ranks $R = \sum_{i=1}^{n} \operatorname{sign}(x_i - M_0)\, r_i$.

   (e) Compute the test statistic $z = \dfrac{R}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$. (For the alternative $M < M_0$, multiply the test statistic by -1.)

   (f) If $z > z_\alpha$, then reject the hypothesis. If $z \le z_\alpha$, then do not reject the hypothesis.

9. Test of the hypothesis $\sigma^2 = \sigma_0^2$ against the alternative $\sigma^2 \ne \sigma_0^2$ of the variance of a normal distribution:

   (a) Determine the critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ such that $F\left(\chi^2_{\alpha/2}\right) = 1 - \alpha/2$ and $F\left(\chi^2_{1-\alpha/2}\right) = \alpha/2$, where $F(\cdot)$ is the chi-square distribution function with $n - 1$ degrees of freedom.

   (b) Compute the standard deviation s of the sample.

   (c) Compute the test statistic $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$.

   (d) If $\chi^2 < \chi^2_{1-\alpha/2}$ or $\chi^2 > \chi^2_{\alpha/2}$, then reject the hypothesis.

   (e) If $\chi^2_{1-\alpha/2} \le \chi^2 \le \chi^2_{\alpha/2}$, then do not reject the hypothesis.

10. Test of the hypothesis $\sigma^2 = \sigma_0^2$ against the alternative $\sigma^2 > \sigma_0^2$ (or $\sigma^2 < \sigma_0^2$) of the variance of a normal distribution:

    (a) Determine the critical value χ^2_α ($\chi^2_{1-\alpha}$ for the alternative $\sigma^2 < \sigma_0^2$) such that $F\left(\chi^2_\alpha\right) = 1 - \alpha$ ($F\left(\chi^2_{1-\alpha}\right) = \alpha$), where $F(\cdot)$ is the chi-square distribution function with $n - 1$ degrees of freedom.

    (b) Compute the standard deviation s of the sample.

    (c) Compute the test statistic $\chi^2 = \dfrac{(n-1)s^2}{\sigma_0^2}$.

    (d) If $\chi^2 > \chi^2_\alpha$ ($\chi^2 < \chi^2_{1-\alpha}$), then reject the hypothesis.

    (e) If $\chi^2 \le \chi^2_\alpha$ ($\chi^2_{1-\alpha} \le \chi^2$), then do not reject the hypothesis.

7.11.2 HYPOTHESIS TESTS: PARAMETERS FROM TWO POPULATIONS

The following hypothesis tests assume a random sample of size n, given by $\{x_1, x_2, \ldots, x_n\}$, and a random sample of size m, given by $\{y_1, y_2, \ldots, y_m\}$. The term "large sample" means that the underlying distribution is approximated well enough by a normal distribution.

1. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x \neq \mu_y$ of the means of independent normal distributions with known variances σ_x^2 and σ_y^2 :
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the means, $\overline{x}$ and $\overline{y}$, of the samples.
   (c) Compute the test statistic $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$.
   (d) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.
2. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x > \mu_y$ (or $\mu_x < \mu_y$) of the means of independent normal distributions with known variances σ_x^2 and σ_y^2 :
   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the means $\overline{x}$ and $\overline{y}$ of the samples.
   (c) Compute the test statistic $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{\sigma_x^2}{n} + \frac{\sigma_y^2}{m}}}$. (For the alternative $\mu_x < \mu_y$, multiply z by -1.)
   (d) If $z > z_\alpha$, then reject the hypothesis. If $z \leq z_\alpha$, then do not reject the hypothesis.
3. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x \neq \mu_y$ of the means of independent normal distributions with unknown variances σ_x^2 and σ_y^2, large sample:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution.
   (b) Compute the means, $\overline{x}$ and $\overline{y}$, and standard deviations, s_x^2 and s_y^2, of the samples.
   (c) Compute the test statistic $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$.
   (d) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.
4. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x > \mu_y$ (or $\mu_x < \mu_y$) of the means of independent normal distributions with unknown variances, σ_x^2 and σ_y^2, large sample:

(a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.
(b) Compute the means, $\overline{x}$ and $\overline{y}$, and standard deviations, s_x^2 and s_y^2, of the samples.
(c) Compute the test statistic $z = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{s_x^2}{n} + \frac{s_y^2}{m}}}$. (For the alternative $\mu_x < \mu_y$, multiply z by -1.)
(d) If $z > z_\alpha$, then reject the hypothesis. If $z \le z_\alpha$, then do not reject the hypothesis.

5. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x \ne \mu_y$ of the means of independent normal distributions with unknown variances $\sigma_x^2 = \sigma_y^2$:

(a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the t-distribution with $n + m - 2$ degrees of freedom.
(b) Compute the means, $\overline{x}$ and $\overline{y}$, and standard deviations, s_x^2 and s_y^2, of the samples.
(c) Compute the test statistic $t = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}}$.
(d) If $|t| > t_{\alpha/2}$, then reject the hypothesis. If $|t| \le t_{\alpha/2}$, then do not reject the hypothesis.

6. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x > \mu_y$ (or $\mu_x < \mu_y$) of the means of independent normal distributions with unknown variances $\sigma_x^2 = \sigma_y^2$:

(a) Determine the critical value t_α such that $F(t_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the t-distribution with $n + m - 2$ degrees of freedom.
(b) Compute the means, $\overline{x}$ and $\overline{y}$, and standard deviations, s_x^2 and s_y^2, of the samples.
(c) Compute the test statistic $t = \dfrac{\overline{x} - \overline{y}}{\sqrt{\frac{(n-1)s_x^2 + (m-1)s_y^2}{n+m-2}}\sqrt{\frac{1}{n} + \frac{1}{m}}}$. (For the alternative $\mu_x < \mu_y$, multiply t by -1.)
(d) If $t > t_\alpha$, then reject the hypothesis. If $t \le t_\alpha$, then do not reject the hypothesis.

7. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x \ne \mu_y$ of the means of paired normal samples:

(a) Determine the critical value $t_{\alpha/2}$ so that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the t-distribution with $n - 1$ degrees of freedom.
(b) Compute the mean, $\hat{\mu}_d$, and standard deviation, s_d, of the differences $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$.
(c) Compute the test statistic $t = \dfrac{\hat{\mu}_d\sqrt{n}}{s_d}$.
(d) If $|t| > t_{\alpha/2}$, then reject the hypothesis. If $|t| \le t_{\alpha/2}$, then do not reject the hypothesis.

8. Test of the hypothesis $\mu_x = \mu_y$ against the alternative $\mu_x > \mu_y$ (or $\mu_x < \mu_y$) of the means of paired normal samples:
   (a) Determine the critical value t_α so that $F(t_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the t-distribution with $n-1$ degrees of freedom.
   (b) Compute the mean, $\hat{\mu}_d$, and standard deviation, s_d, of the differences $x_1 - y_1, x_2 - y_2, \ldots, x_n - y_n$.
   (c) Compute the test statistic $t = \dfrac{\hat{\mu}_d\sqrt{n}}{s_d}$. (For the alternative $\mu_x < \mu_y$, multiply t by -1.)
   (d) If $t > t_\alpha$, then reject the hypothesis. If $t \le t_\alpha$, then do not reject the hypothesis.

9. Test of the hypothesis $p_x = p_y$ against the alternative $p_x \ne p_y$ of the probability of success for a binomial distribution, large sample:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.
   (b) Compute the proportions, $\hat{p}_x$ and $\hat{p}_y$, of "successes" for the samples.
   (c) Compute the test statistic $z = \dfrac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_x(1-p_x)}{n} + \frac{\hat{p}_y(1-p_y)}{m}}}$.
   (d) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \le z_{\alpha/2}$, then do not reject the hypothesis.

10. Test of the hypothesis $p_x = p_y$ against the alternative $p_x > p_y$ (or $p_x < p_y$) of the probability of success for a binomial distribution, large sample:
    (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.
    (b) Compute the proportions, $\hat{p}_x$ and $\hat{p}_y$, of "successes" for the samples.
    (c) Compute the test statistic $z = \dfrac{\hat{p}_x - \hat{p}_y}{\sqrt{\frac{\hat{p}_x(1-p_x)}{n} + \frac{\hat{p}_y(1-p_y)}{m}}}$. (Multiply it by -1 for the alternative $p_x < p_y$.)
    (d) If $z > z_\alpha$, then reject the hypothesis. If $z \le z_\alpha$, then do not reject the hypothesis.

11. Mann–Whitney–Wilcoxon test of the hypothesis $M_x = M_y$ against the alternative $M_x \ne M_y$ of the medians of independent samples, large sample:
    (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution.
    (b) Pool the $N = m + n$ observations, but keep track of which sample the observation was drawn from.
    (c) Order the pooled observations from smallest to largest, assigning rank 1 to the smallest and rank N to the largest; an observation has rank r_i if it is the r_i^{th} entry in the ordered list. If two observations are equal, then assign each the average of their ranks.
    (d) Compute the sum of the ranks from the first sample T_x.

(e) Compute the test statistic $z = \dfrac{T_x - \frac{m(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}}$.

(f) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

12. Mann–Whitney–Wilcoxon test of the hypothesis $M_x = M_y$ against the alternative $M_x > M_y$ (or $M_x < M_y$) of the medians of independent samples, large sample:

(a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution.

(b) Pool the $N = m + n$ observations, but keep track of which sample the observation was drawn from.

(c) Order the pooled observations from smallest to largest, assigning rank 1 to the smallest and rank N to the largest; an observation has rank r_i if it is the r_i^{th} entry in the ordered list. If two observations are equal, then assign each the average of their ranks.

(d) Compute the sum of the ranks from the first sample T_x.

(e) Compute the test statistic $z = \dfrac{T_x - \frac{m(N+1)}{2}}{\sqrt{\frac{mn(N+1)}{12}}}$. (For the alternative $M_x < M_y$, multiply the test statistic by -1.)

(f) If $|z| > z_\alpha$, then reject the hypothesis. If $|z| \leq z_\alpha$, then do not reject the hypothesis.

13. Wilcoxon signed rank test of the hypothesis $M_x = M_y$ against the alternative $M_x \neq M_y$ of the medians of paired samples, large sample:

(a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution.

(b) Compute the paired differences $d_i = x_i - y_i$, for $i = 1, 2, \ldots, n$.

(c) Compute the quantities $|d_i|$ and keep track of the sign of d_i. If $d_i = 0$, then remove it from the list and reduce n by one.

(d) Order the $|d_i|$ from smallest to largest, assigning rank 1 to the smallest and rank n to the largest; $|d_i|$ has rank r_i if it is the r_i^{th} entry in the ordered list. In case of ties (i.e., $|d_i| = |d_j|$ for 2 or more values) assign each the average of their ranks.

(e) Compute the sum of the signed ranks $R = \displaystyle\sum_{i=1}^{n} \text{sign}(d_i)\, r_i$.

(f) Compute the test statistic $z = \dfrac{R}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$.

(g) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

14. Wilcoxon signed rank test of the hypothesis $M_x = M_y$ against the alternative $M_x > M_y$ (or $M_x < M_y$) of the medians of paired samples, large sample:

(a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution.

(b) Compute the paired differences $d_i = x_i - y_i$, for $i = 1, 2, \ldots, n$.

(c) Compute the quantities $|d_i|$ and keep track of the sign of d_i. If $d_i = 0$, then remove it from the list and reduce n by one.

(d) Order the $|d_i|$ from smallest to largest, assigning rank 1 to the smallest and rank n to the largest; $|d_i|$ has rank r_i if it is the r_i^{th} entry in the ordered list. In case of ties (i.e., $|d_i| = |d_j|$ for 2 or more values) assign each the average of their ranks.

(e) Compute the sum of the signed ranks $R = \sum_{i=1}^{n} \text{sign}\,(d_i)\, r_i$.

(f) Compute the test statistic $z = \dfrac{R}{\sqrt{\frac{n(n+1)(2n+1)}{6}}}$. (For the alternative $M_x < M_y$, multiply the test statistic by -1.)

(g) If $z > z_\alpha$, then reject the hypothesis. If $z \leq z_\alpha$, then do not reject the hypothesis.

15. Test of the hypothesis $\sigma_x^2 = \sigma_y^2$ against the alternative $\sigma_x^2 \neq \sigma_y^2$ (or $\sigma_x^2 > \sigma_y^2$) of the variances of independent normal samples:

(a) Determine the critical value $F_{\alpha/2}$ (F_α for the alternative $\sigma_x^2 > \sigma_y^2$) such that $F\left(F_{\alpha/2}\right) = 1 - \alpha/2$ ($F\left(F_\alpha\right) = 1 - \alpha$), where $F(\cdot)$ is the F-distribution function with $n - 1$ and $m - 1$ degrees of freedom.

(b) Compute the standard deviations s_x and s_y of the samples.

(c) Compute the test statistic $F = \dfrac{s_x^2}{s_y^2}$. (For the two-sided test, put the larger value in the numerator.)

(d) If $F > F_{\alpha/2}$ ($F > F_\alpha$), then reject the hypothesis. If $F \leq F_{\alpha/2}$ ($F \leq F_\alpha$), then do not reject the hypothesis.

7.11.3 HYPOTHESIS TESTS: DISTRIBUTION OF A POPULATION

The following hypothesis tests assume a random sample of size n, given by $\{x_1, x_2, \ldots, x_n\}$.

1. Run test for randomness of a sample of binary values, large sample:

(a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi(z)$ is the standard normal distribution function.

(b) Since the data are binary, denote the possible values of x_i by 0 and 1. Count the total number of zeros, and call this n_1; count the total number of ones, and call this n_2. Group the data into maximal sub-sequences of consecutive zeros and ones, and call each such sub-sequence a *run*. Let R be the number of runs in the sample.

(c) Compute $\mu_R = \dfrac{2n_1n_2}{n_1 + n_2} + 1$, and $\sigma_R^2 = \frac{(\mu_R-1)(\mu_R-2)}{n_1+n_2-1}$.

(d) Compute the test statistic $z = \dfrac{R - \mu_R}{\sigma_R}$.

(e) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

2. Run test for randomness against an alternative that a trend (an association between two variables) is present in a sample of binary values, large sample:

   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.

   (b) Since the data are binary, denote the possible values of x_i by 0 and 1. Count the total number of zeros, and call this n_1; count the total number of ones, and call this n_2. Group the data into maximal sub-sequences of consecutive zeros and ones, and call each such sub-sequence a run. Let R be the number of runs in the sample.

   (c) Compute $\mu_R = \dfrac{2n_1n_2}{n_1 + n_2} + 1$, and $\sigma_R^2 = \frac{(\mu_R - 1)(\mu_R - 2)}{n_1 + n_2 - 1}$.

   (d) Compute the test statistic $z = \dfrac{R - \mu_R}{\sigma_R}$.

   (e) If $z < -z_\alpha$, then reject the hypothesis (this suggests the presence of a trend in the data). If $z \geq -z_\alpha$, then do not reject the hypothesis.

3. Run test for randomness against an alternative that the data are periodic for a sample of binary values, large sample:

   (a) Determine the critical value z_α such that $\Phi(z_\alpha) = 1 - \alpha$, where $\Phi(z)$ is the standard normal distribution function.

   (b) Since the data are binary, denote the possible values of x_i by 0 and 1. Count the total number of zeros, and call this n_1; count the total number of ones, and call this n_2. Group the data into maximal sub-sequences of consecutive zeros and ones, and call each such sub-sequence a run. Let R be the number of runs in the sample.

   (c) Compute $\mu_R = \dfrac{2n_1n_2}{n_1 + n_2} + 1$, and $\sigma_R^2 = \frac{(\mu_R - 1)(\mu_R - 2)}{n_1 + n_2 - 1}$.

   (d) Compute the test statistic $z = \dfrac{R - \mu_R}{\sigma_R}$.

   (e) If $z > z_\alpha$, then reject the hypothesis (this suggests the data are periodic). If $z \leq z_\alpha$, then do not reject the hypothesis.

4. Chi-square test that the data are drawn from a specific k-parameter multinomial distribution, large sample:

   (a) Determine the critical value χ^2_α such that $F\left(\chi^2_\alpha\right) = 1 - \alpha$, where $F(x)$ is the chi-square distribution with $k - 1$ degrees of freedom.

   (b) The k-parameter multinomial has k possible outcomes $A_1, A_2, \ldots, A_k$ with probabilities $p_1, p_2, \ldots, p_k$. For $i = 1, 2, \ldots, k$, compute n_i, the number of x_j's corresponding to A_i.

   (c) For $i = 1, 2, \ldots, k$, compute the sample multinomial parameters $\hat{p}_i = n_i/n$.

(d) Compute the test statistic $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i}$.

(e) If $\chi^2 > \chi^2_\alpha$, then reject the hypothesis. If $\chi^2 \le \chi^2_\alpha$, then do not reject the hypothesis.

5. Chi-square test for independence of attributes A and B having possible outcomes $A_1, A_2, \ldots, A_k$ and $B_1, B_2, \ldots, B_m$:

   (a) Determine the critical value χ^2_α such that $F\left(\chi^2_\alpha\right) = 1 - \alpha$, where $F(\cdot)$ is the chi-square distribution with $(k-1)(m-1)$ degrees of freedom.

   (b) For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, m$, define o_{ij} to be the number of observations having attributes A_i and B_j, and define $o_{i\cdot} = \sum_{j=1}^{m} o_{ij}$ and $o_{\cdot j} = \sum_{i=1}^{k} o_{ij}$.

   (c) The variables defined above are often collected into a table, called a *contingency table*:

Attribute	B_1	B_2	$\cdots$	B_m	Totals
A_1	o_{11}	o_{12}	$\cdots$	o_{1m}	$o_{1\cdot}$
A_2	o_{21}	o_{22}	$\cdots$	o_{2m}	$o_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
A_k	o_{k1}	o_{k2}	$\cdots$	o_{km}	$o_{k\cdot}$
Totals	$o_{\cdot 1}$	$o_{\cdot 2}$	$\cdots$	$o_{\cdot m}$	n

   (d) For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, m$, compute the sample mean number of observations in the ij^{th} cell of the contingency table $e_{ij} = \frac{o_{i\cdot} o_{\cdot j}}{n}$.

   (e) Compute the test statistic, $\chi^2 = \sum_{i=1}^{k} \sum_{j=1}^{m} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$.

   (f) If $\chi^2 > \chi^2_\alpha$, then reject the hypothesis (that is, conclude that the attributes are not independent). If $\chi^2 \le \chi^2_\alpha$, then do not reject the hypothesis.

6. Kolmogorov–Smirnov test that $F_0(x)$ is the distribution of the population from which the sample was drawn:

   (a) Determine the critical value D_α such that $Q(D_\alpha) = 1 - \alpha$, where $Q(D)$ is the distribution function for the Kolmogorov–Smirnov test statistic D.

   (b) Compute the sample distribution function $\widehat{F}(x)$.

   (c) Compute the test statistic, given the maximum deviation of the sample and target distribution functions $D = \max \left| \widehat{F}(x) - F_0(x) \right|$.

   (d) If $D > D_\alpha$, then reject the hypothesis (that is, conclude that the data are not drawn from $F_0(x)$). If $D \le D_\alpha$, then do not reject the hypothesis.

7.11.4 HYPOTHESIS TESTS: DISTRIBUTIONS OF TWO POPULATIONS

The following hypothesis tests assume a random sample of size n, given by $\{x_1, x_2, \ldots, x_n\}$, and a random sample of size m, given by $\{y_1, y_2, \ldots, y_m\}$.

1. Chi-square test that two k-parameter multinomial distributions are equal, large sample:
   (a) Determine the critical value χ^2_α such that $F\left(\chi^2_\alpha\right) = 1 - \alpha$, where $F(\cdot)$ is the chi-square distribution with $k - 1$ degrees of freedom.
   (b) The k-parameter multinomials have k possible outcomes A_1, A_2, $\dots$, A_k. For $i = 1, 2, \dots, k$, compute n_i, the number of x_j's corresponding to A_i, and compute m_i, the number of y_j's corresponding to A_i.
   (c) For $i = 1, 2, \dots, k$, compute the sample multinomial parameters $\hat{p}_i = (n_i + m_i)/(n + m)$.
   (d) Compute the test statistic,

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n\hat{p}_i\right)^2}{n\hat{p}_i} + \sum_{i=1}^{k} \frac{\left(m_i - m\hat{p}_i\right)^2}{m\hat{p}_i}. \tag{7.11.1}$$

   (e) If $\chi^2 > \chi^2_\alpha$, then reject the hypothesis. If $\chi^2 \leq \chi^2_\alpha$, then do not reject the hypothesis.

2. Mann–Whitney–Wilcoxon test for equality of independent continuous distributions, large sample:
   (a) Determine the critical value $z_{\alpha/2}$ such that $\Phi\left(z_{\alpha/2}\right) = 1 - \alpha/2$, where $\Phi\left(z\right)$ is the normal distribution function.
   (b) For $i = 1, 2, \dots, n$ and $j = 1, 2, \dots, m$, define $S_{ij} = 1$ if $x_i < y_j$ and $S_{ij} = 0$ if $x_i > y_j$. (Continuous distributions will not produce a tie.)
   (c) Compute $U = \sum_{i=1}^{n} \sum_{j=1}^{m} S_{ij}$.
   (d) Compute the test statistic $z = \left(U - \frac{mn}{2}\right) / \sqrt{\frac{mn(m + n + 1)}{12}}$.
   (e) If $|z| > z_{\alpha/2}$, then reject the hypothesis. If $|z| \leq z_{\alpha/2}$, then do not reject the hypothesis.

3. Spearman rank correlation coefficient for independence of paired samples, large sample:
   (a) Determine the critical value $R_{\alpha/2}$ such that $F\left(R_{\alpha/2}\right) = 1 - \alpha/2$, where $F\left(R\right)$ is the distribution function for the Spearman rank correlation coefficient.
   (b) The samples are ordered, with the smallest x_i assigned the rank r_1 and the largest assigned the rank r_n; for $i = 1, 2, \dots, n$, x_i is assigned rank r_i if it occupies the i^{th} position in the ordered list. Similarly the y_i's are assigned ranks s_i. In case of a tie within a sample, the ranks are averaged.
   (c) Compute the test statistic

$$R = \frac{n \sum_{i=1}^{n} r_i s_i - \left(\sum_{i=1}^{n} r_i\right)\left(\sum_{i=1}^{n} s_i\right)}{\sqrt{\left(n \sum_{i=1}^{n} r_i^2 - \left(\sum_{i=1}^{n} r_i\right)^2\right)\left(n \sum_{i=1}^{n} s_i^2 - \left(\sum_{i=1}^{n} s_i\right)^2\right)}}.$$

   (d) If $|R| > R_{\alpha/2}$, then reject the hypothesis. If $|R| \leq R_{\alpha/2}$, then do not reject the hypothesis.

7.11.5 THE RUNS TEST

A **run** is a maximal subsequence of elements with a common property. The runs test uses the hypotheses

$$\begin{aligned} H_0: &\quad \text{the sequence is random} \\ H_a: &\quad \text{the sequence is not random} \end{aligned} \tag{7.11.2}$$

The decision variable V is the total number of runs. The hypothesis of randomness (H_0) is rejected when $V \geq v_1$ or $V \leq v_2$ where where v_1 and v_2 are critical values for the runs test.

EXAMPLE Flipping a coin gave the following sequence of heads (H) and tails (T): $\{H, H, T, T, H, T, H, T, T, T, T, H\}$. Is the coin biased? Writing the sequence as $HH \mid TT \mid H \mid T \mid H \mid TTTT \mid H \mid$ shows that there are $m = 5$ H's, $n = 7$ T's, and $V = 7$ runs.

The table (for $m = 5$ and $n = 7$) indicates that 65% of the time one would expect there to be 7 runs or fewer. The table (for $m = 5$ and $n = 6$) indicates that 42% of the time one would expect there to be 6 runs or fewer. Hence, 58% (since $1 - 0.42 = 0.58$) of the time there would be 7 runs or more.

Hence, there is no evidence to suggest the coin is biased.

m, n	$v = 2$	3	4	5	6	7	8	9
2, 2	0.3333	0.6667	1.0000					
2, 3	0.2000	0.5000	0.9000	1.0000				
2, 4	0.1333	0.4000	0.8000	1.0000				
2, 5	0.0952	0.3333	0.7143	1.0000				
2, 6	0.0714	0.2857	0.6429	1.0000				
2, 7	0.0556	0.2500	0.5833	1.0000				
2, 8	0.0444	0.2222	0.5333	1.0000				
2, 9	0.0364	0.2000	0.4909	1.0000				
3, 3	0.1000	0.3000	0.7000	0.9000	1.0000			
3, 4	0.0571	0.2000	0.5429	0.8000	0.9714	1.0000		
3, 5	0.0357	0.1429	0.4286	0.7143	0.9286	1.0000		
3, 6	0.0238	0.1071	0.3452	0.6429	0.8810	1.0000		
3, 7	0.0167	0.0833	0.2833	0.5833	0.8333	1.0000		
3, 8	0.0121	0.0667	0.2364	0.5333	0.7879	1.0000		
3, 9	0.0091	0.0545	0.2000	0.4909	0.7455	1.0000		
4, 4	0.0286	0.1143	0.3714	0.6286	0.8857	0.9714	1.0000	
4, 5	0.0159	0.0714	0.2619	0.5000	0.7857	0.9286	0.9921	1.0000
4, 6	0.0095	0.0476	0.1905	0.4048	0.6905	0.8810	0.9762	1.0000
4, 7	0.0061	0.0333	0.1424	0.3333	0.6061	0.8333	0.9545	1.0000
4, 8	0.0040	0.0242	0.1091	0.2788	0.5333	0.7879	0.9293	1.0000
4, 9	0.0028	0.0182	0.0853	0.2364	0.4713	0.7455	0.9021	1.0000

m, n	$v = 2$	3	4	5	6	7	8	9	10
5, 5	0.0079	0.0397	0.1667	0.3571	0.6429	0.8333	0.9603	0.9921	1.0000
5, 6	0.0043	0.0238	0.1104	0.2619	0.5216	0.7381	0.9113	0.9762	0.9978
5, 7	0.0025	0.0152	0.0758	0.1970	0.4242	0.6515	0.8535	0.9545	0.9924
5, 8	0.0016	0.0101	0.0536	0.1515	0.3473	0.5758	0.7933	0.9293	0.9837
5, 9	0.0010	0.0070	0.0390	0.1189	0.2867	0.5105	0.7343	0.9021	0.9720
6, 6	0.0022	0.0130	0.0671	0.1753	0.3918	0.6082	0.8247	0.9329	0.9870
6, 7	0.0012	0.0076	0.0425	0.1212	0.2960	0.5000	0.7331	0.8788	0.9662
6, 8	0.0007	0.0047	0.0280	0.0862	0.2261	0.4126	0.6457	0.8205	0.9371
6, 9	0.0004	0.0030	0.0190	0.0629	0.1748	0.3427	0.5664	0.7622	0.9021
6, 10	0.0002	0.0020	0.0132	0.0470	0.1369	0.2867	0.4965	0.7063	0.8636
7, 7	0.0006	0.0041	0.0251	0.0775	0.2086	0.3834	0.6166	0.7914	0.9225
7, 8	0.0003	0.0023	0.0154	0.0513	0.1492	0.2960	0.5136	0.7040	0.8671
7, 9	0.0002	0.0014	0.0098	0.0350	0.1084	0.2308	0.4266	0.6224	0.8059
8, 8	0.0002	0.0012	0.0089	0.0317	0.1002	0.2145	0.4048	0.5952	0.7855
8, 9	$.0^4823$	0.0007	0.0053	0.0203	0.0687	0.1573	0.3186	0.5000	0.7016

When $m = n$ and $m > 10$ the following table can be used. The columns headed 0.5, 1, 2.5, and 5 give values of v such that v or fewer runs occur with probability less than the indicated percentage. For example, for $m = n = 12$, the probability of 8 or fewer runs is approximately 5%. The columns headed 95, 97.5, 99, and 99.5 give values of v for which the probability of v or more runs is less than 5, 2.5, 1, or 0.5%. The last columns describe the expected number of runs (μ, σ).

$m = n$	0.5	1.0	2.5	5.0	95.0	97.5	99.0	99.5	μ	σ^2	σ
11	5	6	7	7	16	16	17	18	12	5.24	2.29
12	6	7	7	8	17	18	18	19	13	5.74	2.40
13	7	7	8	9	18	19	20	20	14	6.24	2.50
14	7	8	9	10	19	20	21	22	15	6.74	2.60
15	8	9	10	11	20	21	22	23	16	7.24	2.69
20	12	13	14	15	26	27	28	29	21	9.74	3.12
25	16	17	18	19	32	33	34	35	26	12.24	3.50
30	20	21	23	24	37	38	40	41	31	14.75	3.84
40	29	30	31	33	48	50	51	52	41	19.75	4.44
50	37	38	40	42	59	61	63	64	51	24.75	4.97
60	46	47	49	51	70	72	74	75	61	29.75	5.45
70	55	56	58	60	81	83	85	86	71	34.75	5.89
80	64	65	68	70	91	93	96	97	81	39.75	6.30
90	73	74	77	79	102	104	107	108	91	44.75	6.69
100	82	84	86	88	113	115	117	119	101	49.75	7.05

7.11.6 SEQUENTIAL PROBABILITY RATIO TESTS

When using a sequential probability ratio test, the number of samples taken is not fixed a priori, but determined as sampling occurs. Given two simple hypotheses and m observations, compute:

1. $P_{0m} = \text{Prob (observations} \mid H_0)$.
2. $P_{1m} = \text{Prob (observations} \mid H_1)$.
3. $v_m = P_{1m}/P_{0m}$.

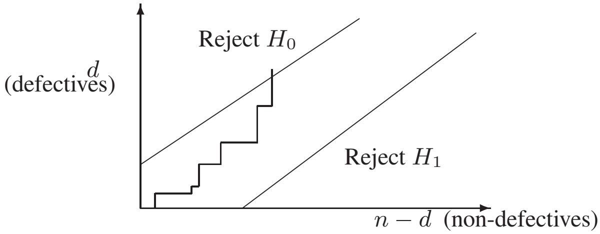
And then make one of the following decisions:

1. If $v_m \geq \dfrac{1-\beta}{\alpha}$ then reject H_0.
2. If $v_m \leq \dfrac{\beta}{1-\alpha}$ then reject H_1.
3. If $\dfrac{\beta}{1-\alpha} < v_m < \dfrac{1-\beta}{\alpha}$ then make another observation.

EXAMPLES

1. Let X be normally distributed with unknown mean μ and known standard deviation σ. Consider the two simple hypotheses, $H_0 : \mu = \mu_0$ and $H_1 : \mu = \mu_1$. If Y is the sum of the first m observations of X, then a (Y, m) control chart is constructed with the two lines:
$$\begin{aligned} Y &= \frac{\mu_0+\mu_1}{2}m + \frac{\sigma^2}{\mu_1-\mu_0}\log\frac{\beta}{1-\alpha}, \\ Y &= \frac{\mu_0+\mu_1}{2}m + \frac{\sigma^2}{\mu_1-\mu_0}\log\frac{1-\beta}{\alpha}. \end{aligned} \tag{7.11.3}$$

2. Let θ denote the fraction of defective items. Two simple hypotheses are H_0: $\theta = \theta_0 = 0.05$ and H_1: $\theta = \theta_1 = 0.15$. Choose $\alpha = 5\%$ and $\beta = 10\%$ (i.e., reject lot with $\theta = \theta_0$ about 5% of the time; accept lot with $\theta = \theta_1$ about 10% of the time). If, after m observations, there are d defective items, then
$$P_{im} = \binom{m}{d}\theta_i^d(1-\theta_i)^{m-d} \qquad \text{and} \qquad v_m = \left(\frac{\theta_1}{\theta_0}\right)^d \left(\frac{1-\theta_1}{1-\theta_0}\right)^{m-d} \tag{7.11.4}$$
or $v_m = 3^d(0.895)^{m-d}$, with the above numbers. We find $\frac{\beta}{1-\alpha} = 0.105$ and $\frac{1-\beta}{\alpha} = 18$. The decision to perform another observation depends on whether or not
$$0.105 \leq 3^d(0.895)^{m-d} \leq 18. \tag{7.11.5}$$
Taking logarithms, a $(m-d, d)$ control chart can be drawn with the following lines: $d = 0.101(m-d) - 2.049$ and $d = 0.101(m-d) + 2.63$. On the figure below, a sample path leading to rejection of H_0 has been indicated:



7.12 LINEAR REGRESSION

1. The general linear statistical model assumes that the observed data values $\{y_1, y_2, \ldots, y_m\}$ are of the form
$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \epsilon_i,$$
for $i = 1, 2, \ldots, m$.
2. For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, the independent variables x_{ij} are known (nonrandom).
3. $\{\beta_0, \beta_1, \beta_2, \cdots, \beta_n\}$ are unknown parameters.
4. For each i, ϵ_i is a zero-mean normal random variable with unknown variance σ^2.

7.12.1 LINEAR MODEL $y_i = \beta_0 + \beta_1 x_i + \epsilon$

1. Point estimate of β_1:
$$\widehat{\beta}_1 = \frac{m \sum_{i=1}^{m} x_i y_i - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} y_i\right)}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}.$$
2. Point estimate of β_0:
$$\widehat{\beta}_0 = \bar{y} - \widehat{\beta}_1 \overline{x}.$$
3. Point estimate of the correlation coefficient:
$$r = \widehat{\rho} = \frac{m \sum_{i=1}^{m} x_i y_i - \left(\sum_{i=1}^{m} x_i\right)\left(\sum_{i=1}^{m} y_i\right)}{\sqrt{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}\sqrt{m\left(\sum_{i=1}^{m} y_i^2\right) - \left(\sum_{i=1}^{m} y_i\right)^2}}.$$
4. Point estimate of error variance σ^2: $\widehat{\sigma^2} = \dfrac{\sum_{i=1}^{m}\left(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i\right)^2}{m-2}$.
5. The standard error of the estimate is defined as $s_e = \sqrt{\widehat{\sigma^2}}$.
6. Least-squares regression line: $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$.
7. Confidence interval for β_0:
   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m-2$ degrees of freedom.

(b) Compute the point estimate $\widehat{\beta}_0$.

(c) Compute $k = t_{\alpha/2} s_e \sqrt{\dfrac{1}{m} + \dfrac{\overline{x}^2}{\sum_{i=1}^{m} (x_i - \overline{x})^2}}$.

(d) The $100(1-\alpha)$ percent confidence interval for β_0 is given by $\left[\widehat{\beta}_0 - k, \widehat{\beta}_0 + k\right]$.

8. Confidence interval for β_1:

(a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m-2$ degrees of freedom.

(b) Compute the point estimate $\widehat{\beta}_1$.

(c) Compute $k = t_{\alpha/2} \dfrac{s_e}{\sqrt{\sum_{i=1}^{m} (x_i - \overline{x})^2}}$.

(d) The $100(1-\alpha)$ percent confidence interval for β_1 is given by $\left[\widehat{\beta}_1 - k, \widehat{\beta}_1 + k\right]$.

9. Confidence interval for σ^2:

(a) Determine the critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ such that $F\left(\chi^2_{\alpha/2}\right) = 1 - \alpha/2$ and $F\left(\chi^2_{1-\alpha/2}\right) = \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the chi-square distribution function with $m-2$ degrees of freedom.

(b) Compute the point estimate $\widehat{\sigma^2}$.

(c) Compute $k_1 = \dfrac{(n-2)\widehat{\sigma^2}}{\chi^2_{\alpha/2}}$ and $k_2 = \dfrac{(n-2)\widehat{\sigma^2}}{\chi^2_{1-\alpha/2}}$.

(d) The $100(1-\alpha)$ percent confidence interval for σ^2 is given by $[k_1, k_2]$.

10. Confidence interval (predictive interval) for y, given x_0:

(a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m-2$ degrees of freedom.

(b) Compute the point estimates $\widehat{\beta}_0$, $\widehat{\beta}_1$, and s_e.

(c) Compute $k = t_{\alpha/2} s_e \sqrt{\dfrac{1}{m} + \dfrac{(x_0 - \overline{x})^2}{\sum_{i=1}^{m} (x_i - \overline{x})^2}}$ and $\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x_0$.

(d) The $100(1-\alpha)$ percent confidence interval for β_1 is given by $[\widehat{y} - k, \widehat{y} + k]$.

11. Test of the hypothesis $\beta_1 = 0$ against the alternative $\beta_1 \neq 0$:
    (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m-2$ degrees of freedom.
    (b) Compute the point estimates $\widehat{\beta}_1$ and s_e.
    (c) Compute the test statistic $t = \dfrac{\widehat{\beta}_1}{s_e}\sqrt{\displaystyle\sum_{i=1}^{m}(x_i - \overline{x})^2}$.
    (d) If $|t| > t_{\alpha/2}$, then reject the hypothesis. If $|t| \leq t_{\alpha/2}$, then do not reject the hypothesis.

7.12.2 GENERAL MODEL $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n + \epsilon$

1. The m equations ($i = 1, 2, \ldots, m$)

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_n x_{in} + \epsilon_i \tag{7.12.1}$$

can be written in matrix notation as $\mathbf{y} = \mathbf{X}\beta + \epsilon$ where

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}, \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \qquad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}, \tag{7.12.2}$$

and

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1n} \\ 1 & x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}. \tag{7.12.3}$$

2. Throughout the remainder of the section, we assume $\mathbf{X}$ has full column rank (see page 93).
3. The least-squares estimate $\widehat{\beta}$ satisfies the *normal equations* $\mathbf{X}^{\mathrm{T}}\mathbf{X}\widehat{\beta} = \mathbf{X}^{\mathrm{T}}\mathbf{y}$. That is, $\widehat{\beta} = \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$.
4. Point estimate of σ^2:

$$\widehat{\sigma^2} = \frac{1}{m-n-1}\left(\mathbf{y}^{\mathrm{T}}\mathbf{y} - \widehat{\beta}^{\mathrm{T}}(\mathbf{X}^{\mathrm{T}}\mathbf{y})\right).$$

5. The standard error of the estimate is defined as $s_e = \sqrt{\widehat{\sigma^2}}$.
6. Least-squares regression line: $\widehat{y} = \mathbf{x}^{\mathrm{T}}\widehat{\beta}$.
7. In the following, let c_{ij} denote the $(i,j)^{\text{th}}$ entry in the matrix $\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}$.
8. Confidence interval for β_i:

(a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m - n - 1$ degrees of freedom.
(b) Compute the point estimate $\widehat{\beta}_i$ by solving the normal equations, and compute s_e.
(c) Compute $k_i = t_{\alpha/2} s_e \sqrt{c_{ii}}$.
(d) The $100(1 - \alpha)$ percent confidence interval for β_i is given by $\left[\widehat{\beta}_i - k_i, \widehat{\beta}_i + k_i\right]$.

9. Confidence interval for σ^2:
   (a) Determine the critical values $\chi^2_{\alpha/2}$ and $\chi^2_{1-\alpha/2}$ such that $F\left(\chi^2_{\alpha/2}\right) = 1 - \alpha/2$ and $F\left(\chi^2_{1-\alpha/2}\right) = \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the chi-square distribution function with $m - n - 1$ degrees of freedom.
   (b) Compute the point estimate $\widehat{\sigma^2}$.
   (c) Compute $k_1 = \dfrac{(m-n-1)\widehat{\sigma^2}}{\chi^2_{\alpha/2}}$ and $k_2 = \dfrac{(m-n-1)\widehat{\sigma^2}}{\chi^2_{1-\alpha/2}}$.
   (d) The $100(1 - \alpha)$ percent confidence interval for σ^2 is given by $[k_1, k_2]$.

10. Confidence interval (predictive interval) for y, given $\mathbf{x}_0$:
    (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $n - m - 1$ degrees of freedom.
    (b) Compute the point estimate $\widehat{\beta}_i$ by solving the normal equations, and compute s_e.
    (c) Compute $k = t_{\alpha/2}\, s_e \sqrt{1 + \mathbf{x}_0^{\mathrm{T}} \left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1} \mathbf{x}_0}$ and $\widehat{y} = \mathbf{x}_0^{\mathrm{T}}\widehat{\beta}$.
    (d) The $100(1 - \alpha)$ percent confidence interval for y_0 is given by $[\widehat{y} - k, \widehat{y} + k]$.

11. Test of the hypothesis $\beta_i = 0$ against the alternative $\beta_i \neq 0$:
    (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $m - n - 1$ degrees of freedom.
    (b) Compute the point estimates $\widehat{\beta}_i$ and s_e by solving the normal equations.
    (c) Compute the test statistic $t = \dfrac{\widehat{\beta}_i}{s_e\sqrt{c_{ii}}}$.
    (d) If $|t| > t_{\alpha/2}$, then reject the hypothesis. If $|t| \leq t_{\alpha/2}$, then do not reject the hypothesis.

7.13 ANALYSIS OF VARIANCE (ANOVA)

Analysis of variance (ANOVA) is a statistical methodology for determining information about means. The analysis uses variances both between and within samples.

7.13.1 ONE-FACTOR ANOVA

1. Suppose we have k samples from k populations, with the j^{th} population consisting of n_j observations,
$$\begin{gathered} y_{11}, y_{21}, \ldots, y_{n_1 1} \\ y_{12}, y_{22}, \ldots, y_{n_2 2} \\ \vdots \\ y_{1k}, y_{2k}, \ldots, y_{n_k k}. \end{gathered}$$
2. One-factor model:
   (a) The one-factor ANOVA assumes that the i^{th} observation from the j^{th} sample is of the form $y_{ij} = \mu + \tau_j + e_{ij}$.
   (b) For $j = 1, 2, \ldots, k$, the parameter $\mu_j = \mu + \tau_j$ is the unknown mean of the j^{th} population, and $\sum_{j=1}^{k} \tau_j = 0$.
   (c) For $i = 1, 2, \ldots, k$ and $j = 1, 2, \ldots, n_j$, the random variables e_{ij} are independent and normally distributed with mean zero and variance σ^2.
   (d) The total number of observations is $n = n_1 + n_2 + \cdots + n_k$.
3. Point estimates of means:
   (a) Total sample mean $$\widehat{y} = \frac{1}{n} \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}.$$
   (b) Sample mean of j^{th} sample $$\widehat{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}.$$
4. Sums of squares:
   (a) Sum of squares between samples $$\text{SS}_\text{b} = \sum_{j=1}^{k} n_j \left(\widehat{y}_j - \widehat{y}\right)^2.$$
   (b) Sum of squares within samples $$\text{SS}_\text{w} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(y_{ij} - \widehat{y}_j\right)^2.$$
   (c) Total sum of squares $$\text{Total SS} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(y_{ij} - \widehat{y}\right)^2.$$
   (d) Partition of total sum of squares $$\text{Total SS} = \text{SS}_\text{b} + \text{SS}_\text{w}.$$
5. Degrees of freedom:
   (a) Between samples, $k - 1$.
   (b) Within samples, $n - k$.
   (c) Total, $n - 1$.

6. Mean squares:
   (a) Obtained by dividing sums of squares by their respective degrees of freedom.
   (b) Between samples, $\text{MS}_\text{b} = \dfrac{\text{SS}_\text{b}}{k-1}$.
   (c) Within samples (also called the *residual mean square*), $\text{MS}_\text{w} = \dfrac{\text{SS}_\text{w}}{n-k}$.

7. Test of the hypothesis $\mu_1 = \mu_2 = \cdots = \mu_k$ against the alternative $\mu_i \neq \mu_j$ for some i and j; equivalently, test the null hypothesis $\tau_1 = \tau_2 = \cdots = \tau_k = 0$ against the hypothesis $\tau_j \neq 0$ for some j:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $k-1$ and $n-k$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$ and $\widehat{y}_j$ for $j = 1, 2, \ldots, k$.
   (c) Compute the sums of squares SS_b and SS_w.
   (d) Compute the mean squares MS_b and MS_w.
   (e) Compute the test statistic $F = \dfrac{\text{MS}_\text{b}}{\text{MS}_\text{w}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis.
   If $F \leq F_\alpha$, then do not reject the hypothesis.
   (g) The above computations are often organized into an ANOVA table:

Source	SS	D.O.F.	MS	F Ratio
Between samples	SS_b	$k-1$	MS_b	$F = \frac{\text{MS}_\text{b}}{\text{MS}_\text{w}}$
Within samples	SS_w	$n-k$	MS_w	
Total	Total SS	$n-1$		

8. Confidence interval for $\mu_i - \mu_j$, for $i \neq j$:
   (a) Determine the critical value $t_{\alpha/2}$ such that $F\left(t_{\alpha/2}\right) = 1 - \alpha/2$, where $F(\cdot)$ is the cumulative distribution function for the t-distribution with $n-k$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}_i$ and $\widehat{y}_j$.
   (c) Compute the residual mean square MS_w.
   (d) Compute $k = t_{\alpha/2}\sqrt{\text{MS}_\text{w}\left(\dfrac{1}{n_i} + \dfrac{1}{n_j}\right)}$.
   (e) The $100(1-\alpha)$ percent confidence interval for $\mu_i - \mu_j$ is given by $[(\widehat{y}_i - y_j) - k, (\widehat{y}_i - y_j) + k]$.

9. Confidence interval for contrast in the means, defined by $C = c_1\mu_1 + c_2\mu_2 + \cdots + c_k\mu_k$, where $c_1 + c_2 + \cdots + c_k = 0$:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $k-1$ and $n-k$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}_j$ for $j = 1, 2, \ldots, k$.

(c) Compute the residual mean square MS_w.

(d) Compute $k = \sqrt{F_\alpha \text{MS}_\text{w} \left(\frac{k-1}{n} \sum_{j=1}^{k} c_j^2 \right)}$.

(e) The $100(1-\alpha)$ percent confidence interval for the contrast C is $\left[\sum_{j=1}^{k} c_j \widehat{y}_j - k, \sum_{j=1}^{k} c_j \widehat{y}_j + k \right]$.

7.13.2 UNREPLICATED TWO-FACTOR ANOVA

1. Suppose we have a sample of observations y_{ij} indexed by two factors $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.
2. Unreplicated two-factor model:
   (a) The unreplicated two-factor ANOVA assumes that the ij^{th} observation is of the form $y_{ij} = \mu + \beta_i + \tau_j + e_{ij}$.
   (b) μ is the overall mean, β_i is the i^{th} differential effect of factor one, τ_j is the j differential effect of factor two, and

$$\sum_{i=1}^{m} \beta_i = \sum_{j=1}^{n} \tau_j = 0.$$

   (c) For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, the random variables e_{ij} are independent and normally distributed with mean zero and variance σ^2.
   (d) Total number of observations is mn.

3. Point estimates of means:
   (a) Total sample mean $\widehat{y} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{j=1}^{n} y_{ij}$.
   (b) i^{th} factor-one sample mean $\widehat{y}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} y_{ij}$.
   (c) j^{th} factor-two sample mean $\widehat{y}_{\cdot j} = \frac{1}{m} \sum_{i=1}^{m} y_{ij}$.
4. Sums of squares:
   (a) Factor-one sum of squares $\text{SS}_1 = n \sum_{i=1}^{m} \left(\widehat{y}_{i\cdot} - \widehat{y} \right)^2$.
   (b) Factor-two sum of squares $\text{SS}_2 = m \sum_{j=1}^{n} \left(\widehat{y}_{\cdot j} - \widehat{y} \right)^2$.
   (c) Residual sum of squares $\text{SS}_\text{r} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left(y_{ij} - \widehat{y}_{i\cdot} - \widehat{y}_{\cdot j} + \widehat{y} \right)^2$.
   (d) Total sum of squares $\text{Total SS} = \sum_{j=1}^{n} \sum_{i=1}^{m} \left(y_{ij} - \widehat{y} \right)^2$.
   (e) Partition of total sum of squares $\text{Total SS} = \text{SS}_1 + \text{SS}_2 + \text{SS}_\text{r}$.

5. Degrees of freedom:
   (a) Factor one, $m-1$.
   (b) Factor two, $n-1$.
   (c) Residual, $(m-1)(n-1)$.
   (d) Total, $mn-1$.

6. Mean squares:

(a) Obtained by dividing sums of squares by their respective degrees of freedom.
(b) Factor-one mean square $\mathrm{MS}_1 = \frac{\mathrm{SS}_1}{m-1}$.
(c) Factor-two mean square $\mathrm{MS}_2 = \frac{\mathrm{SS}_2}{n-1}$.
(d) Residual mean square $\mathrm{MS}_\mathrm{r} = \frac{\mathrm{SS}_\mathrm{r}}{(m-1)(n-1)}$.

7. Test of the null hypothesis $\beta_1 = \beta_2 = \cdots = \beta_m = 0$ (no factor-one effects) against the alternative hypothesis $\beta_i \neq 0$ for some i:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $m-1$ and $(m-1)(n-1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$ and $\widehat{y}_{i\cdot}$ for $i = 1, 2, \ldots, m$.
   (c) Compute the sums of squares SS_1 and SS_r.
   (d) Compute the mean squares MS_1 and MS_r.
   (e) Compute the test statistic $F = \frac{\mathrm{MS}_1}{\mathrm{MS}_\mathrm{r}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis.
   If $F \leq F_\alpha$, then do not reject the hypothesis.

8. Test of the null hypothesis $\tau_1 = \tau_2 = \cdots = \tau_n = 0$ (no factor-two effects) against the alternative hypothesis $\tau_j \neq 0$ for some j:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $n-1$ and $(m-1)(n-1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$ and $\widehat{y}_{\cdot j}$ for $j = 1, 2, \ldots, n$.
   (c) Compute the sums of squares SS_2 and SS_r.
   (d) Compute the mean squares MS_2 and MS_r.
   (e) Compute the test statistic $F = \frac{\mathrm{MS}_2}{\mathrm{MS}_\mathrm{r}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis. If $F \leq F_\alpha$, then do not reject the hypothesis.
   (g) The above computations are often organized into an ANOVA table:

Source	SS	D.O.F.	MS	F Ratio
Factor one	SS_1	$m-1$	MS_1	$F = \frac{\mathrm{MS}_1}{\mathrm{MS}_\mathrm{r}}$
Factor two	SS_2	$n-1$	MS_2	$F = \frac{\mathrm{MS}_2}{\mathrm{MS}_\mathrm{r}}$
Residual	SS_r	$(m-1)(n-1)$	MS_r	
Total	Total SS	$mn-1$		

9. Confidence interval for contrast in the factor-one means, defined by $C = c_1\beta_1 + c_2\beta_2 + \cdots + c_m\beta_m$, where $c_1 + c_2 + \cdots + c_m = 0$:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $m-1$ and $(m-1)(n-1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}_{i\cdot}$ for $i = 1, 2, \ldots, m$.
   (c) Compute the residual mean square MS_r.
   (d) Compute $k = \sqrt{F_\alpha \mathrm{MS}_\mathrm{r} \left(\frac{m-1}{n} \sum_{i=1}^{m} c_i^2\right)}$.

(e) The $100(1-\alpha)$ percent confidence interval for the contrast C is $\left[\sum_{i=1}^{m} c_i \widehat{y}_{i\cdot} - k, \sum_{i=1}^{m} c_i \widehat{y}_{i\cdot} + k\right]$.

10. Confidence interval for contrast in the factor-two means, defined by $C = c_1\tau_1 + c_2\tau_2 + \cdots + c_n\tau_n$, where $c_1 + c_2 + \cdots + c_n = 0$:
    (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $n-1$ and $(m-1)(n-1)$ degrees of freedom.
    (b) Compute the point estimates $\widehat{y}_{\cdot j}$ for $j = 1, 2, \ldots, n$.
    (c) Compute the residual mean square MS_r.
    (d) Compute $k = \sqrt{F_\alpha \text{MS}_\text{r} \left(\frac{n-1}{m} \sum_{j=1}^{n} c_j^2\right)}$.
    (e) The $100(1-\alpha)$ percent confidence interval for the contrast C is $\left[\sum_{j=1}^{n} c_j \widehat{y}_{\cdot j} - k, \sum_{j=1}^{n} c_j \widehat{y}_{\cdot j} + k\right]$.

7.13.3 REPLICATED TWO-FACTOR ANOVA

1. Suppose we have a sample of observations y_{ijk} indexed by two factors $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. Moreover, there are p observations per factor pair (i, j), indexed by $k = 1, 2, \ldots, p$.
2. Replicated two-factor model:
    (a) The replicated two-factor ANOVA assumes that the ijk^{th} observation is of the form $y_{ijk} = \mu + \beta_i + \tau_j + \gamma_{ij} + e_{ijk}$.
    (b) μ is the overall mean, β_i is the i^{th} differential effect of factor one, τ_j is the j differential effect of factor two, and

$$\sum_{i=1}^{m} \beta_i = \sum_{j=1}^{n} \tau_j = 0.$$

    (c) For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, γ_{ij} is the ij^{th} interaction effect of factors one and two.
    (d) For $i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n$, and $k = 1, 2, \ldots, p$, the random variables e_{ijk} are independent and normally distributed with mean zero and variance σ^2.
    (e) Total number of observations is mnp.
3. Point estimates of means:
    (a) Total sample mean $\widehat{y} = \frac{1}{mnp} \sum_{i=1}^{m} \sum_{j=1}^{n} \sum_{k=1}^{p} y_{ijk}$.
    (b) i^{th} factor-one sample mean $\widehat{y}_{i\cdot\cdot} = \frac{1}{np} \sum_{j=1}^{n} \sum_{k=1}^{p} y_{ijk}$.
    (c) j^{th} factor-two sample mean $\widehat{y}_{\cdot j\cdot} = \frac{1}{mp} \sum_{i=1}^{m} \sum_{k=1}^{p} y_{ijk}$.
    (d) ij^{th} interaction mean $\widehat{y}_{ij\cdot} = \frac{1}{p} \sum_{k=1}^{p} y_{ijk}$.
4. Sums of squares:
    (a) Factor-one sum of squares $\text{SS}_1 = np \sum_{i=1}^{m} \left(\widehat{y}_{i\cdot\cdot} - \widehat{y}\right)^2$.
    (b) Factor-two sum of squares $\text{SS}_2 = mp \sum_{j=1}^{n} \left(\widehat{y}_{\cdot j\cdot} - \hat{y}\right)^2$.

(c) Interaction sum of squares $\text{SS}_{12} = p\sum_{i=1}^{m}\sum_{j=1}^{n}\left(\widehat{y}_{ij\cdot} - \widehat{y}\right)^2$.

(d) Residual sum of squares

$$\text{SS}_\text{r} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}\left(y_{ijk} - \widehat{y}_{i\cdot\cdot} - \widehat{y}_{\cdot j\cdot} + \widehat{y}\right)^2.$$

(e) Total sum of squares $\text{Total SS} = \sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{p}\left(y_{ijk} - \widehat{y}\right)^2$.

(f) Partition of total sum of squares $\text{Total SS} = \text{SS}_1 + \text{SS}_2 + \text{SS}_{12} + \text{SS}_\text{r}$.

5. Degrees of freedom:

   (a) Factor one, $m - 1$.
   (b) Factor two, $n - 1$.
   (c) Interaction, $(m - 1)(n - 1)$.
   (d) Residual, $mn(p - 1)$.
   (e) Total, $mnp - 1$.

6. Mean squares:

   (a) Obtained by dividing sums of squares by their respective degrees of freedom.
   (b) Factor-one mean square $\text{MS}_1 = \frac{\text{SS}_1}{m-1}$.
   (c) Factor-two mean square $\text{MS}_2 = \frac{\text{SS}_2}{n-1}$.
   (d) Interaction mean square $\text{MS}_{12} = \frac{\text{SS}_{12}}{(m-1)(n-1)}$.
   (e) Residual mean square $\text{MS}_\text{r} = \frac{\text{SS}_\text{r}}{mn(p-1)}$.

7. Test of the null hypothesis $\beta_1 = \beta_2 = \cdots = \beta_m = 0$ (no factor-one effects) against the alternative hypothesis $\beta_i \neq 0$ for some i:

   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $m - 1$ and $mn(p - 1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$ and $\widehat{y}_{i\cdot\cdot}$ for $i = 1, 2, \ldots, m$.
   (c) Compute the sums of squares SS_1 and SS_r.
   (d) Compute the mean squares MS_1 and MS_r.
   (e) Compute the test statistic $F = \frac{\text{MS}_1}{\text{MS}_\text{r}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis.
   If $F \leq F_\alpha$, then do not reject the hypothesis.

8. Test of the null hypothesis $\tau_1 = \tau_2 = \cdots = \tau_n = 0$ (no factor-two effects) against the alternative hypothesis $\tau_j \neq 0$ for some j:

   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $n - 1$ and $mn(p - 1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$ and $\widehat{y}_{\cdot j\cdot}$ for $j = 1, 2, \ldots, n$.
   (c) Compute the sums of squares SS_2 and SS_r.
   (d) Compute the mean squares MS_2 and MS_r.
   (e) Compute the test statistic $F = \frac{\text{MS}_2}{\text{MS}_\text{r}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis.
   If $F \leq F_\alpha$, then do not reject the hypothesis.

9. Test of the null hypothesis $\gamma_{ij} = 0$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$ (no factor-one effects) against the alternative hypothesis $\gamma_{ij} \neq 0$ for some i and j:
   (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $(m-1)(n-1)$ and $mn(p-1)$ degrees of freedom.
   (b) Compute the point estimates $\widehat{y}$, $\widehat{y}_{i\cdot\cdot}$, $\widehat{y}_{\cdot j\cdot}$, and $\widehat{y}_{ij\cdot}$ for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.
   (c) Compute the sums of squares SS_{12} and SS_r.
   (d) Compute the mean squares MS_{12} and MS_r.
   (e) Compute the test statistic $F = \frac{\text{MS}_{12}}{\text{MS}_\text{r}}$.
   (f) If $F > F_\alpha$, then reject the hypothesis.
   If $F \leq F_\alpha$, then do not reject the hypothesis.
   (g) The above computations are often organized into an ANOVA table:

Source	SS	D.O.F.	MS	F Ratio
Factor one	SS_1	$m-1$	MS_1	$F = \text{MS}_1/\text{MS}_\text{r}$
Factor two	SS_2	$n-1$	MS_2	$F = \text{MS}_2/\text{MS}_\text{r}$
Interaction	SS_{12}	$(m-1)(n-1)$	MS_{12}	$F = \text{MS}_{12}/\text{MS}_\text{r}$
Residual	SS_r	$mn(p-1)$	MS_r	
Total	Total SS	$mnp-1$		

10. Confidence interval for contrast in the factor-one means, defined by $C = c_1\beta_1 + c_2\beta_2 + \cdots + c_m\beta_m$, where $c_1 + c_2 + \cdots + c_m = 0$:
    (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $m-1$ and $mn(p-1)$ degrees of freedom.
    (b) Compute the point estimates $\widehat{y}_{i\cdot\cdot}$ for $i = 1, 2, \ldots, m$.
    (c) Compute the residual mean square MS_r.
    (d) Compute $k = \sqrt{F_\alpha \text{MS}_\text{r} \left(\frac{m-1}{np} \sum_{i=1}^m c_i^2\right)}$.
    (e) The $100(1-\alpha)$ percent confidence interval for the contrast C is $\left[\sum_{i=1}^m c_i \widehat{y}_{i\cdot\cdot} - k, \sum_{i=1}^m c_i \widehat{y}_{i\cdot\cdot} + k\right]$.

11. Confidence interval for contrast in the factor-two means, defined by $C = c_1\tau_1 + c_2\tau_2 + \cdots + c_n\tau_n$, where $c_1 + c_2 + \cdots + c_n = 0$:
    (a) Determine the critical value F_α such that $F(F_\alpha) = 1 - \alpha$, where $F(\cdot)$ is the cumulative distribution function for the F-distribution with $n-1$ and $mn(p-1)$ degrees of freedom.
    (b) Compute the point estimates $\widehat{y}_{\cdot j\cdot}$ for $j = 1, 2, \ldots, n$.
    (c) Compute the residual mean square MS_r.
    (d) Compute $k = \sqrt{F_\alpha \text{MS}_\text{r} \left(\frac{n-1}{mp} \sum_{j=1}^n c_j^2\right)}$.
    (e) The $100(1-\alpha)$ percent confidence interval for the contrast C is $\left[\sum_{j=1}^n c_j \widehat{y}_{\cdot j\cdot} - k, \sum_{j=1}^n c_j \widehat{y}_{\cdot j\cdot} + k\right]$.

7.14 SAMPLE SIZE

7.14.1 CONFIDENCE INTERVALS

To construct a confidence interval of specified width, a priori parameter estimates and a bound on the error of estimation may be used to determine necessary sample sizes. For a $100(1-\alpha)\%$ confidence interval, let E = error of estimation (half the width of the confidence interval). The following table presents some common sample size calculations.

Parameter	Estimate	Sample size
μ	$\overline{x}$	$n = \left(\frac{z_{\alpha/2} \cdot \sigma}{E}\right)^2$
p	$\hat{p}$	$n = \frac{(z_{\alpha/2})^2 \cdot pq}{E^2}$
$\mu_1 - \mu_2$	$\overline{x}_1 - \overline{x}_2$	$n_1 = n_2 = \frac{(z_{\alpha/2})^2(\sigma_1^2 + \sigma_2^2)}{E^2}$
$p_1 - p_2$	$\hat{p}_1 - \hat{p}_2$	$n_1 = n_2 = \frac{(z_{\alpha/2})^2(p_1q_1 + p_2q_2)}{E^2}$

EXAMPLE An experiment will estimate the probability of a success, p, in a binomial distribution. How large a sample is needed to estimate this proportion to within 5% with 99% confidence? That is, what value of n satisfies Prob $(|\hat{p}-p| \leq 0.05) \geq 0.99$.

Since no a priori estimate of p is available, use $p = 0.5$. The bound on the error of estimation is $E = .05$ with $1-\alpha = .99$. Using the table

$$n = \frac{z_{.005}^2 \cdot pq}{E^2} = \frac{(2.5758)(.5)(.5)}{.05^2} = 663.5 \tag{7.14.1}$$

A sample size of at least 664 should be used. This is a conservative sample size since no a priori estimate of p was available. If it was known that p was less than 0.3, then using this value gives

$$n = \frac{z_{.005}^2 \cdot pq}{E^2} = \frac{(2.5758)(.3)(.7)}{.05^2} = 557.3 \tag{7.14.2}$$

and a sample size of only 558 is needed.

7.14.2 BERNOULLI VARIABLES

For a Bernoulli random variable with probability of success p_B, sometimes it is necessary to know how many trials are required to confirm that $p_B > p_T$ for a given threshold value p_T at a significance level of α.

Define the hypotheses:

- $H_0 : p_B < p_T$ *(note: expect more failures)*
- $H_a : p_B \geq p_T$ *(note: expect fewer failures)*

The probability of N Bernoulli trials with probability of success p_B having

1. exactly F failures is $\text{pdf}(F; N, p_B) = \binom{N}{F} p_B^{N-F}(1-p_B)^F$.
2. F or fewer failures is $\text{cdf}(F; N, p_B) = \sum_{f=0,1,\dots,F} \text{pdf}(f; N, p_B)$.

Recall that $\alpha = \text{Prob}\,(\text{reject } H_0 \mid H_0 \text{ is true})$. The probability of having exactly 0 failures is $\text{cdf}(0; N, p_B) = \text{pdf}(0; N, p_B) = p_B^N$. This will be less than α when, using $p_B < p_T$ when H_0 is true, $N \geq N_0 = \left\lceil \frac{\log \alpha}{\log p_T} \right\rceil$. Hence, if $N \geq N_0$ sample values of the Bernoulli random variable are obtained, and no failures occur, then H_0 is rejected at a significance level of α.

If there is a failure within the N_0 trials, then (if the goal is to reject H_0) more sample values must be obtained. Define N_1 to be the least integer so that $\text{cdf}(1; N_1, p_T) < \alpha$. If $N \geq N_1$ sample values of the Bernoulli random variable are obtained, and no more than 1 failure occurs, then H_0 is rejected at a significance level of α.

Likewise if $N \geq N_2$ (with $\text{cdf}(2; N_2, p_T) < \alpha$) sample values of the Bernoulli random variable are obtained, and no more than 2 failures occur, then H_0 is rejected at a significance level of α. The value N_3 is defined analogously.

	$\alpha = 10\%$				$\alpha = 5\%$				$\alpha = 1\%$			
p	N_0	N_1	N_2	N_3	N_0	N_1	N_2	N_3	N_0	N_1	N_2	N_3
0.90	22	38	52	65	29	46	61	76	44	64	81	97
0.91	25	42	58	73	32	51	68	84	49	71	91	109
0.92	28	48	65	82	36	58	77	95	56	81	102	122
0.93	32	55	75	94	42	66	88	109	64	92	117	140
0.94	38	64	88	110	49	78	103	127	75	108	137	164
0.95	45	77	105	132	59	93	124	153	90	130	165	198
0.96	57	96	132	166	74	117	156	192	113	164	207	248
0.97	76	129	176	221	99	157	208	257	152	219	277	332
0.98	114	194	265	333	149	236	313	386	228	330	418	499
0.99	230	388	531	667	299	473	628	773	459	662	838	1001

7.14.3 OBSERVATIONS FOR VARIANCE COMPARISON

Suppose $x_1, x_2, \ldots, x_{n+1}$ is a random sample from a population with variance σ_1^2. The sample variance, s_1^2, has n degrees of freedom and may be used to test the hypothesis that the population variance is σ_0^2. Let R be the ratio of the variances σ_0^2 and σ_1^2. The table below shows the value of R for which a chi-square test, with significance level α, will not be able to detect the difference in the variances with probability β. Note that when R is far from one, few samples are needed to distinguish σ_0^2 from σ_1^2, while for R near one many samples are needed.

EXAMPLE *Testing for an increase in variance.* Let $\alpha = 0.05$, $\beta = 0.01$, and $R = 4$. Using the table below with these values, the value $R = 4$ occurs between the rows corresponding to $n = 15$ and $n = 20$. Using rough interpolation, this indicates that the variance estimate should be based on 19 degrees of freedom.

EXAMPLE *Testing for a decrease in variance.* Let $\alpha = 0.05$, $\beta = 0.01$, and $R = 0.33$. Using the table below with $\alpha' = \beta = 0.01$, $\beta' = \alpha = 0.05$ and $R' = 1/R = 3$, the value $R' = 3$ occurs between the rows corresponding to $n = 24$ and $n = 30$. Using rough interpolation, this indicates that the variance estimate should be based on 26 degrees of freedom.

	$\alpha = 0.01$				$\alpha = 0.05$			
n	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$	$\beta = 0.5$	$\beta = 0.01$	$\beta = 0.05$	$\beta = 0.1$	$\beta = 0.5$
1	42237	1687.4	420.18	14.584	24454	976.94	243.27	8.444
2	458.21	89.781	43.709	6.644	298.07	58.404	28.433	4.322
3	98.796	32.244	19.414	4.795	68.054	22.211	13.373	3.303
4	44.686	18.681	12.483	3.955	31.933	13.349	8.920	2.827
5	27.217	13.170	9.369	3.467	19.972	9.665	6.875	2.544
6	19.278	10.280	7.627	3.144	14.438	7.699	5.713	2.354
7	14.911	8.524	6.521	2.911	11.353	6.490	4.965	2.217
8	12.202	7.352	5.757	2.736	9.418	5.675	4.444	2.112
9	10.377	6.516	5.198	2.597	8.103	5.088	4.059	2.028
10	9.072	5.890	4.770	2.484	7.156	4.646	3.763	1.960
15	5.847	4.211	3.578	2.133	4.780	3.442	2.925	1.743
20	4.548	3.462	3.019	1.943	3.803	2.895	2.524	1.624
25	3.845	3.033	2.690	1.821	3.267	2.577	2.286	1.547
30	3.403	2.752	2.471	1.735	2.927	2.367	2.125	1.492
40	2.874	2.403	2.192	1.619	2.516	2.103	1.919	1.418
50	2.564	2.191	2.021	1.544	2.272	1.942	1.791	1.368
75	2.150	1.898	1.779	1.431	1.945	1.716	1.609	1.294
100	1.938	1.743	1.649	1.367	1.775	1.596	1.510	1.252
150	1.715	1.575	1.506	1.297	1.594	1.464	1.400	1.206
∞	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

7.15 CONTINGENCY TABLES

The general $I \times J$ contingency table has the form:

	Treatment 1	Treatment 1	...	Treatment J	Totals
Sample 1	n_{11}	n_{12}	$\cdots$	n_{1J}	$n_{1.}$
Sample 2	n_{21}	n_{22}	$\cdots$	n_{2J}	$n_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
Sample I	n_{I1}	n_{I2}	$\cdots$	n_{IJ}	$n_{I.}$
Totals	$n_{.1}$	$n_{.2}$	$\cdots$	$n_{.J}$	n

where $n_{k.} = \sum_{j=1}^{J} n_{kj}$ and $n_{.k} = \sum_{i=1}^{I} n_{ik}$. For complete independence the probability of any specific row and column totals $\{n_{.k}, n_{k.}\}$ is

$$\text{Prob}\,(n_{11}, \ldots, n_{IJ} \mid n_{1.}, \ldots, n_{.J}) = \frac{(\Pi_i^I n_{i.}!)(\Pi_j^J n_{.j}!)}{n!\; \Pi_i^I \Pi_j^J n_{ij}!} \tag{7.15.1}$$

Let $\hat{\epsilon}_{ij}$ be the estimated expected count in the $(i,j)^{\text{th}}$ cell:

$$\hat{\epsilon}_{ij} = \frac{(i^{\text{th}}\text{ row total})(j^{\text{th}}\text{ column total})}{\text{grand total}} = \frac{n_{i.}n_{.j}}{n} \tag{7.15.2}$$

The test statistic is

$$\chi^2 = \sum_{\text{all cells}} \frac{(\text{observed} - \text{estimated expected})^2}{\text{estimated expected}} = \sum_{i=1}^{I}\sum_{j=1}^{J} \frac{(n_{ij} - \hat{\epsilon}_{ij})^2}{\hat{\epsilon}_{ij}} \tag{7.15.3}$$

Under the null hypothesis χ^2 is approximately chi-square distributed with $(I-1)(J-1)$ degrees of freedom. The approximation is satisfactory if $\hat{\epsilon}_{ij} \geq 5$ for all i and j.

7.15.1 SIGNIFICANCE TEST IN 2 × 2 CONTINGENCY TABLES

A 2×2 contingency table is a special case that occurs often. Suppose n elements are simultaneously classified as having either property 1 or 2 and as having property I or II. The 2×2 contingency table may be written as:

	I	II	Totals
1	a	$A-a$	A
2	b	$B-b$	B
Totals	r	$n-r$	n

If the marginal totals (r, A, and B) are fixed, the probability of a given configuration may be written as

$$f(a \mid r, A, B) = \frac{\binom{A}{a}\binom{B}{b}}{\binom{n}{r}} = \frac{A!\,B!\,r!\,(n-r)!}{n!\,a!\,b!\,(A-a)!\,(B-b)!} \tag{7.15.4}$$

The table on page 625 can be used to conduct a hypothesis test concerning the difference between observed and expected frequencies in a 2×2 contingency table. Given values of a, A, and B, the table entries show the largest value of b (in bold type, with $b < a$) for which there is a significant difference between a/A and b/B). Critical values of b (probability levels) are presented for $\alpha = 0.05$, 0.025, 0.01, and 0.005. The tables also satisfy the following conditions:

1. Categories 1 and 2 are defined so that $A \geq B$.
2. $\dfrac{a}{A} \geq \dfrac{b}{B}$ or, $aB \geq bA$.
3. If b is less than or equal to the integer in bold type, then a/A is significantly greater than b/B (for a one-tailed test) at the probability level (α) indicated by the column heading. For a two-tailed test, the significance level is 2α.
4. A dash in the body of the table indicates no 2×2 table may show a significant effect at that probability level and combination of a, A, and B.
5. For a given r, the probability b is less than the integer in bold type is shown in small type following an entry.

Note that if A and B are large, this test may be approximated by a two-sample Z test of proportions.

EXAMPLE In order to compare the probability of a success in two populations, the following 2×2 contingency table was obtained.

	Success	Failure	Totals
Sample from population 1	7	2	9
Sample from population 2	3	3	6
Totals	10	5	15

Is there any evidence to suggest the two population proportions are different? Use $\alpha = .05$.

1. In this 2×2 contingency table, $a = 7$, $A = 9$, and $B = 6$. For $\alpha = 0.05$ the table entry is **1**.034.
2. The critical value for b is 1. If $b \leq 1$ then the null hypothesis H_0: $p_1 = p_2$ is rejected.
3. Conclusion: The value of the test statistic does not lie in the rejection region, $b = 3$. There is no evidence to suggest the population proportions are different.
4. Note there are six 2×2 tables with the same marginal totals as the table in this example (that is, $A = 9$, $B = 6$, and $r = 10$):

$$\begin{array}{cc|} 9 & 0 \\ 1 & 5 \\ \hline \end{array} \quad \begin{array}{cc|} 8 & 1 \\ 2 & 4 \\ \hline \end{array} \quad \begin{array}{cc|} 7 & 2 \\ 3 & 3 \\ \hline \end{array} \quad \begin{array}{cc|} 6 & 3 \\ 4 & 2 \\ \hline \end{array} \quad \begin{array}{cc|} 5 & 4 \\ 5 & 1 \\ \hline \end{array} \quad \begin{array}{cc|} 4 & 5 \\ 6 & 0 \\ \hline \end{array}$$

Assuming independence, the probability of obtaining each of these six tables (using Equation (7.15.4), rounded) is $\{.002, .045, .24, .42, .25, .042\}$. That is, the first configuration is the least likely, and the fourth configuration is the most likely.

Contingency tables: 2×2

	a	Probability			
		0.05	0.025	0.01	0.005
$A = 3$ $B = 3$	3	**0**.050	–	–	–
$A = 4$ $B = 4$	4	**0**.014	**0**.014	–	–
$B = 3$	4	**0**.029	–	–	–
$A = 5$ $B = 5$	5	**1**.024	**1**.024	**0**.004	**0**.004
	4	**0**.024	**0**.024	–	–
$B = 4$	5	**1**.048	**0**.008	**0**.008	–
	4	**0**.040	–	–	–
$B = 3$	5	**0**.018	**0**.018	–	–
$B = 2$	5	**0**.048	–	–	–
$A = 6$ $B = 6$	6	**2**.030	**1**.008	**1**.008	**0**.001
	5	**1**.039	**0**.008	**0**.008	–
	4	**0**.030	–	–	–
$B = 5$	6	**1**.015	**1**.015	**0**.002	**0**.002
	5	**0**.013	**0**.013	–	–
	4	**0**.045	–	–	–
$B = 4$	6	**1**.033	**0**.005	**0**.005	**0**.005
	5	**0**.024	**0**.024	–	–
$B = 3$	6	**0**.012	**0**.012	–	–
	5	**0**.048	–	–	–
$B = 2$	6	**0**.036	–	–	–
$A = 7$ $B = 7$	7	**3**.035	**2**.010	**1**.002	**1**.002
	6	**2**.049	**1**.014	**0**.002	**0**.002
	5	**1**.049	**0**.010	–	–
	4	**0**.035	–	–	–
$B = 6$	7	**2**.021	**2**.021	**1**.005	**1**.005
	6	**1**.024	**1**.024	**0**.004	**0**.004
	5	**0**.016	**0**.016	–	–
	4	**0**.049	–	–	–
$B = 5$	7	**2**.045	**1**.010	**0**.001	**0**.001
	6	**1**.044	**0**.008	**0**.008	–
	5	**0**.027	–	–	–
$B = 4$	7	**1**.024	**1**.024	**0**.003	**0**.003
	6	**0**.015	**0**.015	–	–
	5	**0**.045	–	–	–
$B = 3$	7	**0**.008	**0**.008	**0**.008	–
	6	**0**.033	–	–	–
$B = 2$	7	**0**.028	–	–	–
$A = 8$ $B = 8$	8	**4**.038	**3**.013	**2**.003	**2**.003
	7	**2**.020	**2**.020	**1**.005	**1**.005
	6	**1**.020	**1**.020	**0**.003	**0**.003
	5	**0**.013	**0**.013	–	–
	4	**0**.038	–	–	–
$B = 7$	8	**3**.026	**2**.007	**2**.007	**1**.001
	7	**2**.034	**1**.009	**1**.009	**0**.001
	6	**1**.030	**0**.006	**0**.006	–
	5	**0**.019	**0**.019	–	–
$B = 6$	8	**2**.015	**2**.015	**1**.003	**1**.003
	7	**1**.016	**1**.016	**0**.002	**0**.002
	6	**1**.049	**0**.009	**0**.009	–
	5	**0**.028	–	–	–
$B = 5$	8	**2**.035	**1**.007	**1**.007	**0**.001
	7	**1**.031	**0**.005	**0**.005	**0**.005

	a	Probability			
		0.05	0.025	0.01	0.005
$A = 8$ $B = 5$	6	**0**.016	**0**.016	–	–
	5	**0**.044	–	–	–
$B = 4$	8	**1**.018	**1**.018	**0**.002	**0**.002
	7	**0**.010	**0**.010	–	–
	6	**0**.030	–	–	–
$B = 3$	8	**0**.006	**0**.006	**0**.006	–
	7	**0**.024	**0**.024	–	–
$B = 2$	8	**0**.022	**0**.022	–	–
$A = 9$ $B = 9$	9	**5**.041	**4**.015	**3**.005	**3**.005
	8	**3**.024	**3**.024	**2**.007	**1**.002
	7	**2**.027	**1**.007	**1**.007	**0**.001
	6	**1**.024	**1**.024	**0**.005	**0**.005
	5	**0**.015	**0**.015	–	–
	4	**0**.041	–	–	–
$B = 8$	9	**4**.029	**3**.009	**3**.009	**2**.002
	8	**3**.041	**2**.013	**1**.003	**1**.003
	7	**2**.041	**1**.012	**0**.002	**0**.002
	6	**1**.035	**0**.007	**0**.007	–
	5	**0**.020	**0**.020	–	–
$B = 7$	9	**3**.019	**3**.019	**2**.005	**2**.005
	8	**2**.024	**2**.024	**1**.006	**0**.001
	7	**1**.020	**1**.020	**0**.003	**0**.003
	6	**0**.010	**0**.010	–	–
	5	**0**.029	–	–	–
$B = 6$	9	**3**.044	**2**.011	**1**.002	**1**.002
	8	**2**.045	**1**.011	**0**.001	**0**.001
	7	**1**.034	**0**.006	**0**.006	–
	6	**0**.017	**0**.017	–	–
	5	**0**.042	–	–	–
$B = 5$	9	**2**.027	**1**.005	**1**.005	**1**.005
	8	**1**.022	**1**.022	**0**.003	**0**.003
	7	**0**.010	**0**.010	–	–
	6	**0**.028	–	–	–
$B = 4$	9	**1**.014	**1**.014	**0**.001	**0**.001
	8	**0**.007	**0**.007	**0**.007	–
	7	**0**.021	**0**.021	–	–
	6	**0**.049	–	–	–
$B = 3$	9	**1**.045	**0**.005	**0**.005	**0**.005
	8	**0**.018	**0**.018	–	–
	7	**0**.045	–	–	–
$B = 2$	9	**0**.018	**0**.018	–	–

7.16 ACCEPTANCE SAMPLING

The USA's Department of Defense uses the sampling plans described in MIL-STD-1916 (Military Standard 1916); this has replaced the older MIL-STD-105 D. In this standard there are three kinds of sampling plans, seven verification levels (VL), five code letters (CL), and three types of inspection. First, the following are determined:

1. Determine the appropriate sampling type (attributes, variables, or continuous).
   (a) Attributes sampling has a sample size.
   (b) Variables sampling has a sample size as well as "k" and "F" criteria.
   (c) Continuous sampling has "clearance number" (i) and "frequency" (f) criteria.
2. From the contract determine the specified verification level (VL).
3. From the lot size determine the code letter (CL).
4. From the production history determine the appropriate type of inspection (normal, tightened, or reduced)

Then the sampling process to be used can be obtained from MIL-STD-1916.

1. When the lot size is less than or equal to the sample size, 100% inspection is required.
2. For each verification level (VL) the sampling size for the normal type of inspection is obtained from a column and row in the appropriate table. Usually, the sample size for the tightened/relaxed type (using more/fewer samples) is obtained from the same row, but from the column to the left/right.

MIL-STD-1916 contains a description of when to switch the type of inspection. Approximately, when performing lot sampling:

- of the normal type, if 10 consecutive lots pass then the type becomes reduced.
- of the normal type, if 2 lots fail in the last 5 then the type becomes tightened.
- of the reduced type, if one lot fails then the type becomes normal.
- of the tightened type, if 5 consecutive lots pass then the type becomes normal.

Code letters (CL) used in sampling tables

Lot size	**Verification Levels (VL)**						
	VII	**VI**	**V**	**IV**	**III**	**II**	**I**
2–170	A	A	A	A	A	A	A
171–288	A	A	A	A	A	A	B
289–544	A	A	A	A	A	B	C
545–960	A	A	A	A	B	C	D
961–1,632	A	A	A	B	C	D	E
1,633–3,072	A	A	B	C	D	E	E
3,073–5,440	A	B	C	D	E	E	E
5,441–9,216	B	C	D	E	E	E	E
9,217–17,408	C	D	E	E	E	E	E
17,409–30,720	D	E	E	E	E	E	E
$\geq$30,721	E	E	E	E	E	E	E

- Attributes Sampling

Code letter	**Verification Levels (VL)**								
	T	**VII**	**VI**	**V**	**IV**	**III**	**II**	**I**	**R**
	sample size								
A	3,072	1,280	512	192	80	32	12	5	3
B	4,096	1,536	640	256	96	40	16	6	3
C	5,120	2,048	768	320	128	48	20	8	3
D	6,144	2,560	1,024	384	160	64	24	10	4
E	8,192	3,072	1,280	512	192	80	32	12	5

- Variables sampling

Code letter	**Verification Levels (VL)**								
	T	**VII**	**VI**	**V**	**IV**	**III**	**II**	**I**	**R**
	sample size (n_v)								
A	113	87	64	44	29	18	9	4	2
B	122	92	69	49	32	20	11	5	2
C	129	100	74	54	37	23	13	7	2
D	136	107	81	58	41	26	15	8	3
E	145	113	87	64	44	29	18	9	4
	k values (one- or two-sided)								
A	3.51	3.27	3.00	2.69	2.40	2.05	1.64	1.21	1.20
B	3.58	3.32	3.07	2.79	2.46	2.14	1.77	1.33	1.20
C	3.64	3.40	3.12	2.86	2.56	2.21	1.86	1.45	1.20
D	3.69	3.46	3.21	2.91	2.63	2.32	1.93	1.56	1.20
E	3.76	3.51	3.27	3.00	2.69	2.40	2.05	1.64	1.21
	F values (two-sided)								
A	.136	.145	.157	.174	.193	.222	.271	.370	.707
B	.134	.143	.154	.168	.188	.214	.253	.333	.707
C	.132	.140	.152	.165	.182	.208	.242	.301	.707
D	.130	.138	.148	.162	.177	.199	.233	.283	.435
E	.128	.136	.145	.157	.174	.193	.222	.271	.370

- Continuous sampling

Code letter	**Verification Levels (VL)**								
	T	**VII**	**VI**	**V**	**IV**	**III**	**II**	**I**	**R**
	Screening phase: clearance numbers (i)								
A	3867	2207	1134	527	264	125	55	27	NA
B	7061	3402	1754	842	372	180	83	36	NA
C	11337	5609	2524	1237	572	246	116	53	NA
D	16827	8411	3957	1714	815	368	155	73	NA
E	26912	11868	5709	2605	1101	513	228	96	NA
	Sampling phase: frequencies (f)								
A	1/3	4/17	1/6	2/17	1/12	1/17	1/24	1/34	1/48
B	4/17	1/6	2/17	1/12	1/17	1/24	1/34	1/48	1/68
C	1/6	2/17	1/12	1/17	1/24	1/34	1/48	1/68	1/96
D	2/17	1/12	1/17	1/24	1/34	1/48	1/68	1/96	1/136
E	1/12	1/17	1/24	1/34	1/48	1/68	1/96	1/136	1/192

7.17 PROBABILITY TABLES

7.17.1 CRITICAL VALUES

1. The critical value z_α satisfies $\Phi(z_\alpha) = 1 - \alpha$ (where, as usual, $\Phi(z)$ is the distribution function for the standard normal). See Figure 7.3.

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt = \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{z}{\sqrt{2}}\right)\right) \tag{7.17.1}$$

x	1.282	1.645	1.960	2.326	2.576	3.090
$\Phi(x)$	0.90	0.95	0.975	0.99	0.995	0.999
$2[1-\Phi(x)]$	0.20	0.10	0.05	0.02	0.01	0.002

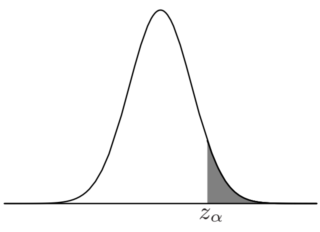
x	3.09	3.72	4.26	4.75	5.20	5.61	6.00	6.36
$1-\Phi(x)$	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}	10^{-10}

For large values of x:

$$\left[\frac{e^{-x^2/2}}{\sqrt{2\pi}}\left(\frac{1}{x} - \frac{1}{x^3}\right)\right] < 1 - \Phi(x) < \left[\frac{e^{-x^2/2}}{\sqrt{2\pi}}\left(\frac{1}{x}\right)\right] \tag{7.17.2}$$

2. The critical value t_α satisfies $F(t_\alpha) = 1 - \alpha$ where $F(\cdot)$ is the distribution function for the t-distribution (for a specified number of degrees of freedom).
3. The critical value χ^2_α satisfies $F(\chi^2_\alpha) = 1 - \alpha$ where $F(\cdot)$ is the distribution function for the χ^2-distribution (for a specified number of degrees of freedom).
4. The critical value F_α satisfies $F(F_\alpha) = 1 - \alpha$ where $F(\cdot)$ is the distribution function for the F-distribution (for a specified number of degrees of freedom).

FIGURE 7.3
The shaded region is defined by $X \geq z_\alpha$ and has area α (here X is $N(0,1)$).

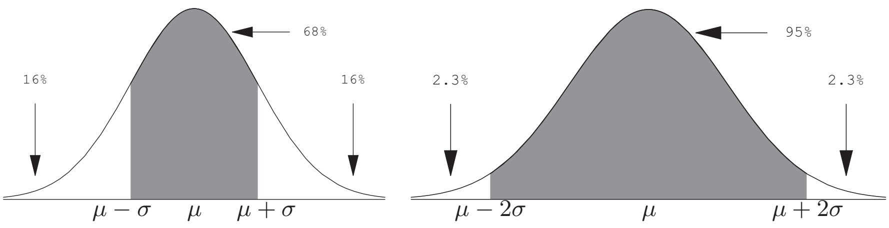


7.17.2 TABLE OF THE NORMAL DISTRIBUTION

For a standard normal random variable (see Figure 7.4):

Limits		Proportion of the total area	Remaining area
$\mu - \lambda\sigma$	$\mu + \lambda\sigma$	(%)	(%)
$\mu - \sigma$	$\mu + \sigma$	68.27	31.73
$\mu - 1.65\sigma$	$\mu + 1.65\sigma$	90	10
$\mu - 1.96\sigma$	$\mu + 1.96\sigma$	95	5
$\mu - 2\sigma$	$\mu + 2\sigma$	95.45	4.55
$\mu - 2.58\sigma$	$\mu + 2.58\sigma$	99.0	0.99
$\mu - 3\sigma$	$\mu + 3\sigma$	99.73	0.27
$\mu - 3.09\sigma$	$\mu + 3.09\sigma$	99.8	0.2
$\mu - 3.29\sigma$	$\mu + 3.29\sigma$	99.9	0.1

FIGURE 7.4
Illustration of σ and 2σ regions of a normal distribution.



x	$F(x)$	$1 - F(x)$	$f(x)$	x	$F(x)$	$1 - F(x)$	$f(x)$
0.01	0.50399	0.49601	0.39892	0.02	0.50798	0.49202	0.39886
0.03	0.51197	0.48803	0.39876	0.04	0.51595	0.48405	0.39862
0.05	0.51994	0.48006	0.39844	0.06	0.52392	0.47608	0.39822
0.07	0.52790	0.47210	0.39797	0.08	0.53188	0.46812	0.39767
0.09	0.53586	0.46414	0.39733	0.10	0.53983	0.46017	0.39695
0.11	0.54380	0.45621	0.39654	0.12	0.54776	0.45224	0.39608
0.13	0.55172	0.44828	0.39559	0.14	0.55567	0.44433	0.39505
0.15	0.55962	0.44038	0.39448	0.16	0.56356	0.43644	0.39387
0.17	0.56749	0.43250	0.39322	0.18	0.57142	0.42858	0.39253
0.19	0.57534	0.42466	0.39181	0.20	0.57926	0.42074	0.39104

x	$F(x)$	$1-F(x)$	$f(x)$	x	$F(x)$	$1-F(x)$	$f(x)$
0.21	0.58317	0.41683	0.39024	0.22	0.58706	0.41294	0.38940
0.23	0.59095	0.40905	0.38853	0.24	0.59484	0.40516	0.38762
0.25	0.59871	0.40129	0.38667	0.26	0.60257	0.39743	0.38568
0.27	0.60642	0.39358	0.38466	0.28	0.61026	0.38974	0.38361
0.29	0.61409	0.38591	0.38251	0.30	0.61791	0.38209	0.38139
0.31	0.62172	0.37828	0.38023	0.32	0.62552	0.37448	0.37903
0.33	0.62930	0.37070	0.37780	0.34	0.63307	0.36693	0.37654
0.35	0.63683	0.36317	0.37524	0.36	0.64058	0.35942	0.37391
0.37	0.64431	0.35569	0.37255	0.38	0.64803	0.35197	0.37115
0.39	0.65173	0.34827	0.36973	0.40	0.65542	0.34458	0.36827
0.41	0.65910	0.34090	0.36678	0.42	0.66276	0.33724	0.36526
0.43	0.66640	0.33360	0.36371	0.44	0.67003	0.32997	0.36213
0.45	0.67365	0.32636	0.36053	0.46	0.67724	0.32276	0.35889
0.47	0.68082	0.31918	0.35723	0.48	0.68439	0.31561	0.35553
0.49	0.68793	0.31207	0.35381	0.50	0.69146	0.30854	0.35207
0.51	0.69497	0.30503	0.35029	0.52	0.69847	0.30153	0.34849
0.53	0.70194	0.29806	0.34667	0.54	0.70540	0.29460	0.34482
0.55	0.70884	0.29116	0.34294	0.56	0.71226	0.28774	0.34105
0.57	0.71566	0.28434	0.33912	0.58	0.71904	0.28096	0.33718
0.59	0.72240	0.27759	0.33521	0.60	0.72575	0.27425	0.33322
0.61	0.72907	0.27093	0.33121	0.62	0.73237	0.26763	0.32918
0.63	0.73565	0.26435	0.32713	0.64	0.73891	0.26109	0.32506
0.65	0.74215	0.25785	0.32297	0.66	0.74537	0.25463	0.32086
0.67	0.74857	0.25143	0.31874	0.68	0.75175	0.24825	0.31659
0.69	0.75490	0.24510	0.31443	0.70	0.75804	0.24196	0.31225
0.71	0.76115	0.23885	0.31006	0.72	0.76424	0.23576	0.30785
0.73	0.76731	0.23270	0.30563	0.74	0.77035	0.22965	0.30339
0.75	0.77337	0.22663	0.30114	0.76	0.77637	0.22363	0.29887
0.77	0.77935	0.22065	0.29659	0.78	0.78231	0.21769	0.29430
0.79	0.78524	0.21476	0.29200	0.80	0.78814	0.21185	0.28969
0.81	0.79103	0.20897	0.28737	0.82	0.79389	0.20611	0.28504
0.83	0.79673	0.20327	0.28269	0.84	0.79955	0.20045	0.28034
0.85	0.80234	0.19766	0.27798	0.86	0.80510	0.19490	0.27562
0.87	0.80785	0.19215	0.27324	0.88	0.81057	0.18943	0.27086
0.89	0.81327	0.18673	0.26848	0.90	0.81594	0.18406	0.26609
0.91	0.81859	0.18141	0.26369	0.92	0.82121	0.17879	0.26129
0.93	0.82381	0.17619	0.25888	0.94	0.82639	0.17361	0.25647
0.95	0.82894	0.17106	0.25406	0.96	0.83147	0.16853	0.25164
0.97	0.83398	0.16602	0.24923	0.98	0.83646	0.16354	0.24681
0.99	0.83891	0.16109	0.24439	1.00	0.84135	0.15865	0.24197

7.17.3 PERCENTAGE POINTS, STUDENT'S t-DISTRIBUTION

For a given value of n and α this table gives the value of $t_{\alpha,n}$ such that

$$F(t_{\alpha,n}) = \int_{-\infty}^{t_{\alpha,n}} \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)} \left(1 + \frac{x^2}{n}\right)^{-(n+1)/2} dx = 1 - \alpha \qquad (7.17.3)$$

The t-distribution is symmetrical, so that $F(-t) = 1 - F(t)$.

EXAMPLE The table gives $t_{\alpha=0.60,n=2} = 0.325$.
Hence, when $n = 2$, $F(0.325) = 0.4$.

	$F(t) =$								
n	0.6000	0.7500	0.9000	0.9500	0.9750	0.9900	0.9950	0.9990	0.9995
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	318.309	636.619
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	22.327	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	10.215	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.552	3.850
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.450	3.725
50	0.255	0.679	1.299	1.676	2.009	2.403	2.678	3.261	3.496
100	0.254	0.677	1.290	1.660	1.984	2.364	2.626	3.174	3.390
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.091	3.291

7.17.4 PERCENTAGE POINTS, CHI-SQUARE DISTRIBUTION

For a given value of n this table gives the value of χ^2 such that

$$F(\chi^2) = \int_0^{\chi^2} \frac{x^{(n-2)/2} e^{-x/2}}{2^{n/2}\Gamma(n/2)}\, dx \tag{7.17.4}$$

is a specified number.

n	$F=$ 0.005	0.010	0.025	0.050	0.100	0.250	0.500	0.750	0.900	0.950	0.975	0.990	0.995
1	0.0000393	0.0001571	0.0009821	0.00393	0.0158	0.102	0.455	1.32	2.71	3.84	5.02	6.63	7.88
2	0.0100	0.0201	0.0506	0.103	0.211	0.575	1.39	2.77	4.61	5.99	7.38	9.21	10.6
3	0.0717	0.115	0.216	0.352	0.584	1.21	2.37	4.11	6.25	7.81	9.35	11.3	12.8
4	0.207	0.297	0.484	0.711	1.06	1.92	3.36	5.39	7.78	9.49	11.1	13.3	14.9
5	0.412	0.554	0.831	1.15	1.61	2.67	4.35	6.63	9.24	11.1	12.8	15.1	16.7
6	0.676	0.872	1.24	1.64	2.20	3.45	5.35	7.84	10.6	12.6	14.4	16.8	18.5
7	0.989	1.24	1.69	2.17	2.83	4.25	6.35	9.04	12.0	14.1	16.0	18.5	20.3
8	1.34	1.65	2.18	2.73	3.49	5.07	7.34	10.2	13.4	15.5	17.5	20.1	22.0
9	1.73	2.09	2.70	3.33	4.17	5.90	8.34	11.4	14.7	16.9	19.0	21.7	23.6
10	2.16	2.56	3.25	3.94	4.87	6.74	9.34	12.5	16.0	18.3	20.5	23.2	25.2
11	2.60	3.05	3.82	4.57	5.58	7.58	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	3.07	3.57	4.40	5.23	6.30	8.44	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	3.57	4.11	5.01	5.89	7.04	9.30	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	4.07	4.66	5.63	6.57	7.79	10.2	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	4.60	5.23	6.26	7.26	8.55	11.0	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	5.14	5.81	6.91	7.96	9.31	11.9	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	5.70	6.41	7.56	8.67	10.1	12.8	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	6.26	7.01	8.23	9.39	10.9	13.7	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	6.84	7.63	8.91	10.1	11.7	14.6	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	7.43	8.26	9.59	10.9	12.4	15.5	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	8.03	8.90	10.3	11.6	13.2	16.3	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	8.64	9.54	11.0	12.3	14.0	17.2	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	9.26	10.2	11.7	13.1	14.8	18.1	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	9.89	10.9	12.4	13.8	15.7	19.0	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	10.5	11.5	13.1	14.6	16.5	19.9	24.3	29.3	34.4	37.7	40.6	44.3	46.9
30	13.8	15.0	16.8	18.5	20.6	24.5	29.3	34.8	40.3	43.8	47.0	50.9	53.7
35	17.2	18.5	20.6	22.5	24.8	29.1	34.3	40.2	46.1	49.8	53.2	57.3	60.3
50	28.0	29.7	32.4	34.8	37.7	42.9	43.9	56.3	63.2	67.5	71.4	76.2	79.5

7.17.5 PERCENTAGE POINTS, F-DISTRIBUTION

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\, dx = \mathbf{0.8}.$$

n	$m=$ 1	2	3	4	5	6	7	8	9	10	50	100	∞
1	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.19	62.69	63.01	63.33
2	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.47	9.48	9.49
3	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.15	5.14	5.13
4	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.80	3.78	3.76
5	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.15	3.13	3.10
6	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.77	2.75	2.72
7	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.52	2.50	2.47
8	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.35	2.32	2.29
9	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.22	2.19	2.16
10	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.12	2.09	2.06
11	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.04	2.01	1.97
12	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	1.97	1.94	1.90
13	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	1.92	1.88	1.85
14	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	1.87	1.83	1.80
15	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	1.83	1.79	1.76
16	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	1.79	1.76	1.72
17	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.76	1.73	1.69
18	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.74	1.70	1.66
19	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.71	1.67	1.63
20	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.69	1.65	1.61
25	2.92	2.53	2.32	2.18	2.09	2.02	1.97	1.93	1.89	1.87	1.61	1.56	1.52
50	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.76	1.73	1.44	1.39	1.34
100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.69	1.66	1.35	1.29	1.20
∞	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.24	1.17	1.00

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\, dx = \mathbf{0.95}.$$

n	$m =$ 1	2	3	4	5	6	7	8	9	10	50	100	∞
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	251.8	253.0	254.3
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.48	19.49	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.58	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.70	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.44	4.41	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.75	3.71	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.32	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.02	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	2.80	2.76	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.64	2.59	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.51	2.46	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.40	2.35	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.31	2.26	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.24	2.19	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.18	2.12	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.12	2.07	2.01
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.08	2.02	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.04	1.98	1.92
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.00	1.94	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	1.97	1.91	1.84
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	1.84	1.78	1.71
50	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.60	1.52	1.45
100	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.48	1.39	1.28
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.35	1.25	1.00

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\, dx = \mathbf{0.975}.$$

n \ $m =$	1	2	3	4	5	6	7	8	9	10	50	100	∞
1	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.7	963.3	968.6	1008	1013	1018
2	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.48	39.49	39.50
3	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.01	13.96	13.90
4	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.38	8.32	8.26
5	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.14	6.08	6.02
6	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	4.98	4.92	4.85
7	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.28	4.21	4.14
8	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	3.81	3.74	3.67
9	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.47	3.40	3.33
10	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.22	3.15	3.08
11	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.03	2.96	2.88
12	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	2.87	2.80	2.72
13	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	2.74	2.67	2.60
14	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	2.64	2.56	2.49
15	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.55	2.47	2.40
16	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.47	2.40	2.32
17	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.41	2.33	2.25
18	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.35	2.27	2.19
19	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.30	2.22	2.13
20	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.25	2.17	2.09
25	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.08	2.00	1.91
50	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	1.75	1.66	1.54
100	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	1.59	1.48	1.37
∞	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.43	1.27	1.00

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\,dx = \mathbf{0.99}.$$

n	$m=$ 1	2	3	4	5	6	7	8	9	10	50	100	∞
1	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6303	6334	6336
2	98.50	99.00	99.17	99.25	99.30	99.33	99.36	99.37	99.39	99.40	99.48	99.49	99.50
3	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.35	27.23	26.35	26.24	26.13
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.55	13.69	13.58	13.46
5	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	10.16	10.05	9.24	9.13	9.02
6	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.09	6.99	6.88
7	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	5.86	5.75	5.65
8	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.07	4.96	4.86
9	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.52	4.41	4.31
10	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.12	4.01	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	3.81	3.71	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	3.57	3.47	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.38	3.27	3.17
14	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.22	3.11	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.08	2.98	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	2.97	2.86	2.75
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	2.87	2.76	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	2.78	2.68	2.57
19	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	2.71	2.60	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	2.64	2.54	2.42
25	7.77	5.57	4.68	4.18	3.85	3.63	3.46	3.32	3.22	3.13	2.40	2.29	2.17
50	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.78	2.70	1.95	1.82	1.70
100	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	1.74	1.60	1.45
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	1.53	1.32	1.00

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\, dx = \mathbf{0.995}.$$

n	$m=$ 1	2	3	4	5	6	7	8	9	10	50	100	∞
1	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	25211	25337	25465
2	198.5	199.0	199.2	199.2	199.3	199.3	199.4	199.4	199.4	199.4	199.5	199.5	199.5
3	55.55	49.80	47.47	46.19	45.39	44.84	44.43	44.13	43.88	43.69	42.21	42.02	41.83
4	31.33	26.28	24.26	23.15	22.46	21.97	21.62	21.35	21.14	20.97	19.67	19.50	19.32
5	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	12.45	12.30	12.14
6	18.63	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	9.17	9.03	8.88
7	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	7.35	7.22	7.08
8	14.69	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	6.22	6.09	5.95
9	13.61	10.11	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	5.45	5.32	5.19
10	12.83	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	4.90	4.77	4.64
11	12.23	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	4.49	4.36	4.23
12	11.75	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.17	4.04	3.90
13	11.37	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	3.91	3.78	3.65
14	11.06	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	3.70	3.57	3.44
15	10.80	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	3.52	3.39	3.26
16	10.58	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	3.37	3.25	3.11
17	10.38	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	3.25	3.12	2.98
18	10.22	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.14	3.01	2.87
19	10.07	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.04	2.91	2.78
20	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	2.96	2.83	2.69
25	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	2.65	2.52	2.38
50	8.63	5.90	4.83	4.23	3.85	3.58	3.38	3.22	3.09	2.99	2.10	1.95	1.81
100	8.24	5.59	4.54	3.96	3.59	3.33	3.13	2.97	2.85	2.74	1.84	1.68	1.51
∞	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	1.60	1.36	1.00

Given n and m this gives the value of f such that

$$F(f) = \int_0^f \frac{\Gamma((n+m)/2)}{\Gamma(m/2)\Gamma(n/2)} m^{m/2} n^{n/2} x^{m/2-1} (n+mx)^{-(m+n)/2}\, dx = \mathbf{0.999}.$$

n	$m=$ 1	2	3	4	5	6	7	8	9	10	50	100	∞
2	998.5	999.0	999.2	999.2	999.3	999.3	999.4	999.4	999.4	999.4	999.5	999.5	999.5
3	167.0	148.5	141.1	137.1	134.6	132.8	131.6	130.6	129.9	129.2	124.7	124.1	123.5
4	74.14	61.25	56.18	53.44	51.71	50.53	49.66	49.00	48.47	48.05	44.88	44.47	44.05
5	47.18	37.12	33.20	31.09	29.75	28.83	28.16	27.65	27.24	26.92	24.44	24.12	23.79
6	35.51	27.00	23.70	21.92	20.80	20.03	19.46	19.03	18.69	18.41	16.31	16.03	15.75
7	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	14.33	14.08	12.20	11.95	11.70
8	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.77	11.54	9.80	9.57	9.33
9	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	10.11	9.89	8.26	8.04	7.81
10	21.04	14.91	12.55	11.28	10.48	9.93	9.52	9.20	8.96	8.75	7.19	6.98	6.76
11	19.69	13.81	11.56	10.35	9.58	9.05	8.66	8.35	8.12	7.92	6.42	6.21	6.00
12	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.48	7.29	5.83	5.63	5.42
13	17.82	12.31	10.21	9.07	8.35	7.86	7.49	7.21	6.98	6.80	5.37	5.17	4.97
14	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.58	6.40	5.00	4.81	4.60
15	16.59	11.34	9.34	8.25	7.57	7.09	6.74	6.47	6.26	6.08	4.70	4.51	4.31
16	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.19	5.98	5.81	4.45	4.26	4.06
17	15.72	10.66	8.73	7.68	7.02	6.56	6.22	5.96	5.75	5.58	4.24	4.05	3.85
18	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.56	5.39	4.06	3.87	3.67
19	15.08	10.16	8.28	7.27	6.62	6.18	5.85	5.59	5.39	5.22	3.90	3.71	3.51
20	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	5.24	5.08	3.77	3.58	3.38
25	13.88	9.22	7.45	6.49	5.89	5.46	5.15	4.91	4.71	4.56	3.28	3.09	2.89
50	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.82	3.67	2.44	2.25	2.06
100	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.44	3.30	2.08	1.87	1.65
∞	10.83	6.91	5.42	4.62	4.10	3.74	3.47	3.27	3.10	2.96	1.75	1.45	1.00

7.17.6 CUMULATIVE TERMS, BINOMIAL DISTRIBUTION

$$B(n, x; p) = \sum_{k=0}^{x} \binom{n}{k} p^k (1-p)^{n-k}.$$

Note that $B(n, x; p) = B(n, n-x; 1-p)$.

If p is the probability of success, then $B(n, x; p)$ is the probability of x or fewer successes in n independent trials. For example, if a biased coin has a probability $p = 0.4$ of being a head, and the coin is independently flipped 5 times, then there is a 68% chance that there will be 2 or fewer heads (since $B(5, 2; 0.4) = 0.6826$).

		p							
n	x	0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
2	0	0.9025	0.8100	0.7225	0.6400	0.5625	0.4900	0.3600	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9375	0.9100	0.8400	0.7500
3	0	0.8574	0.7290	0.6141	0.5120	0.4219	0.3430	0.2160	0.1250
	1	0.9928	0.9720	0.9393	0.8960	0.8438	0.7840	0.6480	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9844	0.9730	0.9360	0.8750
4	0	0.8145	0.6561	0.5220	0.4096	0.3164	0.2401	0.1296	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.7383	0.6517	0.4752	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9492	0.9163	0.8208	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9961	0.9919	0.9744	0.9375
5	0	0.7738	0.5905	0.4437	0.3277	0.2373	0.1681	0.0778	0.0312
	1	0.9774	0.9185	0.8352	0.7373	0.6328	0.5282	0.3370	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8965	0.8369	0.6826	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9844	0.9692	0.9130	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9990	0.9976	0.9898	0.9688
6	0	0.7351	0.5314	0.3771	0.2621	0.1780	0.1177	0.0467	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.5339	0.4202	0.2333	0.1094
	2	0.9978	0.9841	0.9527	0.9011	0.8306	0.7443	0.5443	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9624	0.9295	0.8208	0.6562
	4	1.0000	1.0000	0.9996	0.9984	0.9954	0.9891	0.9590	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9998	0.9993	0.9959	0.9844
7	0	0.6983	0.4783	0.3206	0.2097	0.1335	0.0824	0.0280	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.4450	0.3294	0.1586	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.7564	0.6471	0.4199	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.9294	0.8740	0.7102	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9871	0.9712	0.9037	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9987	0.9962	0.9812	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9999	0.9998	0.9984	0.9922
8	0	0.6634	0.4305	0.2725	0.1678	0.1001	0.0576	0.0168	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.3671	0.2553	0.1064	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.6785	0.5518	0.3154	0.1445
	3	0.9996	0.9950	0.9787	0.9437	0.8862	0.8059	0.5941	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9727	0.9420	0.8263	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9958	0.9887	0.9502	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9996	0.9987	0.9915	0.9648
	7	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961

n	x	p 0.05	0.10	0.15	0.20	0.25	0.30	0.40	0.50
9	0	0.6302	0.3874	0.2316	0.1342	0.0751	0.0403	0.0101	0.0019
	1	0.9288	0.7748	0.5995	0.4362	0.3003	0.1960	0.0705	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.6007	0.4628	0.2318	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.8343	0.7297	0.4826	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9511	0.9012	0.7334	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9900	0.9747	0.9006	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9987	0.9957	0.9750	0.9102
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9996	0.9962	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980
10	0	0.5987	0.3487	0.1969	0.1074	0.0563	0.0283	0.0060	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.2440	0.1493	0.0464	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.5256	0.3828	0.1673	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.7759	0.6496	0.3823	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.9219	0.8497	0.6331	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9803	0.9526	0.8338	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9965	0.9894	0.9452	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9996	0.9984	0.9877	0.9453
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990
11	0	0.5688	0.3138	0.1673	0.0859	0.0422	0.0198	0.0036	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1971	0.1130	0.0302	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.4552	0.3127	0.1189	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.7133	0.5696	0.2963	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.8854	0.7897	0.5328	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9657	0.9218	0.7535	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9924	0.9784	0.9006	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9988	0.9957	0.9707	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9994	0.9941	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
12	0	0.5404	0.2824	0.1422	0.0687	0.0317	0.0138	0.0022	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.1584	0.0850	0.0196	0.0032
	2	0.9804	0.8891	0.7358	0.5584	0.3907	0.2528	0.0834	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.6488	0.4925	0.2253	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.8424	0.7237	0.4382	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.9456	0.8821	0.6652	0.3872
	6	1.0000	1.0000	0.9993	0.9961	0.9858	0.9614	0.8418	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9972	0.9905	0.9427	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9996	0.9983	0.9847	0.9270
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9972	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998

7.17.7 CUMULATIVE TERMS, POISSON DISTRIBUTION

$$F(x;\lambda) = \sum_{k=0}^{x} e^{-\lambda}\frac{\lambda^k}{k!}.$$

If λ is the rate of Poisson arrivals, then $F(x;\lambda)$ is the probability of x or fewer arrivals occurring in a unit of time. For example, if customers arrive at the rate of $\lambda = 0.5$ customers per hour, then the probability of having no customers in any specified hour is 0.61 (the probability of one or fewer customers is 0.91).

	x									
λ	0	1	2	3	4	5	6	7	8	9
0.02	0.980	1.000								
0.04	0.961	0.999	1.000							
0.06	0.942	0.998	1.000							
0.08	0.923	0.997	1.000							
0.10	0.905	0.995	1.000							
0.15	0.861	0.990	1.000	1.000						
0.20	0.819	0.983	0.999	1.000						
0.25	0.779	0.974	0.998	1.000						
0.30	0.741	0.963	0.996	1.000						
0.35	0.705	0.951	0.995	1.000						
0.40	0.670	0.938	0.992	0.999	1.000					
0.45	0.638	0.925	0.989	0.999	1.000					
0.50	0.607	0.910	0.986	0.998	1.000					
0.60	0.549	0.878	0.977	0.997	1.000					
0.70	0.497	0.844	0.966	0.994	0.999	1.000				
0.80	0.449	0.809	0.953	0.991	0.999	1.000				
0.90	0.407	0.772	0.937	0.987	0.998	1.000				
1.0	0.368	0.736	0.920	0.981	0.996	0.999	1.000			
1.1	0.333	0.699	0.900	0.974	0.995	0.999	1.000			
1.2	0.301	0.663	0.879	0.966	0.992	0.999	1.000			
1.3	0.273	0.627	0.857	0.957	0.989	0.998	1.000			
1.4	0.247	0.592	0.834	0.946	0.986	0.997	0.999	1.000		
1.5	0.223	0.558	0.809	0.934	0.981	0.996	0.999	1.000		
1.6	0.202	0.525	0.783	0.921	0.976	0.994	0.999	1.000		
1.7	0.183	0.493	0.757	0.907	0.970	0.992	0.998	1.000		
1.8	0.165	0.463	0.731	0.891	0.964	0.990	0.997	0.999	1.000	
1.9	0.150	0.434	0.704	0.875	0.956	0.987	0.997	0.999	1.000	
2.0	0.135	0.406	0.677	0.857	0.947	0.983	0.996	0.999	1.000	

λ	x: 0	1	2	3	4	5	6	7	8	9
3	0.050	0.199	0.423	0.647	0.815	0.916	0.967	0.988	0.996	0.999
4	0.018	0.092	0.238	0.433	0.629	0.785	0.889	0.949	0.979	0.992
5	0.007	0.040	0.125	0.265	0.441	0.616	0.762	0.867	0.932	0.968
6	0.003	0.017	0.062	0.151	0.285	0.446	0.606	0.744	0.847	0.916
7	0.001	0.007	0.030	0.082	0.173	0.301	0.450	0.599	0.729	0.831
8	0.000	0.003	0.014	0.042	0.100	0.191	0.313	0.453	0.593	0.717
9	0.000	0.001	0.006	0.021	0.055	0.116	0.207	0.324	0.456	0.587
10	0.000	0.001	0.003	0.010	0.029	0.067	0.130	0.220	0.333	0.458
11	0.000	0.000	0.001	0.005	0.015	0.037	0.079	0.143	0.232	0.341
12	0.000	0.000	0.001	0.002	0.008	0.020	0.046	0.089	0.155	0.242
13	0.000	0.000	0.000	0.001	0.004	0.011	0.026	0.054	0.100	0.166
14	0.000	0.000	0.000	0.001	0.002	0.005	0.014	0.032	0.062	0.109
15	0.000	0.000	0.000	0.000	0.001	0.003	0.008	0.018	0.037	0.070

λ	x: 10	11	12	13	14	15	16	17	18	19
3	1.000									
4	0.997	0.999	1.000							
5	0.986	0.995	0.998	0.999	1.000					
6	0.957	0.980	0.991	0.996	0.999	1.000	1.000			
7	0.901	0.947	0.973	0.987	0.994	0.998	0.999	1.000		
8	0.816	0.888	0.936	0.966	0.983	0.992	0.996	0.998	0.999	1.000
9	0.706	0.803	0.876	0.926	0.959	0.978	0.989	0.995	0.998	0.999
10	0.583	0.697	0.792	0.865	0.916	0.951	0.973	0.986	0.993	0.997
11	0.460	0.579	0.689	0.781	0.854	0.907	0.944	0.968	0.982	0.991
12	0.347	0.462	0.576	0.681	0.772	0.844	0.899	0.937	0.963	0.979
13	0.252	0.353	0.463	0.573	0.675	0.764	0.836	0.890	0.930	0.957
14	0.176	0.260	0.358	0.464	0.570	0.669	0.756	0.827	0.883	0.923
15	0.118	0.185	0.268	0.363	0.466	0.568	0.664	0.749	0.820	0.875

λ	x: 20	21	22	23	24	25	26	27	28	29
9	1.000									
10	0.998	0.999	1.000							
11	0.995	0.998	0.999	1.000						
12	0.988	0.994	0.997	0.999	0.999	1.000				
13	0.975	0.986	0.992	0.996	0.998	0.999	1.000			
14	0.952	0.971	0.983	0.991	0.995	0.997	0.999	0.999	1.000	
15	0.917	0.947	0.967	0.981	0.989	0.994	0.997	0.998	0.999	1.000

7.17.8 CRITICAL VALUES, KOLMOGOROV–SMIRNOV TEST

One-sided test Two-sided test	$p = 0.90$ $p = 0.80$	0.95 0.90	0.975 0.95	0.99 0.98	0.995 0.99
$n = 1$	0.900	0.950	0.975	0.990	0.995
2	0.684	0.776	0.842	0.900	0.929
3	0.565	0.636	0.708	0.785	0.829
4	0.493	0.565	0.624	0.689	0.734
5	0.447	0.509	0.563	0.627	0.669
6	0.410	0.468	0.519	0.577	0.617
7	0.381	0.436	0.483	0.538	0.576
8	0.358	0.410	0.454	0.507	0.542
9	0.339	0.387	0.430	0.480	0.513
10	0.323	0.369	0.409	0.457	0.489
11	0.308	0.352	0.391	0.437	0.468
12	0.296	0.338	0.375	0.419	0.449
13	0.285	0.325	0.361	0.404	0.432
14	0.275	0.314	0.349	0.390	0.418
15	0.266	0.304	0.338	0.377	0.404
20	0.232	0.265	0.294	0.329	0.352
25	0.208	0.238	0.264	0.295	0.317
30	0.190	0.218	0.242	0.270	0.290
35	0.177	0.202	0.224	0.251	0.269
40	0.165	0.189	0.210	0.235	0.252
Approximation for $n > 40$:	$\frac{1.07}{\sqrt{n}}$	$\frac{1.22}{\sqrt{n}}$	$\frac{1.36}{\sqrt{n}}$	$\frac{1.52}{\sqrt{n}}$	$\frac{1.63}{\sqrt{n}}$

7.17.9 CRITICAL VALUES, TWO SAMPLE KOLMOGOROV–SMIRNOV TEST

The value of $D = \max |F_{n_1}(x) - F_{n_2}(x)|$ shown below is so large that the hypothesis H_0, the two distributions are the same, is to be rejected at the indicated level of significance. Here, n_1 and n_2 are assumed to be large.

Level of significance	Value of D
$\alpha = 0.10$	$1.22\sqrt{\frac{n_1+n_2}{n_1n_2}}$
$\alpha = 0.05$	$1.36\sqrt{\frac{n_1+n_2}{n_1n_2}}$
$\alpha = 0.025$	$1.48\sqrt{\frac{n_1+n_2}{n_1n_2}}$
$\alpha = 0.01$	$1.63\sqrt{\frac{n_1+n_2}{n_1n_2}}$
$\alpha = 0.005$	$1.73\sqrt{\frac{n_1+n_2}{n_1n_2}}$
$\alpha = 0.001$	$1.95\sqrt{\frac{n_1+n_2}{n_1n_2}}$

7.17.10 CRITICAL VALUES, SPEARMAN'S RANK CORRELATION

Spearman's coefficient of rank correlation, ρ_s, measures the correspondence between two rankings. Let d_i be the difference between the ranks of the i^{th} pair of a set of n pairs of elements. Then Spearman's rho is defined as

$$\rho_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n} = 1 - \frac{6S_r}{n^3 - n}$$

where $S_r = \sum_{i=1}^{n} d_i^2$. The table below gives critical values for S_r when there is complete independence.

n	$p = 0.90$	$p = 0.95$	$p = 0.99$	$p = 0.999$
4	0.8000	0.8000	–	–
5	0.7000	0.8000	0.9000	–
6	0.6000	0.7714	0.8857	–
7	0.5357	0.6786	0.8571	0.9643
8	0.5000	0.6190	0.8095	0.9286
9	0.4667	0.5833	0.7667	0.9000
10	0.4424	0.5515	0.7333	0.8667
11	0.4182	0.5273	0.7000	0.8364
12	0.3986	0.4965	0.6713	0.8182
13	0.3791	0.4780	0.6429	0.7912
14	0.3626	0.4593	0.6220	0.7670
15	0.3500	0.4429	0.6000	0.7464
20	0.2977	0.3789	0.5203	0.6586
25	0.2646	0.3362	0.4654	0.5962
30	0.2400	0.3059	0.4251	0.5479

Chapter 8

Scientific Computing

The text *Numerical Analysis*, Seventh Edition, Brooks/Cole, Pacific Grove, CA, 2001, by R. L. Burden and J. D. Faires, was the primary reference for most of the information presented in this chapter.

8.1 BASIC NUMERICAL ANALYSIS

8.1.1 APPROXIMATIONS AND ERRORS

Numerical methods involve finding approximate solutions to mathematical problems. Errors of approximation can result from two sources: error inherent in the method or formula used and round-off error. *Round-off error* results when a calculator or computer is used to perform real-number calculations with a finite number of significant digits. All but the first specified number of digits are either *chopped* or *rounded* to that number of digits.

If p^* is an approximation to p, the *absolute error* is defined to be $|p - p^*|$ and the *relative error* is $|p - p^*|/|p|$, provided that $p \neq 0$.

Iterative techniques often generate sequences that (ideally) converge to an exact solution. It is sometimes desirable to describe the *rate of convergence*.

Definition Suppose $\lim \beta_n = 0$ and $\lim \alpha_n = \alpha$. If a positive constant K exists with $|\alpha_n - \alpha| < K|\beta_n|$ for large n, then $\{\alpha_n\}$ is said to converge to α with a *rate of convergence* $O(\beta_n)$. This is read "big oh of β_n" and written $\alpha_n = \alpha + O(\beta_n)$.

Definition Suppose $\{p_n\}$ is a sequence that converges to p, with $p_n \neq p$, for all n. If positive constants λ and α exist with $\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^\alpha} = \lambda$, then $\{p_n\}$ *converges to* p *of order* α, with *asymptotic error constant* λ.

In general, a higher order of convergence yields a more rapid rate of convergence. A sequence has *linear convergence* if $\alpha = 1$ and *quadratic convergence* if $\alpha = 2$.

8.1.1.1 Aitken's $\triangle^2$ method

Definition Given $\{p_n\}_{n=0}^{\infty}$, the *forward difference* $\triangle p_n$ is defined by $\triangle p_n = p_{n+1} - p_n$, for $n \geq 0$. Higher powers $\triangle^k p_n$ are defined recursively by $\triangle^k p_n = \triangle(\triangle^{k-1} p_n)$, for $k \geq 2$. In particular, $\triangle^2 p_n = \triangle(p_{n+1} - p_n) = p_{n+2} - 2p_{n+1} + p_n$.

If a sequence $\{p_n\}$ converges linearly to p and $(p_n - p)(p_{n-1} - p) > 0$, for sufficiently large n, then the new sequence $\{\widehat{p}_n\}$ generated by *Aitken's* $\triangle^2$ *method*,

$$\widehat{p}_n = p_n - \frac{(\triangle p_n)^2}{\triangle^2 p_n} \tag{8.1.1}$$

for all $n \geq 0$, satisfies $\lim_{n\to\infty} \frac{\widehat{p}_n - p}{p_n - p} = 0$.

8.1.1.2 Richardson's extrapolation

Improved accuracy can be achieved by combining *extrapolation* with a low-order formula. Suppose the unknown value M is approximated by a formula $N(h)$ for which

$$M = N(h) + K_1 h + K_2 h^2 + K_3 h^3 + \dots \tag{8.1.2}$$

for some unspecified constants $K_1, K_2, K_3, \ldots$. To apply extrapolation, set $N_1(h) = N(h)$, and generate new approximations $N_j(h)$ by

$$N_j(h) = N_{j-1}\left(\frac{h}{2}\right) + \frac{N_{j-1}\left(\frac{h}{2}\right) - N_{j-1}(h)}{2^{j-1} - 1}. \tag{8.1.3}$$

Then $M = N_j(h) + O(h^j)$. A table of the following form is generated, one row at a time:

$$\begin{array}{llll} N_1(h) & & & \\ N_1(^h/_2) & N_2(h) & & \\ N_1(^h/_4) & N_2(^h/_2) & N_3(h) & \\ N_1(^h/_8) & N_2(^h/_4) & N_3(^h/_2) & N_4(h). \end{array}$$

Extrapolation can be applied whenever the truncation error for a formula has the form $\sum_{j=1}^{m-1} K_j h^{\alpha_j} + O(h^{\alpha_m})$ for constants K_j and $\alpha_1 < \alpha_2 < \cdots < \alpha_m$. In particular, if $\alpha_j = 2j$, the following computation can be used:

$$N_j(h) = N_{j-1}\left(\frac{h}{2}\right) + \frac{N_{j-1}\left(\frac{h}{2}\right) - N_{j-1}(h)}{4^{j-1} - 1}, \tag{8.1.4}$$

and the entries in the j^{th} column of the table have order $O(h^{2j})$.

8.1.2 SOLUTION TO ALGEBRAIC EQUATIONS

Iterative methods generate sequences $\{p_n\}$ that converge to a solution p.

Definition A solution p of $f(x) = 0$ is a *zero of multiplicity* m if $f(x)$ can be written as $f(x) = (x - p)^m q(x)$, for $x \neq p$, where $\lim_{x \to p} q(x) \neq 0$. A zero is called *simple* if $m = 1$.

8.1.2.1 Fixed-point iteration

A *fixed point* p for a function g satisfies $g(p) = p$. Given p_0, generate $\{p_n\}$ by

$$p_{n+1} = g(p_n) \quad \text{for } n \geq 0. \tag{8.1.5}$$

If $\{p_n\}$ converges, then it will converge to a fixed point of g and the value p_n can be used as an approximation for p. The following theorem gives conditions that guarantee convergence.

THEOREM 8.1.1 *(Fixed-point theorem)*
Let $g \in C[a,b]$ and suppose that $g(x) \in [a,b]$ for all x in $[a,b]$. Suppose also that g' exists on (a,b) with $|g'(x)| \leq k < 1$, for all $x \in (a,b)$. If p_0 is any number in $[a,b]$, then the sequence defined by Equation (8.1.5) converges to the (unique) fixed point p in $[a,b]$. Both of the error estimates $|p_n - p| \leq \frac{k^n}{1-k}|p_0 - p_1|$ and $|p_n - p| \leq k^n max\{p_0 - a, b - p_0\}$ hold, for all $n \geq 1$.

The iteration sometimes converges even if the conditions are not all satisfied.

THEOREM 8.1.2

Suppose g is a function that satisfies the conditions of Theorem 8.1.1 and g' is also continuous on (a, b). If $g'(p) \neq 0$, then for any number p_0 in $[a, b]$, the sequence generated by Equation (8.1.5) converges only linearly to the unique fixed point p in $[a, b]$.

THEOREM 8.1.3

Let p be a solution of the equation $x = g(x)$. Suppose that $g'(p) = 0$ and g'' are continuous and bounded by a constant on an open interval I containing p. Then there exists a $\delta > 0$ such that, for $p_0 \in [p - \delta, p + \delta]$, the sequence defined by Equation (8.1.5) converges at least quadratically to p.

8.1.2.2 Steffensen's method

For a linearly convergent fixed-point iteration, convergence can be accelerated by applying Aitken's $\triangle^2$ method. This is called *Steffensen's method*. Define $p_0^{(0)} = p_0$, compute $p_1^{(0)} = g(p_0^{(0)})$ and $p_2^{(0)} = g(p_1^{(0)})$. Set $p_0^{(1)} = \widehat{p}_0$ which is computed using Equation (8.1.1) applied to $p_0^{(0)}, p_1^{(0)}$ and $p_2^{(0)}$. Use fixed-point iteration to compute $p_1^{(1)}$ and $p_2^{(1)}$ and then Equation (8.1.1) to find $p_0^{(2)}$. Continuing, generate $\{p_0^{(n)}\}$.

THEOREM 8.1.4

Suppose that $x = g(x)$ has the solution p with $|g'(p)| < 1$ and $g \in C^3$ in a neighborhood of p. Then there exists a $\delta > 0$ such that Steffensen's method gives quadratic convergence for the sequence $\{p_0^{(n)}\}$ for any $p_0 \in [p - \delta, p + \delta]$.

8.1.2.3 Newton–Raphson method (Newton's method)

To solve $f(x) = 0$, given an initial approximation p_0, generate $\{p_n\}$ using

$$p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}, \quad \text{for } n \geq 0. \tag{8.1.6}$$

Figure 8.1 illustrates the method geometrically. Each value p_{n+1} represents the x-intercept of the tangent line to the graph of $f(x)$ at the point $[p_n, f(p_n)]$.

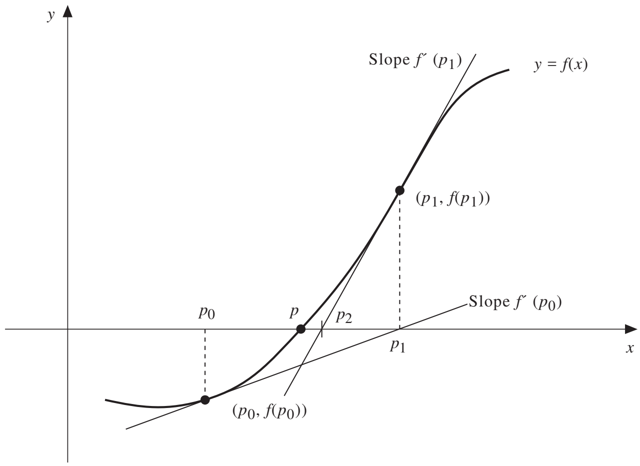
THEOREM 8.1.5

Let $f \in C^2[a, b]$. If $p \in [a, b]$ is such that $f(p) = 0$ and $f'(p) \neq 0$, then there exists a $\delta > 0$ such that Newton's method generates a sequence $\{p_n\}$ converging to p for any initial approximation $p_0 \in [p - \delta, p + \delta]$.

Note:

1. Generally the conditions of the theorem cannot be checked. Therefore one usually generates the sequence $\{p_n\}$ and observes whether or not it converges.
2. An obvious limitation is that the iteration terminates if $f'(p_n) = 0$.
3. For *simple* zeros of f, Theorem 8.1.5 implies that Newton's method converges quadratically. Otherwise, the convergence is much slower.

FIGURE 8.1
Illustration of Newton's method.[1]



8.1.2.4 Modified Newton's method

Newton's method converges only linearly if p has multiplicity larger than one. However, the function $u(x) = \frac{f(x)}{f'(x)}$ has a simple zero at p. Hence, the Newton iteration formula applied to $u(x)$ yields quadratic convergence to a root of $f(x) = 0$. The iteration simplifies to

$$p_{n+1} = p_n - \frac{f(p_n)f'(p_n)}{[f'(p_n)]^2 - f(p_n)f''(p_n)}, \qquad \text{for } n \geq 0. \tag{8.1.7}$$

8.1.2.5 Root-bracketing methods

Suppose $f(x)$ is continuous on $[a, b]$ and $f(a)f(b) < 0$. The Intermediate Value Theorem guarantees a number $p \in (a, b)$ exists with $f(p) = 0$. A *root-bracketing method* constructs a sequence of nested intervals $[a_n, b_n]$, each containing a solution of $f(x) = 0$. At each step, compute $p_n \in [a_n, b_n]$ and proceed as follows:

> If $f(p_n) = 0$, stop the iteration and $p = p_n$.
> Else,
>
> > if $f(a_n)f(p_n) < 0$, then set $a_{n+1} = a_n$, $b_{n+1} = p_n$.
> > Else, set $a_{n+1} = p_n$, $b_{n+1} = b_n$.

[1]From R. L. Burden and J. D. Faires, *Numerical Analysis*, 7th ed., Brooks/Cole, Pacific Grove, CA, 2001. With permission.

8.1.2.6 Secant method

To solve $f(x) = 0$, the *secant method* uses the x-intercept of the secant line passing through $(p_n, f(p_n))$ and $(p_{n-1}, f(p_{n-1}))$. The derivative of f is not needed. Given p_0 and p_1, generate the sequence with

$$p_{n+1} = p_n - f(p_n)\frac{(p_n - p_{n-1})}{f(p_n) - f(p_{n-1})}, \qquad \text{for } n \geq 1. \tag{8.1.8}$$

8.1.2.7 Bisection method

This is a special case of the root-bracketing method. The values p_n are computed by

$$p_n = a_n + \frac{b_n - a_n}{2} = \frac{a_n + b_n}{2}, \qquad \text{for } n \geq 1. \tag{8.1.9}$$

Clearly, $|p_n - p| \leq (b - a)/2^n$ for $n \geq 1$. The rate of convergence is $O(2^{-n})$. Although convergence is slow, the exact number of iterations for a specified accuracy ϵ can be determined. To guarantee that $|p_N - p| < \epsilon$, use

$$N > \log_2\left(\frac{b-a}{\epsilon}\right) = \frac{\ln(b-a) - \ln\epsilon}{\ln 2}. \tag{8.1.10}$$

8.1.2.8 False position (*regula falsi*)

$$p_n = b_n - f(b_n)\frac{b_n - a_n}{f(b_n) - f(a_n)}, \qquad \text{for } n \geq 1. \tag{8.1.11}$$

This root-bracketing method also converges if the initial criteria are satisfied.

8.1.2.9 Horner's method with deflation

If Newton's method is used to solve for roots of the polynomial $P(x) = 0$, then the polynomials P and P' are repeatedly evaluated. Horner's method efficiently evaluates a polynomial of degree n using only n multiplications and n additions.

8.1.2.10 Horner's algorithm

To evaluate $P(x) = a_n x^n + a_{n-1}x^{n-1} + \ldots + a_0$ and its derivative at x_0:

INPUT: degree n, coefficients $\{a_0, a_1, \ldots, a_n\}$; x_0,
OUTPUT: $y = P(x_0)$; $z = P'(x_0)$.
Algorithm:

1. Set $y = a_n$; $z = a_n$.
2. For $j = n-1, n-2, \ldots, 1$,
   set $y = x_0 y + a_j$; $z = x_0 z + y$.
3. Set $y = x_0 y + a_0$.
4. OUTPUT (y, z). STOP.

When satisfied with the approximation $\widehat{x}_1$ for a root x_1 of P, use synthetic division to compute $Q_1(x)$ so that $P(x) \approx (x - \widehat{x}_1)Q_1(x)$. Estimate a root of $Q_1(x)$ and write $P(x) \approx (x - \widehat{x}_1)(x - \widehat{x}_2)Q_2(x)$, and so on. Eventually, $Q_{n-2}(x)$ will be a quadratic, and the quadratic formula can be applied. This procedure, finding one root at a time, is called *deflation*.

Note: Care must be taken since $\widehat{x}_1$ is an *approximation* for x_1. Some inaccuracy occurs when computing the coefficients of $Q_1(x)$, etc. Although the estimate $\widehat{x}_2$ of a root of $Q_1(x)$ can be very accurate, it may not be as accurate when estimating a root of $P(x)$.

8.1.3 INTERPOLATION

Interpolation involves fitting a function to a set of data points (x_0, y_0), (x_1, y_1), $\ldots$, (x_n, y_n). The x_i are unique and the y_i may be regarded as the values of some function $f(x)$, that is, $y_i = f(x_i)$ for $i = 0, 1, \ldots, n$. The following are polynomial interpolation methods.

8.1.3.1 Lagrange interpolation

The *Lagrange interpolating polynomial*, denoted $P_n(x)$, is the unique polynomial of degree at most n for which $P_n(x_k) = f(x_k)$ for $k = 0, 1, \ldots, n$. It is given by

$$P(x) = \sum_{k=0}^{n} f(x_k) L_{n,k}(x) \tag{8.1.12}$$

where $\{x_0, \ldots, x_n\}$ are called *node points*, and

$$\begin{aligned} L_{n,k}(x) &= \frac{(x - x_0)(x - x_1)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_n)}{(x_k - x_0)(x_k - x_1)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)} \\ &= \prod_{i=0, i\neq k}^{n} \frac{(x - x_i)}{(x_k - x_i)}, \qquad \text{for } k = 0, 1, \ldots, n. \end{aligned} \tag{8.1.13}$$

THEOREM 8.1.6 *(Error formula)*

If $x_0, x_1, \ldots, x_n$ are distinct numbers in $[a, b]$ and $f \in C^{n+1}[a, b]$, then, for each x in $[a, b]$, a number $\xi(x)$ in (a, b) exists with

$$f(x) = P(x) + \frac{f^{(n+1)}(\xi(x))}{(n+1)!}(x - x_0)(x - x_1)\cdots(x - x_n), \tag{8.1.14}$$

where P is the interpolating polynomial given in Equation (8.1.12).

Although the Lagrange polynomial is unique, it can be expressed and evaluated in several ways. Equation (8.1.12) is tedious to evaluate, and including more nodes affects the *entire* expression. *Neville's method* evaluates the Lagrange polynomial at a single point *without* explicitly finding the polynomial and the method adapts easily when new nodes are included.

8.1.3.2 Neville's method

Let $P_{m_1,m_2,\ldots,m_k}$ denote the Lagrange polynomial using distinct nodes $\{x_{m_1}, x_{m_2}, \ldots, x_{m_k}\}$. If $P(x)$ denotes the Lagrange polynomial using nodes $\{x_0, x_1, \ldots, x_k\}$ and x_i and x_j are two distinct numbers in this set, then

$$P(x) = \frac{(x-x_j)P_{0,1,\ldots,j-1,j+1,\ldots,k}(x) - (x-x_i)P_{0,1,\ldots,i-1,i+1,\ldots,k}(x)}{(x_i - x_j)}.$$

8.1.3.3 Neville's algorithm

Generate a table of entries $Q_{i,j}$ for $j \geq 0$ and $0 \leq i \leq j$ where the terms are $Q_{i,j} = P_{i-j,i-j+1,\ldots,i-1,i}$. Calculations use Equation (8.1.3.2) for a specific value of x as shown:

$$\begin{array}{lllll} x_0 & Q_{0,0} = P_0 & & & \\ x_1 & Q_{1,0} = P_1 & Q_{1,1} = P_{0,1} & & \\ x_2 & Q_{2,0} = P_2 & Q_{2,1} = P_{1,2} & Q_{2,2} = P_{0,1,2} & \\ x_3 & Q_{3,0} = P_3 & Q_{3,1} = P_{2,3} & Q_{3,2} = P_{1,2,3} & Q_{3,3} = P_{0,1,2,3} \end{array}$$

Note that $P_k = P_k(x) = f(x_k)$ and $\{Q_{i,i}\}$ represents successive estimates of $f(x)$ using Lagrange polynomials. Nodes may be added until $|Q_{i,i} - Q_{i-1,i-1}| < \epsilon$ as desired.

8.1.3.4 Divided differences

Some interpolation formulas involve *divided differences*. Given an ordered sequence of values, $\{x_i\}$. and the corresponding function values $f(x_i)$, the *zero*^*th* *divided difference* is $f[x_i] = f(x_i)$. The *first divided difference* is defined by

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{f[x_{i+1}] - f[x_i]}{x_{i+1} - x_i}. \tag{8.1.15}$$

The k^{th} *divided difference* is defined by

$$f[x_i, x_{i+1}, \ldots, x_{i+k-1}, x_{i+k}] = \frac{f[x_{i+1}, x_{i+2}, \ldots, x_{i+k}] - f[x_i, x_{i+1}, \ldots, x_{i+k-1}]}{x_{i+k} - x_i}. \tag{8.1.16}$$

Divided differences are usually computed by forming a triangular table.

x	$f(x)$	First divided differences	Second divided differences	Third divided differences
x_0	$f[x_0]$			
		$f[x_0, x_1]$		
x_1	$f[x_1]$		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		$f[x_0, x_1, x_2, x_3]$
x_2	$f[x_2]$		$f[x_1, x_2, x_3]$	
		$f[x_2, x_3]$		
x_3	$f[x_3]$			

8.1.3.5 Newton's interpolatory divided-difference formula

$$P_n(x) = f[x_0] + \sum_{k=1}^{n} f[x_0, x_1, \ldots, x_k](x - x_0)\cdots(x - x_{k-1}). \tag{8.1.17}$$

Labeling the nodes as $\{x_n, x_{n-1}, \ldots, x_0\}$, a formula similar to Equation (8.1.17) results in *Newton's backward divided-difference formula*,

$$\begin{aligned} P_n(x) = f[x_n] &+ f[x_n, x_{n-1}](x - x_n) \\ &+ f[x_n, x_{n-1}, x_{n-2}](x - x_n)(x - x_{n-1}) \\ &+ \ldots + f[x_n, x_{n-1}, \ldots, x_0](x - x_n)\cdots(x - x_1). \end{aligned} \tag{8.1.18}$$

If the nodes are equally spaced (that is, $x_i - x_{i-1} = h$), define the parameter s by $x = x_0 + sh$. The following formulas evaluate $P_n(x)$ at a single point:

1. Newton's interpolatory divided-difference formula,

$$P_n(x) = P_n(x_0 + sh) = \sum_{k=0}^{n} \binom{s}{k} k! h^k f[x_0, x_1, \ldots, x_k]. \tag{8.1.19}$$

2. Newton's forward-difference formula (Newton–Gregory),

$$P_n(x) = P_n(x_0 + sh) = \sum_{k=0}^{n} \binom{s}{k} \triangle^k f(x_0). \tag{8.1.20}$$

3. Newton–Gregory backward formula (fits nodes x_{-n} to x_0),

$$P_n(x) = f(x_0) + \binom{s}{1} \triangle f(x_{-1}) + \binom{s+1}{2} \triangle^2 f(x_{-2}) + \ldots + \binom{s+n-1}{n} \triangle^n f(x_{-n}). \tag{8.1.21}$$

4. Newton's backward-difference formula,

$$P_n(x) = \sum_{k=0}^{n} (-1)^k \binom{-s}{k} \nabla^k f(x_n), \tag{8.1.22}$$

   where $\nabla^k f(x_n)$ is the k^{th} *backward difference*, defined for a sequence $\{p_n\}$, by $\nabla p_n = p_n - p_{n-1}$ for $n \geq 1$. Higher powers are defined recursively by $\nabla^k p_n = \nabla(\nabla^{k-1} p_n)$ for $k \geq 2$. For notation, set $\nabla^0 p_n = p_n$.

5. Stirling's formula (for equally spaced nodes $x_{-m}, \ldots, x_{-1}, x_0, x_1, \ldots, x_m$),

$$\begin{aligned} P_n(x) = P_{2m+1}(x) &= f[x_0] + \frac{sh}{2}(f[x_{-1}, x_0] + f[x_0, x_1]) + \\ s^2h^2 f[x_{-1}, x_0, x_1] &+ \frac{s(s^2-1)h^3}{2}(f[x_{-2}, x_{-1}, x_0, x_1] + f[x_{-1}, x_0, x_1, x_2]) \\ &+ \ldots + s^2(s^2-1)(s^2-4)\cdots(s^2-(m-1)^2)h^{2m} f[x_{-m}, \ldots, x_m] \\ &+ \frac{s(s^2-1)\cdots(s^2-m^2)h^{2m+1}}{2}(f[x_{-m-1}, \ldots, x_m] + f[x_{-m}, \ldots, x_{m+1}]). \end{aligned}$$

Use the entire formula if $n = 2m + 1$ is odd, and omit the last term if $n = 2m$ is even. The following table identifies the desired divided differences used in Stirling's formula:

x	$f(x)$	First divided differences	Second divided differences	Third divided differences
x_{-2}	$f[x_{-2}]$			
		$f[x_{-2}, x_{-1}]$		
x_{-1}	$f[x_{-1}]$		$f[x_{-2}, x_{-1}, x_0]$	
		$f[x_{-1}, x_0]$		$f[x_{-2}, x_{-1}, x_0, x_1]$
x_0	$f[x_0]$		$f[x_{-1}, x_0, x_1]$	
		$f[x_0, x_1]$		$f[x_{-1}, x_0, x_1, x_2]$
x_1	$f[x_1]$		$f[x_0, x_1, x_2]$	
		$f[x_1, x_2]$		
x_2	$f[x_2]$			

8.1.3.6 Inverse interpolation

Any method of interpolation which does not require the nodes to be equally spaced may be applied by interchanging the nodes (x values) and the function values (y values).

8.1.3.7 Hermite interpolation

Given distinct numbers $\{x_0, x_1, \ldots, x_n\}$, the *Hermite interpolating polynomial* for a function f is the unique polynomial $H(x)$ of degree at most $2n + 1$ that satisfies $H(x_i) = f(x_i)$ and $H'(x_i) = f'(x_i)$ for each $i = 0, 1, \ldots, n$.

A technique and formula similar to Equation (8.1.17) can be used. For distinct nodes $\{x_0, x_1, \ldots, x_n\}$, define $\{z_0, z_1, \ldots, z_{2n+1}\}$ by $z_{2i} = z_{2i+1} = x_i$ for $i = 0, 1, \ldots, n$. Construct a divided difference table for the ordered pairs $(z_i, f(z_i))$ using $f'(x_i)$ in place of $f[z_{2i}, z_{2i+1}]$, which would be undefined. Denote the Hermite polynomial by $H_{2n+1}(x)$.

8.1.3.8 Hermite interpolating polynomial

$$
\begin{aligned}
H_{2n+1}(x) &= f[z_0] + \sum_{k=1}^{2n+1} f[z_0, z_1, \ldots, z_k](x - z_0) \cdots (x - z_{k-1}) \\
&= f[z_0] + f[z_0, z_1](x - x_0) + f[z_0, z_1, z_2](x - x_0)^2 \\
&\qquad + f[z_0, z_1, z_2, z_3](x - x_0)^2(x - x_1) \\
&\qquad + \ldots + f[z_0, \ldots, z_{2n+1}](x - x_0)^2 \cdots (x - x_{n-1})^2(x - x_n).
\end{aligned}
\tag{8.1.23}
$$

THEOREM 8.1.7 *(Error formula)*
If $f \in C^{2n+2}[a, b]$, then

$$
f(x) = H_{2n+1}(x) + \frac{f^{(2n+2)}(\xi(x))}{(2n+2)!}(x - x_0)^2 \cdots (x - x_n)^2 \tag{8.1.24}
$$

for some $\xi(x) \in (a, b)$ and where $x_i \in [a, b]$ for each $i = 0, 1, \ldots, n$.

8.1.4 DATA FITTING

8.1.4.1 Piecewise polynomial approximation

An interpolating polynomial has large degree and tends to oscillate greatly for large data sets. *Piecewise polynomial approximation* divides the interval into a collection of subintervals and constructs an approximating polynomial on each subinterval. *Piecewise linear interpolation* consists of simply joining the data points with line segments. This collection is continuous but not differentiable at the node points. *Cubic spline interpolation* is popular since no derivative information is needed.

Definition Given a function f defined on $[a, b]$ and a set of numbers $a = x_0 < x_1 < \ldots < x_n = b$, a *cubic spline interpolant*, S, for f is a function that satisfies

1. S is a cubic polynomial, denoted S_j, on $[x_j, x_{j+1}]$ for $j = 0, 1, \ldots, n-1$.
2. For $j = 0, 1, \ldots, n$: $S(x_j) = f(x_j)$
3. For $j = 0, 1, \ldots, n-2$.
   - $S_{j+1}(x_{j+1}) = S_j(x_{j+1})$
   - $S'_{j+1}(x_{j+1}) = S'_j(x_{j+1})$
   - $S''_{j+1}(x_{j+1}) = S''_j(x_{j+1})$
4. One of the following sets of boundary conditions is satisfied:
   - $S''(x_0) = S''(x_n) = 0$ (free or natural boundary),
   - $S'(x_0) = f'(x_0)$ and $S'(x_n) = f'(x_n)$ (clamped boundary).

If a function f is defined at all node points, then f has a unique natural spline interpolant. If, in addition, f is differentiable at a and b, then f has a unique clamped spline interpolant. To construct a cubic spline, set

$$S_j(x) = a_j + b_j(x - x_j) + c_j(x - x_j)^2 + d_j(x - x_j)^3$$

for each $j = 0, 1, \ldots, n-1$. The constants $\{a_j, b_j, c_j, d_j\}$ are found by solving a triadiagonal system of linear equations, which is included in the following algorithms.

8.1.4.2 Algorithm for natural cubic splines

INPUT: n, $\{x_0, x_1, \ldots, x_n\}$, where $a_0 = f(x_0), \ldots, a_n = f(x_n)$.
OUTPUT: $\{a_j, b_j, c_j, d_j\}$ for $j = 0, 1, \ldots, n-1$.
Algorithm:

1. For $i = 0, 1, \ldots, n-1$, set $h_i = x_{i+1} - x_i$.
2. For $i = 1, 2, \ldots, n-1$, set $\alpha_i = \dfrac{3}{h_i}(a_{i+1} - a_i) - \dfrac{3}{h_{i-1}}(a_i - a_{i-1})$.
3. Set $\ell_0 = 1$, $\mu_0 = 0$, $z_0 = 0$.
4. For $i = 1, 2, \ldots, n-1$,
   set $\ell_i = 2(x_{i+1} - x_{i-1}) - h_{i-1}\mu_{i-1}$;
   set $\mu_i = h_i/\ell_i$;
   set $z_i = (\alpha_i - h_{i-1}z_{i-1})/\ell_i$.
5. Set $\ell_n = 1$, $z_n = 0$, $c_n = 0$.

6. For $j = n-1, n-2, \ldots, 0$,
   set $c_j = z_j - \mu_j c_{j+1}$;
   set $b_j = (a_{j+1} - a_j)/h_j - h_j(c_{j+1} + 2c_j)/3$;
   set $d_j = (c_{j+1} - c_j)/(3h_j)$.
7. OUTPUT (a_j, b_j, c_j, d_j for $j = 0, 1, \ldots, n-1$). STOP.

8.1.4.3 Algorithm for clamped cubic splines

INPUT: n, $\{x_0, x_1, \ldots, x_n\}$, $F_0 = f'(x_0)$, $F_n = f'(x_n)$.
$a_0 = f(x_0)$, $a_1 = f(x_1)$, …, $a_n = f(x_n)$,
OUTPUT: $\{a_j, b_j, c_j, d_j\}$ for $j = 0, 1, \ldots, n-1$.
Algorithm

1. For $i = 0, 1, \ldots, n-1$, set $h_i = x_{i+1} - x_i$.
2. Set $\alpha_0 = 3(a_1 - a_0)/h_0 - 3F_0$, $\alpha_n = 3F_n - 3(a_n - a_{n-1})/h_{n-1}$.
3. For $i = 1, 2, \ldots, n-1$, set $\alpha_i = \dfrac{3}{h_i}(a_{i+1} - a_i) - \dfrac{3}{h_{i-1}}(a_i - a_{i-1})$.
4. Set $\ell_0 = 2h_0$, $\mu_0 = 0.5$, $z_0 = \alpha_0/\ell_0$.
5. For $i = 1, 2, \ldots, n-1$,
   set $\ell_i = 2(x_{i+1} - x_{i-1}) - h_{i-1}\mu_{i-1}$;
   set $\mu_i = h_i/\ell_i$;
   set $z_i = (\alpha_i - h_{i-1}z_{i-1})/\ell_i$.
6. Set $\ell_n = h_{n-1}(2 - \mu_{n-1})$, $z_n = (\alpha_n - h_{n-1}z_{n-1})/\ell_n$, $c_n = z_n$.
7. For $j = n-1, n-2, \ldots, 0$,
   set $c_j = z_j - \mu_j c_{j+1}$;
   set $b_j = (a_{j+1} - a_j)/h_j - h_j(c_{j+1} + 2c_j)/3$;
   set $d_j = (c_{j+1} - c_j)/(3h_j)$.
8. OUTPUT (a_j, b_j, c_j, d_j for $j = 0, 1, \ldots, n-1$). STOP.

8.1.4.4 Discrete approximation

Another approach to fit a function to a set of data points $\{(x_i, y_i) \mid i = 1, 2, \ldots, m\}$ is *approximation*. If a polynomial of degree n is used, then $P_n(x) = \sum_{k=0}^{n} a_k x^k$ is found that minimizes the *least squares error* $E = \sum_{i=1}^{m} [y_i - P_n(x_i)]^2$.

- **Normal equations**
  To find $\{a_0, a_1, \ldots, a_n\}$, solve the linear system, called the *normal equations*, created by setting partial derivatives of E taken with respect to each a_k equal to zero. The coefficient of a_0 in the first equation is actually the number of data points, m.

$$
\begin{aligned}
a_0 \sum_{i=1}^{m} x_i^0 + a_1 \sum_{i=1}^{m} x_i^1 + a_2 \sum_{i=1}^{m} x_i^2 + \ldots + a_n \sum_{i=1}^{m} x_i^n &= \sum_{i=1}^{m} y_i x_i^0, \\
a_0 \sum_{i=1}^{m} x_i^1 + a_1 \sum_{i=1}^{m} x_i^2 + a_2 \sum_{i=1}^{m} x_i^3 + \ldots + a_n \sum_{i=1}^{m} x_i^{n+1} &= \sum_{i=1}^{m} y_i x_i^1, \\
&\;\;\vdots \\
a_0 \sum_{i=1}^{m} x_i^n + a_1 \sum_{i=1}^{m} x_i^{n+1} + a_2 \sum_{i=1}^{m} x_i^{n+2} + \ldots + a_n \sum_{i=1}^{m} x_i^{2n} &= \sum_{i=1}^{m} y_i x_i^n.
\end{aligned}
\tag{8.1.25}
$$

Note: $P_n(x)$ can be replaced by a function f of specified form. Unfortunately, to minimize E, the resulting system is generally *not* linear. Although these systems can be solved, one technique is to "linearize" the data. For example, if $y = f(x) = be^{ax}$, then $\ln y = \ln b + ax$. The method applied to the data points $(x_i, \ln y_i)$ produces a linear system. Note that this technique does *not* find the approximation for the original problem but, instead, minimizes the least-squares for the "linearized" data.

- **Best-fit line** Given the points $P_1 = (x_1, y_1)$, $P_2 = (x_2, y_2)$, ..., $P_n = (x_n, y_n)$, the *line of best-fit*, which minimizes E, is given by $y - \overline{y} = m(x - \overline{x})$ where

$$\begin{aligned}
\overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{(x_1 + x_2 + \ldots + x_n)}{n}, \\
\overline{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{(y_1 + y_2 + \ldots + y_n)}{n}, \\
m &= \frac{(x_1y_1 + x_2y_2 + \ldots + x_ny_n) - n\overline{x}\,\overline{y}}{(x_1^2 + x_2^2 + \ldots + x_n^2) - n\overline{x}^2} = \frac{\overline{xy} - \overline{x}\,\overline{y}}{\overline{x^2} - \overline{x}^2}.
\end{aligned} \tag{8.1.26}$$

If **x** and **y** are column vectors containing $\{x_i\}$ and $\{y_i\}$, and **j** is an $n \times 1$ vector of ones, then: $\overline{x} = \mathbf{x}^{\mathrm{T}}\mathbf{j}/n$, $\overline{y} = \mathbf{y}^{\mathrm{T}}\mathbf{j}/n$, $\overline{xy} = \mathbf{x}^{\mathrm{T}}\mathbf{y}/n$, and $\overline{x^2} = \mathbf{x}^{\mathrm{T}}\mathbf{x}/n$.

8.1.5 BÉZIER CURVES

Bézier curves are widely used in computer graphics and image processing. Use of these curves does not depend on understanding the underlying mathematics. This section summarizes Bézier polynomials; the *rational Bézier curves* are not covered.

Bézier curves are defined using parametric equations; the curves are not required to describe a function of one coordinate variable in terms of the other(s). A point P is interpreted as a vector and a Bézier curve has the form $B(t) = [x(t), y(t)]$ (or $B(t) = [x(t), y(t), z(t)]$) where the parameter t satisfies $0 \le t \le 1$.

A Bézier curve $B(t)$ is a weighted average of the $n + 1$ *control points* $\{P_0, P_1, \ldots, P_n\}$, where the weights are the *Bernstein basis polynomials*:

$$B(t) = \sum_{i=0}^{n} \binom{n}{i} (1-t)^{n-i} t^i P_i, \qquad 0 \le t \le 1 \tag{8.1.27}$$

The curve starts at P_0 and ends at P_n. It is said to have *order* $n + 1$ (it has degree n) since it uses $n + 1$ control points. The first three orders are

$$\begin{aligned}
&\text{(Linear)} & B(t) &= (1-t)\ P_0 + tP_1 \\
&\text{(Quadratic)} & B(t) &= (1-t)^2 P_0 + 2(1-t)\ tP_1 + t^2 P_2 \\
&\text{(Cubic)} & B(t) &= (1-t)^3 P_0 + 3(1-t)^2 t P_1 + 3(1-t)t^2 P_2 + t^3 P_3
\end{aligned}$$

Notes:

- Bézier curves of degree n can be viewed as a linear interpolation of two curves of degree $n-1$ since

$$B_{P_0P_1\dots P_n}(t) = (1-t)B_{P_0P_1\dots P_{n-1}}(t) + tB_{P_1P_2\dots P_n}(t)$$

  where $B_{M_1M_2\dots M_k}$ denotes the Bézier curve using nodes $\{M_1,\ M_2,\dots M_k\}$.

- The cubic Bézier curve is essentially the cubic Hermite interpolating polynomial. In general, the curve will not pass through $\{P_i\}$ for $i \neq 0$ and $i \neq n$. These points are sometimes called *guidepoints* since they each provide a direction for the curve. The line P_0P_1 (resp. P_2P_3) is the tangent to the curve at P_0 (resp. P_1).

- Bézier curves are often generated interactively so the user can vary the guidepoints to modify the appearance and shape of the curve.

The method to reproduce a given continuous curve is similar to spline interpolation. The given curve is divided into segments by choosing several points on the curve and a Bézier curve is used to interpolate between each pair of consecutive points. For example, using a cubic curve between each pair of nodes (each requires four nodes), two additional points must be selected to serve as guidepoints. The entire curve is then described using the collection of Bézier curves, creating a *Bézier spline*. Note that any changes made in one of the Bézier curves will only change the nodes of the adjacent Bézier curves.

Bézier curves are infinitely differentiable so any continuity requirement imposed is satisfied, except where two Bézier curves are joined. Suppose two connecting Bézier curves, B_1 and B_2, of equal order are defined by the points $\{P_0, P_1, \dots, P_n\}$ and $\{Q_0, Q_1, \dots Q_n\}$, respectively, where $Q_0 = P_n$. The method of constructing the spline guarantees continuity (i.e., C^0-continuity). C^1-continuity requires $P'(1) = Q'(0)$, so we must have $Q_1 - Q_0 = P_n - P_{n-1}$. Combined with the previous condition this requires $Q_1 = 2P_n - P_{n-1} = P_n + (P_n - P_{n-1})$. Similarly, for C^2-continuity, we would need $P''(1) = Q''(0)$, which requires Q_2 to satisfy $Q_2 = 4P_n - 4P_{n-1} + P_{n-2} = P_{n-2} + 4(P_n - P_{n-1})$. In this way, nodes Q_k can be determined for continuity restrictions of higher order.

The methods of constructing Bézier curves can be extended to define *Bézier surfaces*. A tensor product Bézier surface, called a *Bézier patch* of order $n+1$ is defined by the $(n+1)^2$ control points $\{P_{i,j} \mid 0 \le i, j \le n\}$:

$$P(s,t) = \sum_{i=0}^{n} \binom{n}{i} (1-s)^{n-1} s^i \sum_{j=0}^{n} \binom{n}{j} (1-t)^{n-j} t^j P_{i,j}, \qquad \text{for} \quad \begin{array}{c} 0 \le s \le 1 \\ 0 \le t \le 1 \end{array}$$

Each patch can be viewed as a continuous set of Bézier curves. Adjacent patches are linked in a manner similar to the way Bézier splines are formed from the individual Bézier curves so that desired continuity conditions are satisfied.

8.2 NUMERICAL LINEAR ALGEBRA

8.2.1 SOLVING LINEAR SYSTEMS

The solution of systems of linear equations using *Gaussian elimination* with backward substitution is described in Section 8.2.2. The algorithm is highly sensitive to round-off error. *Pivoting strategies* can reduce round-off error when solving an $n \times n$ system. For a linear system $\mathbf{Ax} = \mathbf{b}$, assume that the equivalent matrix equation $\mathbf{A}^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$ has been constructed. Call the entry, $a_{kk}^{(k)}$, the *pivot element*.

8.2.2 GAUSSIAN ELIMINATION

To solve the system $A\mathbf{x} = \mathbf{b}$, Gaussian elimination creates the *augmented matrix*

$$A' = [A \vdots \mathbf{b}] = \begin{bmatrix} a_{11} & \dots & a_{1n} & b_1 \\ \vdots & & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} & b_n \end{bmatrix}. \tag{8.2.1}$$

This matrix is turned into an upper-triangular matrix by a sequence of (1) row permutations, and (2) subtracting a multiple of one row from another. The result is a matrix of the form (the primes denote that the quantities have been modified)

$$\begin{bmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} & b'_1 \\ 0 & a'_{22} & \dots & a'_{2n} & b'_2 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 0 & a'_{nn} & b'_n \end{bmatrix}. \tag{8.2.2}$$

This matrix represents a linear system that is equivalent to the original system. If the solution exists and is unique, then back substitution can be used to successively determine $\{x_n, x_{n-1}, \dots\}$.

8.2.2.1 Gaussian elimination algorithm

INPUT: number of unknowns and equations n, matrix A, and vector $\mathbf{b}$.
OUTPUT: solution $\mathbf{x} = (x_1, \dots, x_n)^{\mathrm{T}}$ to the linear system $A\mathbf{x} = \mathbf{b}$,
or message that the system does not have a unique solution.
Algorithm:

1. Construct the augmented matrix $A' = [A \vdots \mathbf{b}] = (a'_{ij})$
2. For $i = 1, 2, \dots, n-1$ do (a)–(c): *(Elimination process)*
   (a) Let p be the least integer with $i \leq p \leq n$ and $a'_{pi} \neq 0$
   If no integer can be found, then
   OUTPUT(“no unique solution exists”). STOP.
   (b) If $p \neq i$ interchange rows p and i in A'. The new matrix is A'.
   (c) For $j = i+1, \dots, n$ do i.–ii:

i. Set $m_{ij} = a'_{ji}/a'_{ii}$.
ii. Subtract from row j the quantity (m_{ij} times row i). Replace row j with this result.

3. If $a'_{nn} = 0$ then OUTPUT ("no unique solution exists"). STOP.
4. Set $x_n = a'_{n,n+1}/a'_{nn}$. *(Start backward substitution).*
5. For $i = n-1, \ldots, 2, 1$ set $x_i = \left[a'_{i,n+1} - \sum_{j=i+1}^{n} a'_{ij} x_j\right] / a'_{ii}$.
6. OUTPUT $(x_1, \ldots, x_n)$, *(Procedure completed successfully).* STOP.

8.2.2.2 Pivoting

Maximal column pivoting

Maximal column pivoting (often called *partial pivoting*) finds, at each step, the element in the same column as the pivot element that lies on or below the main diagonal having the largest magnitude and moves it to the pivot position. Determine the least $p \geq k$ such that $\left|a_{pk}^{(k)}\right| = \max_{k \leq i \leq n} \left|a_{ik}^{(k)}\right|$ and interchange the k^{th} equation with the p^{th} equation before performing the elimination step.

Scaled-column pivoting

Scaled-column pivoting sometimes produces better results, especially when the elements of A differ greatly in magnitude. The desired pivot element is chosen to have the largest magnitude *relative* to the other values in its row. For row i define a *scale factor* s_i by $s_i = \max_{1 \leq j \leq n} |a_{ij}|$. The desired pivot element at the k^{th} step is determined by choosing the least integer p with $\left|a_{pk}^{(k)}\right| / s_p = \max_{k \leq j \leq n} \left|a_{jk}^{(k)}\right| / s_j$.

Maximal (or complete) pivoting

The desired pivot element at the k^{th} step is the entry of largest magnitude among $\{a_{ij}\}$ with $i = k, k+1, \ldots, n$ and $j = k, k+1, \ldots, n$. Both row and column interchanges are necessary and additional comparisons are required, resulting in additional execution time.

8.2.3 EIGENVALUE COMPUTATION

8.2.3.1 Power method

Assume that the $n \times n$ matrix A has n eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ with linearly independent eigenvectors $\{\mathbf{v}^{(1)}, \mathbf{v}^{(2)}, \ldots, \mathbf{v}^{(n)}\}$. Assume further that A has a unique dominant eigenvalue λ_1, that is $|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \cdots \geq |\lambda_n|$. Note that for any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \sum_{j=1}^{n} \alpha_j \mathbf{v}^{(j)}$.

The algorithm is called the *power method* because powers of the input matrix are taken: $A^k \mathbf{x} \sim \lambda_1^k \alpha_1 \mathbf{v}^{(1)}$ as $k \to \infty$. However, this sequence converges to zero if $|\lambda_1| < 1$ and diverges if $|\lambda_1| > 1$, provided $\alpha_1 \neq 0$. Appropriate scaling of $A^k \mathbf{x}$ is necessary to obtain a meaningful limit. Begin by choosing a unit vector $\mathbf{x}^{(0)}$ having a component $x_{p_0}^{(0)}$ such that $x_{p_0}^{(0)} = 1 = \left\|\mathbf{x}^{(0)}\right\|_\infty$.

The algorithm inductively constructs sequences of vectors $\{\mathbf{x}^{(m)}\}_{m=0}^{\infty}$ and $\{\mathbf{y}^{(m)}\}_{m=0}^{\infty}$ and a sequence of scalars $\{\mu^{(m)}\}_{m=1}^{\infty}$ by

$$\mathbf{y}^{(m)} = A\mathbf{x}^{(m-1)}, \qquad \mu^{(m)} = y^{(m)}_{p_{m-1}}, \qquad \mathbf{x}^{(m)} = \frac{\mathbf{y}^{(m)}}{y^{(m)}_{p_m}}, \tag{8.2.3}$$

where, at each step, p_m represents the least integer for which $\left|y^{(m)}_{p_m}\right| = \left\|\mathbf{y}^{(m)}\right\|_\infty$.

The sequence of scalars satisfies $\lim_{m\to\infty} \mu^{(m)} = \lambda_1$, provided $\alpha_1 \neq 0$, and the sequence of vectors $\{\mathbf{x}^{(m)}\}_{m=0}^{\infty}$ converges to an eigenvector associated with λ_1 that has l_∞ norm one.

8.2.3.2 Power method algorithm

INPUT: dimension n, matrix A, vector **x**, tolerance TOL, and maximum number of iterations N.

OUTPUT: approximate eigenvalue μ,
approximate eigenvector **x** (with $\|\mathbf{x}\|_\infty = 1$),
or a message that the maximum number of iterations was exceeded.

Algorithm:

1. Set k=1.
2. Find the smallest integer p with $1 \le p \le n$ and $|x_p| = \|\mathbf{x}\|_\infty$.
3. Set $\mathbf{x} = \mathbf{x}/x_p$.
4. While $(k \le N)$ do (a)–(g):
   (a) Set $\mathbf{y} = A\mathbf{x}$.
   (b) Set $\mu = y_p$.
   (c) Find the smallest integer p with $1 \le p \le n$ and $|y_p| = \|\mathbf{y}\|_\infty$.
   (d) If $y_p = 0$ then OUTPUT (“Eigenvector,” **x**, “corresponds to eigenvalue 0. Select a new vector **x** and restart.”); STOP.
   (e) Set ERR $= \|\mathbf{x} - \mathbf{y}/y_p\|_\infty$; $\mathbf{x} = \mathbf{y}/y_p$.
   (f) If ERR $<$ TOL then OUTPUT $(\mu, \mathbf{x})$
   (procedure successful) STOP.
   (g) Set $k = k + 1$.
5. OUTPUT (“Maximum number of iterations exceeded”). STOP.

Notes:

1. The method does not require that λ_1 be unique. If the multiplicity is greater than one, the eigenvector obtained depends on the choice of $\mathbf{x}^{(0)}$.
2. The sequence constructed converges linearly, so that Aitken’s $\triangle^2$ method (Equation (8.1.1)) can be applied to accelerate convergence.

8.2.3.3 Inverse power method

The *inverse power method* modifies the power method to yield faster convergence by finding the eigenvalue of A that is closest to a specified number q. Assume that A satisfies the conditions as before. If $q \neq \lambda_i$, for $i = 1, 2, \dots, n$, the eigenvalues of $(A - qI)^{-1}$ are $\frac{1}{\lambda_1 - q}, \frac{1}{\lambda_2 - q}, \dots, \frac{1}{\lambda_n - q}$, with the same eigenvectors $\mathbf{v}^{(1)}, \dots, \mathbf{v}^{(n)}$. Apply the power method to $(A - qI)^{-1}$. At each step, $\mathbf{y}^{(m)} = (A - qI)^{-1}\mathbf{x}^{(m-1)}$. Generally, $\mathbf{y}^{(m)}$ is found by solving $(A - qI)\mathbf{y}^{(m)} = \mathbf{x}^{(m-1)}$ using Gaussian elimination with pivoting. Choose the value q from an initial approximation to the eigenvector $\mathbf{x}^{(0)}$ by $q = \mathbf{x}^{(0)\mathrm{T}}\mathrm{A}\mathbf{x}^{(0)} / \left(\mathbf{x}^{(0)\mathrm{T}}\mathbf{x}^{(0)}\right)$.

The only changes in the algorithm for the power method are to set an initial value q as described (do this prior to step 1), determine $\mathbf{y}$ in step (4a) by solving the linear system $(A - qI)\mathbf{y} = \mathbf{x}$. (If the system does not have a unique solution, state that q is an eigenvalue and stop.), delete step (4d), and replace step (4f) with

$$\text{if} \quad \text{ERR} < \text{TOL then set } \mu = \frac{1}{\mu} + q; \qquad \text{OUTPUT}(\mu, \mathbf{x}); \qquad \text{STOP.}$$

8.2.3.4 Wielandt deflation

Once the dominant eigenvalue (λ_1) has been found, the remaining eigenvalues can be found by using *deflation techniques*. A new matrix B is formed having the same eigenvalues as A, except that the dominant eigenvalue of A is replaced by 0.

One method is *Wielandt deflation* which defines $\mathbf{x} = \frac{1}{\lambda_1 v_i^{(1)}}[a_{i1}\ a_{i2}\ \dots\ a_{in}]^{\mathrm{T}}$, where $v_i^{(1)}$ is a coordinate of $\mathbf{v}^{(1)}$ that is non-zero, and the values $\{a_{i1}, a_{i2}, \dots, a_{in}\}$ are the entries in the i^{th} row of A. Then the matrix $B = A - \lambda_1 \mathbf{v}^{(1)}\mathbf{x}^{\mathrm{T}}$ has eigenvalues $0, \lambda_2, \lambda_3, \dots, \lambda_n$ with associated eigenvectors $\{\mathbf{v}^{(1)}, \mathbf{w}^{(2)}, \mathbf{w}^{(3)}, \dots, \mathbf{w}^{(n)}\}$, where

$$\mathbf{v}^{(i)} = (\lambda_i - \lambda_1)\mathbf{w}^{(i)} + \lambda_1(\mathbf{x}^{\mathrm{T}}\mathbf{w}^{(i)})\mathbf{v}^{(1)} \tag{8.2.4}$$

for $i = 2, 3, \dots, n$.

The i^{th} row of B consists entirely of zero entries and B may be replaced with an $(n-1) \times (n-1)$ matrix B' obtained by deleting the i^{th} row and i^{th} column of B. The power method can be applied to B' to find *its* dominant eigenvalue and so on.

8.2.4 HOUSEHOLDER'S METHOD

Definition Two $n \times n$ matrices A and B are *similar* if a non-singular matrix S exists with $A = S^{-1}BS$. (If A and B are similar, then they have the same set of eigenvalues).

Householder's method constructs a symmetric tridiagonal matrix B that is similar to a given symmetric matrix A. After applying this method, the *QR algorithm* can be used efficiently to approximate the eigenvalues of the resulting symmetric tridiagonal matrix.

8.2.4.1 Algorithm for Householder's method

To construct a symmetric tridiagonal matrix $A^{(n-1)}$ similar to the symmetric matrix $A = A^{(1)}$, construct matrices $A^{(2)}, A^{(3)}, \dots, A^{(n-1)}$, where $A^{(k)} = (a_{ij}^{(k)})$ for $k = 1, 2, \dots, n-1$.

INPUT: dimension n, matrix A.
OUTPUT: $A^{(n-1)}$. *(At each step, A can be overwritten.)*
Algorithm:

1. For $k = 1, 2, \dots, n-2$ do (a)–(k).
   (a) Set $q = \sum_{j=k+1}^{n} (a_{jk}^{(k)})^2$.
   (b) If $a_{k+1,k}^{(k)} = 0$ then set $\alpha = -q^{\frac{1}{2}}$
   else set $\alpha = -q^{\frac{1}{2}} a_{k+1,k}^{(k)} / \left| a_{k+1,k}^{(k)} \right|$.

(c) Set RSQ $= \alpha^2 - \alpha a_{k+1,k}^{(k)}$.
(d) Set $v_k = 0$. (Note: $v_1 = \cdots = v_{k-1} = 0$, but are not needed.)
 set $v_{k+1} = a_{k+1,k}^{(k)} - \alpha$;
 for $j = k+2, \ldots, n$ set $v_j = a_{jk}^{(k)}$.
(e) For $j = k, k+1, \ldots, n$ set $u_j = \left(\sum_{i=k+1}^{n} a_{ji}^{(k)} v_i\right)/\text{RSQ}$.
(f) Set PROD $= \sum_{i=k+1}^{n} v_i u_i$.
(g) For $j = k, k+1, \ldots, n$ set $z_j = u_j - (\text{PROD}/2\,\text{RSQ})v_j$.
(h) For $\ell = k+1, k+2, \ldots, n-1$ do i.–ii.
 i. For $j = \ell+1, \ldots, n$ set $a_{j\ell}^{(k+1)} = a_{j\ell}^{(k)} - v_\ell z_j - v_j z_\ell$; $a_{\ell j}^{(k+1)} = a_{j\ell}^{(k+1)}$.
 ii. Set $a_{\ell\ell}^{(k+1)} = a_{\ell\ell}^{(k)} - 2v_\ell z_\ell$.
(i) Set $a_{nn}^{(k+1)} = a_{nn}^{(k)} - 2v_n z_n$.
(j) For $j = k+2, \ldots, n$ set $a_{kj}^{(k+1)} = a_{jk}^{(k+1)} = 0$.
(k) Set $a_{k+1,k}^{(k+1)} = a_{k+1,k}^{(k)} - v_{k+1} z_k$; $a_{k,k+1}^{(k+1)} = a_{k+1,k}^{(k+1)}$.
 (Note: The other elements of $A^{(k+1)}$ are the same as $A^{(k)}$.)

2. OUTPUT $(A^{(n-1)})$. STOP.
 ($A^{(n-1)}$ is symmetric, tridiagonal, and similar to A.)

8.2.5 QR ALGORITHM

The *QR algorithm* is generally used (instead of deflation) to determine all of the eigenvalues of a symmetric matrix. The matrix must be symmetric and tridiagonal. If necessary, first apply Householder's method. Suppose the matrix A has the form

$$A = \begin{bmatrix} a_1 & b_2 & 0 & \cdots & 0 & 0 & 0 \\ b_2 & a_2 & b_3 & & 0 & 0 & 0 \\ 0 & b_3 & a_3 & & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & & a_{n-2} & b_{n-1} & 0 \\ 0 & 0 & 0 & & b_{n-1} & a_{n-1} & b_n \\ 0 & 0 & 0 & \cdots & 0 & b_n & a_n \end{bmatrix}. \tag{8.2.5}$$

If $b_2 = 0$ or $b_n = 0$, then A has the eigenvalue a_1 or a_n, respectively. If $b_j = 0$ for some j, $2 < j < n$, the problem is reduced to considering the smaller matrices

$$\begin{bmatrix} a_1 & b_2 & 0 & \cdots & 0 \\ b_2 & a_2 & b_3 & & 0 \\ 0 & b_3 & a_3 & & 0 \\ \vdots & & & & \\ 0 & 0 & & a_{j-2} & b_{j-1} \\ 0 & 0 & & b_{j-1} & a_{j-1} \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} a_j & b_{j+1} & 0 & \cdots & 0 \\ b_{j+1} & a_{j+1} & b_{j+2} & & 0 \\ 0 & b_{j+2} & a_{j+2} & & 0 \\ \vdots & & & & \\ 0 & 0 & & a_{n-1} & b_n \\ 0 & 0 & & b_n & a_n \end{bmatrix}. \tag{8.2.6}$$

If no b_j equals zero, the algorithm constructs $A^{(1)}$, $A^{(2)}$, $A^{(3)}$, $\ldots$ as follows:

1. $A^{(1)} = A$ is factored as $A^{(1)} = Q^{(1)}R^{(1)}$, with $Q^{(1)}$ orthogonal and $R^{(1)}$ upper-triangular.
2. $A^{(2)}$ is defined as $A^{(2)} = R^{(1)}Q^{(1)}$.

In general, $A^{(i+1)} = R^{(i)}Q^{(i)} = ({Q^{(i)}}^{\mathrm{T}}A^{(i)})Q^{(i)} = {Q^{(i)}}^{\mathrm{T}}A^{(i)}Q^{(i)}$. Each $A^{(i+1)}$ is symmetric and tridiagonal with the same eigenvalues as $A^{(i)}$ and, hence, has the same eigenvalues as A.

8.2.5.1 Algorithm for QR

To obtain eigenvalues of the symmetric, tridiagonal $n \times n$ matrix

$$A \equiv A_1 = \begin{bmatrix} a_1^{(1)} & b_2^{(1)} & 0 & \cdots & 0 & 0 & 0 \\ b_2^{(1)} & a_2^{(1)} & b_3^{(1)} & & 0 & 0 & 0 \\ 0 & b_3^{(1)} & a_3^{(1)} & & 0 & 0 & 0 \\ \vdots & & & \ddots & & & \vdots \\ 0 & 0 & 0 & & a_{n-2}^{(1)} & b_{n-1}^{(1)} & 0 \\ 0 & 0 & 0 & & b_{n-1}^{(1)} & a_{n-1}^{(1)} & b_n^{(1)} \\ 0 & 0 & 0 & \cdots & 0 & b_n^{(1)} & a_n^{(1)} \end{bmatrix}. \tag{8.2.7}$$

INPUT: n; $\{a_1^{(1)}, \ldots, a_n^{(1)}, b_2^{(1)}, \ldots, b_n^{(1)}\}$, tolerance TOL, and maximum number of iterations M.

OUTPUT: eigenvalues of A, or recommended splitting of A, or a message that the maximum number of iterations was exceeded.

Algorithm:

1. Set $k = 1$; SHIFT $= 0$. (Accumulated shift)
2. While $k \le M$, do steps 3–12.
3. Test for success:
   (a) If $\left|b_n^{(k)}\right| \le$ TOL, then set $\lambda = a_n^{(k)} +$ SHIFT; OUTPUT (λ);
   set $n = n - 1$.
   (b) If $\left|b_2^{(k)}\right| \le$ TOL then set $\lambda = a_1^{(k)} +$ SHIFT; OUTPUT (λ);
   set $n = n - 1$; $a_1^{(k)} = a_2^{(k)}$;
   for $j = 2, \ldots, n$ set $a_j^{(k)} = a_{j+1}^{(k)}$; $b_j^{(k)} = b_{j+1}^{(k)}$.
   (c) If $n = 0$ then STOP.
   (d) If $n = 1$ then set $\lambda = a_1^{(k)} +$ SHIFT; OUTPUT(λ); STOP.
   (e) For $j = 3, \ldots, n - 1$
   if $\left|b_j^{(k)}\right| \le$ TOL then
   OUTPUT (“split into,” $\{a_1^{(k)}, \ldots, a_{j-1}^{(k)}, b_2^{(k)}, \ldots, b_{j-1}^{(k)}\}$,
   “and” $\{a_j^{(k)}, \ldots, a_n^{(k)}, b_{j+1}^{(k)}, \ldots, b_n^{(k)}\}$, SHIFT); STOP.
4. Compute shift:
   Set $b = -(a_{n-1}^{(k)} + a_n^{(k)})$; $c = a_n^{(k)}a_{n-1}^{(k)} - [b_n^{(k)}]^2$; $d = (b^2 - 4c)^{\frac{1}{2}}$.

5. If $b > 0$, then set $\mu_1 = -2c/(b+d)$; $\mu_2 = -(b+d)/2$;
   else set $\mu_1 = (d-b)/2$; $\mu_2 = 2c/(d-b)$.
6. If $n = 2$, then set $\lambda_1 = \mu_1 + \text{SHIFT}$; $\lambda_2 = \mu_2 + \text{SHIFT}$;
   OUTPUT (λ_1, λ_2); STOP.
7. Choose s so that $\left|s - a_n^{(k)}\right| = \min\left(\left|\mu_1 - a_n^{(k)}\right|, \left|\mu_2 - a_n^{(k)}\right|\right)$.
8. Accumulate shift: Set SHIFT $=$ SHIFT $+ s$.
9. Perform shift: For $j = 1, \ldots, n$ set $d_j = a_j^{(k)} - s$.
10. Compute $R^{(k)}$:
    (a) Set $x_1 = d_1$; $y_1 = b_2$.
    (b) For $j = 2, \ldots, n$
        set $z_{j-1} = (x_{j-1}^2 + [b_j^{(k)}]^2)^{\frac{1}{2}}$; $c_j = x_{j-1}/z_{j-1}$;
        set $s_j = b_j^{(k)}/z_{j-1}$; $q_{j-1} = c_j y_{j-1} + s_j d_j$;
        set $x_j = -s_j y_{j-1} + c_j d_j$.
        If $j \neq n$ then set $r_{j-1} = s_j b_{j+1}^{(k)}$; $y_j = c_j b_{j+1}^{(k)}$.

    (At this point, $A_j^{(k)} = P_j A_{j-1}^{(k)}$ has been computed (P_j is a rotation matrix) and $R^{(k)} = A_n^{(k)}$.)
11. Compute $A^{(k+1)}$.
    (a) Set $z_n = x_n$; $a_1^{(k+1)} = s_2 q_1 + c_2 z_1$; $b_2^{(k+1)} = s_2 z_2$.
    (b) For $j = 2, 3 \ldots, n-1$,
        set $a_j^{(k+1)} = s_{j+1} q_j + c_j c_{j+1} z_j$;
        set $b_{j+1}^{(k+1)} = s_{j+1} z_{j+1}$.
    (c) Set $a_n^{(k+1)} = c_n z_n$.
12. Set $k = k + 1$.
13. OUTPUT (“Maximum number of iterations exceeded”);
    (Procedure unsuccessful.) STOP.

8.2.6 NON-LINEAR SYSTEMS AND NUMERICAL OPTIMIZATION

8.2.6.1 Newton’s method

Many iterative methods exist for solving systems of non-linear equations. *Newton’s method* is a natural extension from solving a single equation in one variable. Convergence is generally quadratic but usually requires an initial approximation that is near the true solution. Assume $\mathbf{F}(\mathbf{x}) = \mathbf{0}$ where $\mathbf{x}$ is an n-dimensional vector, $\mathbf{F} : \mathbb{R}^n \to \mathbb{R}^n$, and $\mathbf{0}$ is the zero vector. That is,

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}(x_1, x_2, \ldots, x_n) = [f_1(x_1, x_2, \ldots, x_n), \ldots, f_n(x_1, x_2, \ldots, x_n)]^{\mathrm{T}}.$$

A fixed-point iteration is performed on $\mathbf{G}(\mathbf{x}) = \mathbf{x} - (J(\mathbf{x}))^{-1}\mathbf{F}(\mathbf{x})$ where $J(\mathbf{x})$ is the *Jacobian matrix*,

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \frac{\partial f_n(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix}. \tag{8.2.8}$$

The iteration is given by

$$\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}) = \mathbf{x}^{(k-1)} - \left[J(\mathbf{x}^{(k-1)})\right]^{-1} \mathbf{F}(\mathbf{x}^{(k-1)}). \tag{8.2.9}$$

The algorithm avoids calculating $(J(\mathbf{x}))^{-1}$ at each step. Instead, it finds a vector $\mathbf{y}$ so that $J(\mathbf{x}^{(k-1)})\mathbf{y} = -\mathbf{F}(\mathbf{x}^{(k-1)})$, and then sets $\mathbf{x}^{(k)} = \mathbf{x}^{(k-1)} + \mathbf{y}$.

For the special case of a two-dimensional system (the equations $f(x, y) = 0$ and $g(x, y) = 0$ are to be satisfied), Newton's iteration becomes:

$$\begin{aligned} x_{n+1} &= x_n - \left.\frac{f g_y - f_y g}{f_x g_y - f_y g_x}\right|_{x=x_n, y=y_n} \\ y_{n+1} &= y_n - \left.\frac{f_x g - f g_x}{f_x g_y - f_y g_x}\right|_{x=x_n, y=y_n} \end{aligned} \tag{8.2.10}$$

8.2.6.2 Method of steepest-descent

The method of *steepest-descent* determines the *local minimum* for a function of the form $g : \mathbb{R}^n \to \mathbb{R}$. It can also be used to solve a system $\{f_i\}$ of non-linear equations. The system has a solution $\mathbf{x} = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$ when the function

$$g(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} [f_i(x_1, x_2, \ldots, x_n)]^2 \tag{8.2.11}$$

has the minimal value zero.

This method converges only linearly to the solution but it usually converges even for poor initial approximations. It can be used to locate initial approximations that are close enough so that Newton's method will converge. Intuitively, a local minimum for a function $g : \mathbb{R}^n \to \mathbb{R}$ can be found as follows:

1. Evaluate g at an initial approximation $\mathbf{x}^{(0)} = (x_1^{(0)}, \ldots, x_n^{(0)})^{\mathrm{T}}$.
2. Determine a direction from $\mathbf{x}^{(0)}$ that results in a decrease in the value of g.
3. Move an appropriate distance in this direction and call the new vector $\mathbf{x}^{(1)}$.
4. Repeat steps 1 through 3 with $\mathbf{x}^{(0)}$ replaced by $\mathbf{x}^{(1)}$.

The direction of greatest decrease in the value of g at $\mathbf{x}$ is the direction given by $-\nabla g(\mathbf{x})$ where $\nabla g(\mathbf{x})$ is the *gradient* of g.

Definition If $g : \mathbb{R}^n \to \mathbb{R}$, the *gradient* of g at $\mathbf{x} = (x_1, \ldots, x_n)^{\mathrm{T}}$, denoted $\nabla g(\mathbf{x})$, is

$$\nabla g(\mathbf{x}) = \left(\frac{\partial g}{\partial x_1}(\mathbf{x}), \frac{\partial g}{\partial x_2}(\mathbf{x}), \ldots, \frac{\partial g}{\partial x_n}(\mathbf{x})\right)^{\mathrm{T}}. \tag{8.2.12}$$

Thus, set $\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - \alpha\nabla g(\mathbf{x}^{(0)})$ for some constant $\alpha > 0$. Ideally the value of α minimizes the function $h(\alpha) = g\big(\mathbf{x}^{(0)} - \alpha\nabla g(\mathbf{x}^{(0)})\big)$. Instead of tedious direct calculation, the method interpolates h with a quadratic polynomial using nodes α_1, α_2, and α_3 that are hopefully close to the minimum value of h.

8.2.6.3 Algorithm for steepest-descent

To approximate a solution to the minimization problem $\min_{\mathbf{x}\in\mathbb{R}^n} g(\mathbf{x})$, given an initial approximation $\mathbf{x}$.

INPUT: number n of variables, initial approximation $\mathbf{x} = (x_1, \dots, x_n)^{\mathrm{T}}$, tolerance TOL, and maximum number of iterations N.
OUTPUT: approximate solution $\mathbf{x} = (x_1, x_2, \dots, x_n)^{\mathrm{T}}$ or a failure message.

Algorithm:

1. Set $k = 1$.
2. While $(k \le N)$, do steps (a)–(k).
   (a) Set: $g_1 = g(x_1, \dots, x_n)$; (Note: $g_1 = g(\mathbf{x}^{(k)})$.)
   $\mathbf{z} = \nabla g(x_1, \dots, x_n)$; (Note: $\mathbf{z} = \nabla g(\mathbf{x}^{(k)})$.)
   $z_0 = \|\mathbf{z}\|_2$.
   (b) If $z_0 = 0$ then OUTPUT ("Zero gradient");
   (Procedure completed, may have a minimum.)
   OUTPUT $(x_1, \dots, x_n, g_1)$; STOP.
   (c) Set $\mathbf{z} = \mathbf{z}/z_0$. (Make $\mathbf{z}$ a unit vector.)
   Set $\alpha_1 = 0$; $\alpha_3 = 1$; $g_3 = g(\mathbf{x} - \alpha_3\mathbf{z})$.
   (d) While $(g_3 \ge g_1)$, do steps i.–ii.
      i. Set $\alpha_3 = \alpha_3/2$; $g_3 = g(\mathbf{x} - \alpha_3\mathbf{z})$.
      ii. If $\alpha_3 < \mathrm{TOL}/2$, then
      (Procedure completed, may have a minimum.)
      OUTPUT ("No likely improvement");
      OUTPUT $(x_1, \dots, x_n, g_1)$; STOP.
   (e) Set $\alpha_2 = \alpha_3/2$; $g_2 = g(\mathbf{x} - \alpha_2\mathbf{z})$.
   (f) Set: $h_1 = (g_2 - g_1)/\alpha_2$; $h_2 = (g_3 - g_2)/(\alpha_3 - \alpha_2)$;
   $h_3 = (h_2 - h_1)/\alpha_3$.
   (g) Set: $\alpha_0 = (\alpha_2 - h_1/h_3)/2$; (Critical point occurs at α_0.)
   $g_0 = g(\mathbf{x} - \alpha_0\mathbf{z})$.
   (h) Find α from $\{\alpha_0, \alpha_3\}$ so that $g = g(\mathbf{x} - \alpha\mathbf{z}) = \min\{g_0, g_3\}$.
   (i) Set $\mathbf{x} = \mathbf{x} - \alpha\mathbf{z}$.
   (j) If $|g - g_1| < \mathrm{TOL}$ then OUTPUT $(x_1, \dots, x_n, g)$;
   (Procedure completed successfully.) STOP.
   (k) Set $k = k + 1$.
3. OUTPUT ("Maximum iterations exceeded");
   (Procedure unsuccessful.) STOP.

8.3 NUMERICAL INTEGRATION AND DIFFERENTIATION

8.3.1 NUMERICAL INTEGRATION

Numerical quadrature involves estimating $\int_a^b f(x)\,dx$ using a formula of the form

$$\int_a^b f(x)\,dx \approx \sum_{i=0}^{n} c_i f(x_i). \tag{8.3.1}$$

8.3.1.1 Newton–Cotes formulas

A *closed Newton–Cotes formula* uses nodes $x_i = x_0 + ih$ for $i = 0, 1, \ldots, n$, where $h = (b-a)/n$. Note that $x_0 = a$ and $x_n = b$.

An *open Newton–Cotes formula* uses nodes $x_i = x_0 + ih$ for $i = 0, 1, \ldots, n$, where $h = (b-a)/(n+2)$. Here $x_0 = a + h$ and $x_n = b - h$. Set $x_{-1} = a$ and $x_{n+1} = b$. The nodes actually used lie in the *open interval* (a, b).

In all formulas, ξ is a number for which $a < \xi < b$ and f_i denotes $f(x_i)$.

8.3.1.2 Closed Newton–Cotes formulas

1. $(n = 1)$ Trapezoidal rule
$$\int_a^b f(x)\,dx = \frac{h}{2}[f(x_0) + f(x_1)] - \frac{h^3}{12} f''(\xi).$$
2. $(n = 2)$ Simpson's rule
$$\int_a^b f(x)\,dx = \frac{h}{3}[f(x_0) + 4f(x_1) + f(x_2)] - \frac{h^5}{90} f^{(4)}(\xi).$$
3. $(n = 3)$ Simpson's three-eighths rule
$$\int_a^b f(x)\,dx = \frac{3h}{8}[f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)] - \frac{3h^5}{80} f^{(4)}(\xi).$$
4. $(n = 4)$ Milne's rule (also called Boole's rule)
$$\int_a^b f(x)\,dx = \frac{2h}{45}[7f_0 + 32f_1 + 12f_2 + 32f_3 + 7f_4] - \frac{8h^7}{945} f^{(6)}(\xi).$$
5. $(n = 5)$
$$\int_a^b f(x)\,dx = \frac{5h}{288}[19f_0 + 75f_1 + 50f_2 + 50f_3 + 75f_4 + 19f_5] - \frac{275h^7}{12096} f^{(6)}(\xi).$$
6. $(n = 6)$ Weddle's rule
$$\int_a^b f(x)\,dx = \frac{h}{140}[41f_0 + 216f_1 + 27f_2 + 272f_3 + 27f_4 + 216f_5 + 41f_6] - \frac{9h^9}{1400} f^{(8)}(\xi).$$

8.3.1.3 Open Newton–Cotes formulas

1. ($n = 0$) Midpoint rule
$$\int_a^b f(x)\,dx = 2hf(x_0) + \frac{h^3}{3} f''(\xi).$$

2. ($n = 1$)
$$\int_a^b f(x)\,dx = \frac{3h}{2}[f(x_0) + f(x_1)] + \frac{3h^3}{4} f''(\xi).$$

3. ($n = 2$)
$$\int_a^b f(x)\,dx = \frac{4h}{3}[2f(x_0) - f(x_1) + 2f(x_2)] + \frac{14h^5}{45} f^{(4)}(\xi).$$

4. ($n = 3$)
$$\int_a^b f(x)\,dx = \frac{5h}{24}[11f(x_0) + f(x_1) + f(x_2) + 11f(x_3)] + \frac{95h^5}{144} f^{(4)}(\xi).$$

5. ($n = 4$)
$$\int_a^b f(x)\,dx = \frac{3h}{10}[11f_0 - 14f_1 + 26f_2 - 14f_3 + 11f_4] + \frac{41h^7}{140} f^{(6)}(\xi).$$

8.3.1.4 Composite rules

Some Newton–Cotes formulas extend to *composite formulas*. This consists of dividing the interval into sub-intervals and using a Newton–Cotes formulas on each sub-interval, In the following, note that $a < \mu < b$.

1. Composite trapezoidal rule for n subintervals: If $f \in C^2[a, b]$, $h = (b - a)/n$, and $x_j = a + jh$, for $j = 0, 1, \ldots, n$, then
$$\int_a^b f(x)\,dx = \frac{h}{2}\left[f(a) + 2\sum_{j=1}^{n-1} f(x_j) + f(b)\right] - \frac{b-a}{12} h^2 f''(\mu).$$

2. Composite Simpson's rule for n subintervals: If $f \in C^4[a, b]$, n is even, $h = (b - a)/n$, and $x_j = a + jh$, for $j = 0, 1, \ldots, n$, then
$$\int_a^b f(x)\,dx = \frac{h}{3}\left[f(a) + 2\sum_{j=1}^{(n/2)-1} f(x_{2j}) + 4\sum_{j=1}^{n/2} f(x_{2j-1}) + f(b)\right] - \frac{b-a}{180} h^4 f^{(4)}(\mu).$$

3. Composite midpoint rule for $n + 2$ subintervals: If $f \in C^2[a, b]$, n is even, $h = (b - a)/(n + 2)$, and $x_j = a + (j + 1)h$, for $j = -1, 0, 1, \ldots, n + 1$, then
$$\int_a^b f(x)\,dx = 2h\sum_{j=0}^{n/2} f(x_{2j}) + \frac{b-a}{6} h^2 f''(\mu).$$

8.3.1.5 Romberg integration

Romberg integration uses the composite trapezoidal rule beginning with $h_1 = b - a$ and $h_k = (b-a)/2^{k-1}$, for $k = 1, 2, \ldots$, to give preliminary estimates for $\int_a^b f(x)\,dx$ and improves these estimates using Richardson's extrapolation. Since many function evaluations would be repeated, the first column of the extrapolation table (with entries denoted $R_{i,j}$) can be more efficiently determined by the following recursion formula:

$$\begin{aligned} R_{1,1} &= \frac{h_1}{2}[f(a)+f(b)] = \frac{b-a}{2}[f(a)+f(b)], \\ R_{k,1} &= \frac{1}{2}\left[R_{k-1,1} + h_{k-1}\sum_{i=1}^{2^{k-2}} f(a+(2i-1)h_k)\right], \end{aligned} \tag{8.3.2}$$

for $k = 2, 3, \ldots$. Now apply Equation (8.1.4) to complete the extrapolation table.

8.3.1.6 Gregory's formula

Using f_j to represent $f(x_0 + jh)$,

$$\begin{aligned} \int_{x_0}^{x_0+nh} f(y)\,dy &= h(\frac{1}{2}f_0 + f_1 + \ldots + f_{n-1} + \frac{1}{2}f_n) \\ &\quad + \frac{h}{12}(\triangle f_0 - \triangle f_{n-1}) - \frac{h}{24}(\triangle^2 f_0 + \triangle^2 f_{n-2}) \\ &\quad + \frac{19h}{720}(\triangle^3 f_0 - \triangle^3 f_{n-3}) - \frac{3h}{160}(\triangle^4 f_0 + \triangle^4 f_{n-4}) + \ldots. \end{aligned} \tag{8.3.3}$$

where $\triangle$'s represent forward differences. The first expression on the right is the composite trapezoidal rule, and additional terms provide improved approximations. Do not carry this process too far; Gregory's formula is only asymptotically convergent and round-off error can be significant when computing higher differences.

8.3.1.7 Gaussian quadrature

A quadrature formula, whose nodes (abscissae) x_i and coefficients w_i are chosen to achieve a maximum order of accuracy, is called a *Gaussian quadrature formula*. The integrand usually involves a *weight function* w. An integral in t on an interval (a, b) must be converted into an integral in x over the interval (α, β) specified for the weight function involved. This can be accomplished by the transformation $x = \frac{(b\alpha - a\beta)}{(b-a)} + \frac{(\beta-\alpha)t}{(b-a)}$. Gaussian quadrature formulas generally take the form

$$\int_\alpha^\beta w(x)f(x)\,dx = \sum_i w_i f(x_i) + E_n \tag{8.3.4}$$

where the error is $E_n = K_n f^{(2n)}(\xi)$ for some $\alpha < \xi < \beta$ and K_n is a constant. The next table contains many weight functions and their associated intervals.

The subsequent tables give abscissae and weights for selected formulas. If some x_i are specified (such as one or both end points), then use the Radau or Lobatto methods.

FIGURE 8.2
Formulae for integration rules with various weight functions

Weight $w(x)$	Interval (α, β)	Abcissas are zeros of	x_i	w_i	K_n
1	$(-1, 1)$	$P_n(x)$	See table on page 672	$\dfrac{-2}{(n+1)P_{n+1}(x_i)P_n'(x_i)}$	$\dfrac{2^{2n+1}(n!)^4}{(2n+1)[(2n)!]^3}$
e^{-x}	$(0, \infty)$	$L_n(x)$	See table on page 672	$\dfrac{(n!)^2 x_i}{(n+1)^2 L_{n+1}^2(x_i)}$	$\dfrac{(n!)^2}{(2n)!}$
e^{-x^2}	$(-\infty, \infty)$	$H_n(x)$	See table on page 672	$\dfrac{2^{n-1} n! \sqrt{\pi}}{n^2 H_{n-1}^2(x_i)}$	$\dfrac{n! \sqrt{\pi}}{2^n (2n)!}$
$\dfrac{1}{\sqrt{1-x^2}}$	$(-1, 1)$	$T_n(x)$	$\cos \dfrac{(2i-1)\pi}{2n}$	$\dfrac{\pi}{n}$	$\dfrac{2\pi}{2^{2n}(2n)!}$
$\sqrt{1-x^2}$	$(-1, 1)$	$U_n(x)$	$\cos\left(\dfrac{i\pi}{n+1}\right)$	$\dfrac{\pi}{n+1} \sin^2\left(\dfrac{i\pi}{n+1}\right)$	$\dfrac{\pi}{2^{2n+1}(2n)!}$
$\sqrt{\dfrac{x}{1-x}}$	$(0, 1)$	$\dfrac{T_{2n+1}(\sqrt{x})}{\sqrt{x}}$	$\cos^2\left(\dfrac{(2i-1)\pi}{4n+2}\right)$	$\dfrac{2\pi}{2n+1} \cos^2\left(\dfrac{(2i-1)\pi}{4n+2}\right)$	$\dfrac{\pi}{2^{4n+1}(2n)!}$
$\sqrt{\dfrac{1-x}{1+x}}$	$(0, 1)$	$J_n\left(x, \dfrac{1}{2}, -\dfrac{1}{2}\right)$	$\cos\left(\dfrac{2i\pi}{2n+1}\right)$	$\dfrac{4\pi}{2n+1} \sin^2\left(\dfrac{i\pi}{2n+1}\right)$	$\dfrac{\pi}{2^{2n}(2n)!}$
$\dfrac{1}{\sqrt{x}}$	$(0, 1)$	$P_{2n}(\sqrt{x})$	$\left(x_i^+\right)^2$	$2h_i$	$\dfrac{2^{4n+1}[(2n)!]^3}{(4n+1)[(4n)!]^2}$
$\sqrt{x}$	$(0, 1)$	$\dfrac{P_{2n+1}(\sqrt{x})}{\sqrt{x}}$	$\left(x_i^+\right)^2$	$2h_i \left(x_i^+\right)^2$	$\dfrac{2^{4n+3}[(2n+1)!]^4}{(4n+3)[(4n+2)!]^2(2n)!}$

In this table, P_n, L_n, H_n, T_n, U_n, and J_n denote the n^{th} Legendre, Laguerre, Hermite, Chebyshev (first kind T_n, second kind U_n), and Jacobi polynomials, respectively. Also, x_i^+ denotes the i^{th} positive root of $P_{2n}(x)$ or $P_{2n+1}(x)$ of the previous column, and h_i denotes the corresponding weight for x_i^+ in the Gauss–Legendre formula $(w(x) = 1)$.

8.3.1.8 Gauss–Legendre quadrature

Weight function is $w(x) = 1$.

$$\int_{-1}^{1} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
2	0.5773502692	1
3	0	0.8888888889
	0.7745966692	0.5555555556
4	0.3399810436	0.6521451549
	0.8611363116	0.3478548451

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
5	0	0.5688888889
	0.5384693101	0.4786286705
	0.9061798459	0.2369268851
6	0.2386191861	0.4679139346
	0.6612093865	0.3607615730
	0.9324695142	0.1713244924

8.3.1.9 Gauss–Laguerre quadrature

Weight function is $w(x) = e^{-x}$.

$$\int_{0}^{\infty} e^{-x} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

n	Nodes $\{x_i\}$	Weights $\{w_i\}$
2	0.5857864376	0.8535533905
	3.4142135623	0.1464466094
3	0.4157745567	0.7110930099
	2.2942803602	0.2785177335
	6.2899450829	0.0103892565

n	Nodes $\{x_i\}$	Weights $\{w_i\}$
4	0.3225476896	0.6031541043
	1.7457611011	0.3574186924
	4.5366202969	0.0388879085
	9.3950709123	0.0005392947
5	0.2635603197	0.5217556105
	1.4134030591	0.3986668110
	3.5964257710	0.0759424496
	7.0858100058	0.0036117586
	12.6408008442	0.0000233699

8.3.1.10 Gauss–Hermite quadrature

Weight function is $w(x) = e^{-x^2}$.

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)\,dx \approx \sum_{i=1}^{n} w_i f(x_i).$$

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
2	0.7071067811	0.8862269254
3	0	1.1816359006
	1.2247448713	0.2954089751
4	0.5246476232	0.8049140900
	1.6506801238	0.0813128354

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
5	0	0.9453087204
	0.9585724646	0.3936193231
	2.0201828704	0.0199532420
6	0.4360774119	0.7246295952
	1.3358490740	0.1570673203
	2.3506049736	0.0045300099

8.3.1.11 Radau quadrature

$$\int_{-1}^{1} f(x)\,dx = w_1 f(-1) + \sum_{i=2}^{n} w_i f(x_i) + \frac{2^{2n-1}[n(n-1)!]^4}{[(2n-1)!]^3} f^{(2n-1)}(\xi),$$

where each free node x_i is the i^{th} root of $\frac{P_{n-1}(x)+P_n(x)}{x+1}$ and $w_i = \frac{1-x_i}{n^2[P_{n-1}(x_i)]^2}$ for $i = 2, \ldots, n$, see the following table. Note that $x_1 = -1$ and $w_1 = 2/n^2$.

n	Nodes	Weights $\{w_i\}$
3	−0.2898979485	1.0249716523
	0.6898979485	0.7528061254
4	−0.5753189235	0.6576886399
	0.1810662711	0.7763869376
	0.8228240809	0.4409244223
5	−0.7204802713	0.4462078021
	−0.1671808647	0.6236530459
	0.4463139727	0.5627120302
	0.8857916077	0.2874271215

n	Nodes	Weights $\{w_i\}$
6	−0.8029298284	0.3196407532
	−0.3909285467	0.4853871884
	0.1240503795	0.5209267831
	0.6039731642	0.4169013343
	0.9203802858	0.2015883852
7	−0.8538913426	0.2392274892
	−0.5384677240	0.3809498736
	−0.1173430375	0.4471098290
	0.3260306194	0.4247037790
	0.7038428006	0.3182042314
	0.9413671456	0.1489884711

8.3.1.12 Lobatto quadrature

$$\int_{-1}^{1} f(x)\,dx = w_1 f(-1) + w_n f(1) + \sum_{i=2}^{n-1} w_i f(x_i) - \frac{n(n-1)^3 2^{2n-1}[(n-2)!]^4}{(2n-1)[(2n-2)!]^3} f^{(2n-2)}(\xi)$$

where x_i is the $(i-1)^{\text{st}}$ root of $P'_{n-1}(x)$ and $w_i = \frac{2}{n(n-1)[P_{n-1}(x_i)]^2}$ for $i = 2, \ldots, n-1$. Note that $x_1 = -1$, $x_n = 1$, and $w_1 = w_n = 2/(n(n-1))$.

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
3	0	1.3333333333
	1	0.3333333333
4	0.4472135954	0.8333333333
	1	0.1666666666
5	0	0.7111111111
	0.6546536707	0.5444444444
	1	0.1000000000

n	Nodes $\{\pm x_i\}$	Weights $\{w_i\}$
6	0.2852315164	0.5548583770
	0.7650553239	0.3784749562
	1	0.0666666666
7	0	0.4876190476
	0.4688487934	0.4317453812
	0.8302238962	0.2768260473
	1	0.0476190476

8.3.1.13 Chebyshev quadrature

$$\int_{-1}^{1} f(x)\,dx \approx \frac{2}{n}\sum_{i=1}^{n} f(x_i).$$

n	Nodes $\{\pm x_i\}$
2	0.5773502691
3	0
	0.7071067811
4	0.1875924740
	0.7946544722

n	Nodes $\{\pm x_i\}$
5	0
	0.3745414095
	0.8324974870
6	0.2666354015
	0.4225186537
	0.8662468181

n	Nodes $\{\pm x_i\}$
7	0
	0.3239118105
	0.5296567752
	0.8838617007

8.3.1.14 Multiple integrals

Quadrature methods can be extended to multiple integrals. The general idea, using a double integral as an example, involves writing the double integral in the form of an iterated integral, applying the quadrature method to the "inner integral" and then applying the method to the "outer integral."

8.3.1.15 Simpson's double integral over a rectangle

To integrate $f(x, y)$ over a rectangular region $R = \{(x, y) \mid a \le x \le b, c \le y \le d\}$ using the composite Simpson's Rule produces the formula given below. Intervals $[a, b]$ and $[c, d]$ must be partitioned using even integers n and m for evenly spaced mesh points $x_0, x_1, \ldots, x_n$ and $y_0, y_1, \ldots, y_m$, respectively.

$$\int\!\!\int_R f(x,y)\,dydx = \int_a^b \int_c^d f(x,y)\,dydx = \frac{hk}{9}\sum_{i=0}^{n}\sum_{j=0}^{m} c_{i,j} f(x_i, y_j) + E \tag{8.3.5}$$

where the error term E is (here, $h = (b-a)/n$ and $k = (d-c)/m$)

$$E = -\frac{(d-c)(b-a)}{180}\left[h^4 \frac{\partial^4 f}{\partial x^4}(\bar\eta, \bar\mu) + k^4 \frac{\partial^4 f}{\partial y^4}(\hat\eta, \hat\mu)\right] \tag{8.3.6}$$

for some $(\bar\eta, \bar\mu)$ and $(\hat\eta, \hat\mu)$ in R. The coefficients $\{c_{i,j}\}$ are:

m	1	4	2	4	2	4	$\ldots$	2	4	1
$m-1$	4	16	8	16	8	16	$\ldots$	8	16	4
$m-2$	2	8	4	8	4	8	$\ldots$	4	8	2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$
2	2	8	4	8	4	8	$\ldots$	4	8	2
1	4	16	8	16	8	16	$\ldots$	8	16	4
0	1	4	2	4	2	4	$\ldots$	2	4	1
j / i	0	1	2	3	4	5	$\ldots$	$n-2$	$n-1$	n

Similarly, Simpson's Rule can be extended for regions that are not rectangular. It is simpler to give the following algorithm than to state a general formula.

8.3.1.16 Simpson's double integral algorithm

To approximate the integral $I = \int_a^b \int_{c(x)}^{d(x)} f(x,y)\,dy\,dx$:

INPUT endpoints a, b; even positive integers $m,\ n$;
functions $c(x)$, $d(x)$, and $f(x,y)$
OUTPUT approximation J to I.
Algorithm:

1. Set $h = (b-a)/n$; $J_1 = 0$; (End terms.)
   $J_2 = 0$; (Even terms.) $J_3 = 0$. (Odd terms.)
2. For $i = 0, 1, ..., n$ do (a)–(d).
   (a) Set $x = a + ih$; (Composite Simpson's method for x.)
   $HX = (d(x) - c(x))/m$;
   $K_1 = f(x, c(x)) + f(x, d(x))$; (End terms.)
   $K_2 = 0$; (Even terms.)
   $K_3 = 0$. (Odd terms.)
   (b) For $j = 1, 2, ..., m-1$ do i.–ii.
   i. Set $y = c(x) + jHX$; $Q = f(x,y)$.
   ii. If j is even then set $K_2 = K_2 + Q$ else set $K_3 = K_3 + Q$.
   (c) Set $L = (K_1 + 2K_2 + 4K_3)HX/3$.
   $\left(L \approx \int_{c(x_i)}^{d(x_i)} f(x_i, y)\,dy \text{by composite Simpson's method.}\right)$
   (d) If $i = 0$ or $i = n$ then set $J_1 = J_1 + L$;
   else if i is even then set $J_2 = J_2 + L$;
   else set $J_3 = J_3 + L$.
3. Set $J = h(J_1 + 2J_2 + 4J_3)/3$.
4. OUTPUT(J); STOP.

8.3.1.17 Gaussian double integral

To apply Gaussian quadrature to $I = \int_a^b \int_{c(x)}^{d(x)} f(x,y)\,dy\,dx$, first requires transforming, for each x in $[a,b]$, the interval $[c(x), d(x)]$ to $[-1, 1]$ and then applying Gaussian quadrature. This is performed in the following algorithm.

8.3.1.18 Gauss–Legendre double integral

To approximate the integral $I = \int_a^b \int_{c(x)}^{d(x)} f(x,y)\,dy\,dx$:

INPUT endpoints a, b; positive integers $m,\ n$; the needed roots $r_{i,j}$ and coefficients $c_{i,j}$ for $i = \max\{m,n\}$ and for $1 \le j \le i$.
(From table on page 672)
OUTPUT approximation J to I.

Algorithm:

1. Set $h_1 = (b-a)/2$; $h_2 = (b+a)/2$; $J = 0$.
2. For $i = 1, 2, ..., m$ do (a)–(c).
   (a) Set $JX = 0$;
   $x = h_1 r_{m,i} + h_2$;
   $d_1 = d(x)$; $c_1 = c(x)$;
   $k_1 = (d_1 - c_1)/2$; $k_2 = (d_1 + c_1)/2$.
   (b) For $j = 1, 2, ..., n$ do
   set $y = k_1 r_{n,j} + k_2$;
   $Q = f(x, y)$;
   $JX = JX + c_{n,j} Q$.
   (c) Set $J = J + c_{m,i} k_1 JX$.
3. Set $J = h_1 J$.
4. OUTPUT(J); STOP.

8.3.1.19 Double integrals of polynomials over polygons

If the vertices of the polygon A are $\{(x_1, y_1), (x_2, y_2), \ldots, (x_p, y_p)\}$, and we define $w_i = x_i y_{i+1} - x_{i+1} y_i$ (with $x_{p+1} = x_1$ and $y_{p+1} = y_1$) then (see Marin, page 807)

$$\int\int_A x^m y^n \, dA = \tag{8.3.7}$$
$$\frac{m!n!}{(m+n+2)!} \sum_{i=1}^{p} w_i \sum_{j=0}^{m} \sum_{k=0}^{n} \binom{j+k}{j} \binom{m+n-j-k}{n-k} x_i^{m-j} x_{i+1}^{j} y_i^{n-k} y_{i+1}^{k}.$$

8.3.1.20 Monte–Carlo methods

Monte–Carlo methods, in general, involve the generation of random numbers (actually *pseudorandom* when computer-generated) to represent independent, uniform random variables over $[0, 1]$. Section 7.6 describes random number generation. Such a simulation can provide insight into the solutions of very complex problems.

Monte–Carlo methods are generally not competitive with other numerical methods of this section. However, if the function fails to have continuous derivatives of moderate order, those methods may not be applicable. One advantage of Monte–Carlo methods is that they extend to multidimensional integrals quite easily, although here only a few techniques for one-dimensional integrals $I = \int_a^b g(x)\, dx$ are given.

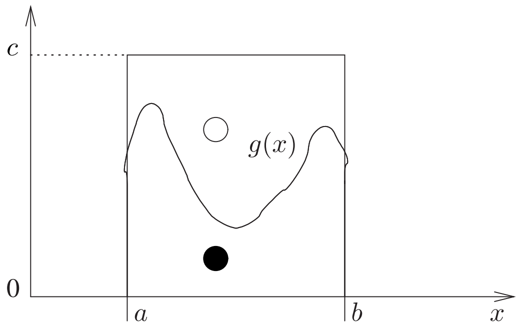
8.3.1.21 Hit or miss method

Suppose $0 \le g(x) \le c$, $a \le x \le b$, and $\Omega = \{(x, y) \mid a \le x \le b, 0 \le y \le c\}$. If (X, Y) is a random vector which is uniformly distributed over Ω, then the probability p that (X, Y) lies in S (see Figure 8.3) is $p = I/(c(b-a))$.

If N independent random vectors $\{(X_i, Y_i)\}_{i=1}^{N}$ are generated, the parameter p can be estimated by $\hat{p} = N_H/N$ where N_H is the number of times $Y_i \le g(X_i)$, $i = 1, 2, \ldots, N$, called the number of *hits*. (Likewise $N - N_H$ is the number of *misses*.) The value of I is then estimated by the unbiased estimator $\theta_1 = c(b-a)N_H/N$.

FIGURE 8.3

Illustration of the Monte–Carlo method. The sample points are shown as circles; the solid circle is counted as a "hit," the empty circle is counted as a "miss."



8.3.1.22 Hit or miss algorithm

1. Generate $\{U_j\}_{j=1}^{2N}$ of $2N$ random numbers, uniformly distributed in $[0, 1)$.
2. Arrange the sequence into N pairs (U_1, U'_1), (U_2, U'_2), ..., (U_N, U'_N), so that each U_j is used exactly once.
3. Compute $X_i = a + U_i(b - a)$ and $g(X_i)$ for $i = 1, 2, \ldots, N$.
4. Count the number of cases N_H for which $g(X_i) \geq cU'_i$.
5. Compute $\theta_1 = c(b - a)N_H/N$. (This is an estimate of I.)

The number of trials N necessary for $P(|\theta_1 - I| < \epsilon) \geq \alpha$ is given by

$$N \geq \frac{(1-p)p[c(b-a)]^2}{(1-\alpha)\epsilon^2}. \tag{8.3.8}$$

With the usual notation of z_α for the value of the standard normal random variable Z for which $P(Z > z_\alpha) = \alpha$ (see page 628), a confidence interval for I with confidence level $1 - \alpha$ is

$$\theta_1 \pm z_{\frac{\alpha}{2}} \frac{\sqrt{\hat{p}(1-\hat{p})}(b-a)c}{\sqrt{N}}. \tag{8.3.9}$$

8.3.1.23 The sample-mean Monte–Carlo method

Write the integral $I = \int_a^b g(x)\,dx$ as $\int_a^b \frac{g(x)}{f_X(x)} f_X(x)\,dx$, where f is any probability density function for which $f_X(x) > 0$ when $g(x) \neq 0$. Then $I = \mathrm{E}\left[\frac{g(X)}{f_X(X)}\right]$ where the random variable X is distributed according to $f_X(x)$. Values from this distribution can be generated by the methods discussed in Section 7.6.2. For the case where $f_X(x)$ is the uniform distribution on $[0, 1]$, $I = (b - a)\mathrm{E}[g(X)]$. An unbiased estimator of I is its sample mean

$$\theta_2 = (b-a)\frac{1}{N}\sum_{i=1}^{N} g(X_i). \tag{8.3.10}$$

It follows that the variance of θ_2 is less than or equal to the variance of θ_1. In fact,

$$\begin{aligned} \operatorname{var}\theta_1 &= \frac{I}{N}[c(b-a)-I], \\ \operatorname{var}\theta_2 &= \frac{1}{N}\left[(b-a)\int_a^b g^2(x)\,dx - I^2\right]. \end{aligned} \tag{8.3.11}$$

Note that to estimate I with θ_1 or θ_2, $g(x)$ is not needed explicitly. It is only necessary to evaluate $g(x)$ at any point x.

8.3.1.24 Sample-mean algorithm

1. Generate $\{U_i\}_{i=1}^N$ of N random numbers, uniformly distributed in $[0,1)$.
2. Compute $X_i = a + U_i(b-a)$, for $i = 1,2,\ldots,N$.
3. Compute $g(X_i)$ for $i = 1,2,\ldots,N$.
4. Compute θ_2 according to Equation (8.3.10) (This is an estimate of I).

8.3.1.25 Integration in the presence of noise

Suppose $g(x)$ is measured with some error: $\tilde{g}(x_i) = g(x_i) + \epsilon_i$, for $i = 1,2,\ldots,N$, where ϵ_i are independent identically distributed random variables with $\mathrm{E}[\epsilon_i] = 0$, $\operatorname{var}(\epsilon_i) = \sigma^2$, and $|\epsilon_i| < k < \infty$.

If (X,Y) is uniformly distributed on the rectangle $a \le x \le b$, $0 \le y \le c_1$, where $c_1 \ge g(x) + k$, set $\tilde{\theta_1} = c_1(b-a)N_H/N$ as in the hit or miss method. Similarly, set $\tilde{\theta_2} = \frac{1}{N}(b-a)\sum_{i=1}^N \tilde{g}(X_i)$ as in the sample-mean method. Then both $\tilde{\theta_1}$ and $\tilde{\theta_2}$ are unbiased and converge almost surely to I. Again, $\operatorname{var}\tilde{\theta_2} \le \operatorname{var}\tilde{\theta_1}$.

8.3.1.26 Weighted Monte–Carlo integration

Estimate the integral $I = \int_0^1 g(x)\,dx$ according to the following algorithm:

1. Generate numbers $\{U_1, U_2, \ldots, U_N\}$ from the uniform distribution on $[0,1)$.
2. Arrange $U_1, U_2, \ldots, U_N$ in the increasing order $U_{(1)}, U_{(2)}, \ldots, U_{(N)}$.
3. Compute $\theta_3 = \dfrac{1}{2}\left[\sum_{i=0}^{N}(g(U_{(i)}) + g(U_{(i+1)})(U_{(i+1)} - U_{(i)})\right]$, where $U_{(0)} \equiv 0$, $U_{(N+1)} \equiv 1$. This is an estimate of I.

If $g(x)$ has a continuous second derivative on $[0,1]$, then the estimator θ_3 satisfies $\operatorname{var}\theta_3 = \mathrm{E}\left[(\theta_3 - I)^2\right] \le k/N^4$, where k is some positive constant.

8.3.2 NUMERICAL DIFFERENTIATION

8.3.2.1 Derivative estimates

Selected formulas to estimate the derivative of a function at a single point, with error terms, are given. Nodes are equally spaced with $x_i - x_{i-1} = h$; h may be positive or negative and, in the error formulas, ξ lies between the smallest and largest nodes. To shorten some of the formulas, f_j is used to denote $f(x_0 + jh)$ and some error formulas are expressed as $O(h^k)$.

1. Two-point formula for $f'(x_0)$

$$f'(x_0) = \frac{1}{h}(f(x_0+h) - f(x_0)) - \frac{h}{2}f''(\xi). \tag{8.3.12}$$

   This is called the *forward-difference formula* if $h > 0$ and the *backward-difference formula* if $h < 0$.

2. Three-point formulas for $f'(x_0)$

$$\begin{aligned} f'(x_0) &= \frac{1}{2h}[-3f(x_0) + 4f(x_0+h) - f(x_0+2h)] + \frac{h^2}{3}f^{(3)}(\xi) \\ &= \frac{1}{2h}[f(x_0+h) - f(x_0-h)] - \frac{h^2}{6}f^{(3)}(\xi). \end{aligned} \tag{8.3.13}$$

3. Four-point formula (or five uniformly spaced points) for $f'(x_0)$

$$f'(x_0) = \frac{1}{12h}[f_{-2} - 8f_{-1} + 8f_1 - f_2] + \frac{h^4}{30}f^{(5)}(\xi). \tag{8.3.14}$$

4. Five-point formula for $f'(x_0)$

$$f'(x_0) = \frac{1}{12h}[-25f_0 + 48f_1 - 36f_2 + 16f_3 - 3f_4] + \frac{h^4}{5}f^{(5)}(\xi). \tag{8.3.15}$$

5. Formulas for the second derivative

$$\begin{aligned} f''(x_0) &= \frac{1}{h^2}[f_{-1} - 2f_0 + f_1] - \frac{h^2}{12}f^{(4)}(\xi), \\ &= \frac{1}{h^2}[f_0 - 2f_1 + f_2] + \frac{h^2}{6}f^{(4)}(\xi_1) - hf^{(3)}(\xi_2). \end{aligned} \tag{8.3.16}$$

6. Formulas for the third derivative

$$\begin{aligned} f^{(3)}(x_0) &= \frac{1}{h^3}[f_3 - 3f_2 + 3f_1 - f_0] + O(h), \\ &= \frac{1}{2h^3}[f_2 - 2f_1 + 2f_{-1} - f_{-2}] + O(h^2). \end{aligned} \tag{8.3.17}$$

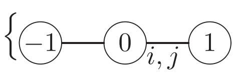
7. Formulas for the fourth derivative

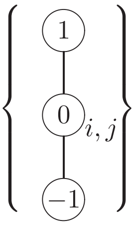
$$\begin{aligned} f^{(4)}(x_0) &= \frac{1}{h^4}[f_4 - 4f_3 + 6f_2 - 4f_1 + f_0] + O(h), \\ &= \frac{1}{h^4}[f_2 - 4f_1 + 6f_0 - 4f_{-1} + f_{-2}] + O(h^2). \end{aligned} \tag{8.3.18}$$

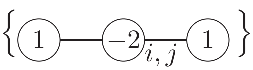
Richardson's extrapolation can be applied to improve estimates. The error term of the formula must satisfy Equation (8.1.2) and an extrapolation procedure must be developed. As a special case, however, Equation (8.1.4) may be used when first-column entries are generated by Equation (8.3.13).

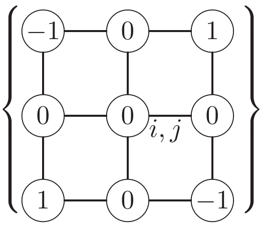
8.3.2.2 Computational molecules

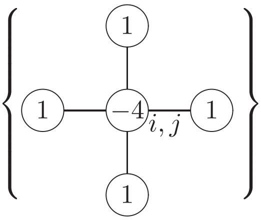
A computational molecule is a graphical depiction of an approximate partial derivative formula. The following computational molecules are for $h = \Delta x = \Delta y$:

(a) $\left.\dfrac{du}{dx}\right|_{i,j} = \dfrac{1}{h^2}$  $+ O\left(h^2\right)$

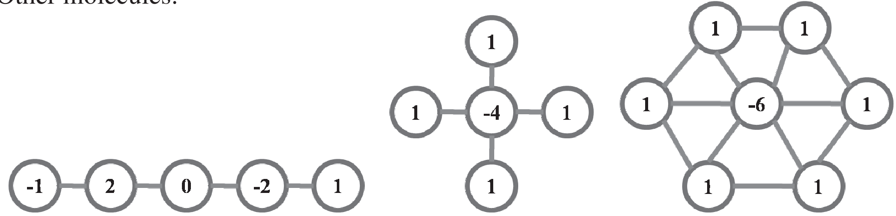
(b) $\left.\dfrac{du}{dy}\right|_{i,j} = \dfrac{1}{h^2}$  $+ O\left(h^2\right)$

(c) $\left.\dfrac{d^2u}{dx^2}\right|_{i,j} = \dfrac{1}{4h^2}$  $+ O\left(h^2\right)$

(d) $\left.\dfrac{d^2u}{dx\,dy}\right|_{i,j} = \dfrac{1}{h^2}$  $+ O\left(h^2\right)$

(e) $\left.\nabla^2 u\right|_{i,j} = \dfrac{1}{h^2}$  $+ O\left(h^2\right)$

Other molecules:



$2h^2 \dfrac{\partial^2}{\partial x^2}$ $\qquad h^2\nabla^2$ $\qquad \dfrac{3}{2}h^2\nabla^2$

8.3.2.3 Numerical solution of differential equations

Numerical methods to solve differential equations depend on whether small changes in the statement of the problem cause small changes in the solution.

Definition The initial value problem,

$$\frac{dy}{dt} = f(t, y), \quad a \le t \le b, \quad y(a) = \alpha, \tag{8.3.19}$$

is said to be *well-posed* if

1. A unique solution, $y(t)$, to the problem exists.
2. For any $\epsilon > 0$, there exists a positive constant $k(\epsilon)$ with the property that, whenever $|\epsilon_0| < \epsilon$ and $\delta(t)$ is continuous with $|\delta(t)| < \epsilon$ on $[a, b]$, a unique solution, $z(t)$, to the problem,
$$\frac{dz}{dt} = f(t, z) + \delta(t), \quad a \le t \le b, \quad z(a) = \alpha + \epsilon_0,$$
exists and satisfies $|z(t) - y(t)| < k(\epsilon)\epsilon$, for all $a \le t \le b$.

This is called the *perturbed problem* associated with the original problem. Although other criteria exist, the following result gives conditions that are easy to check to guarantee that a problem is well-posed.

THEOREM 8.3.1 *(Well-posed condition)*
Suppose that f and f_y (its first partial derivative with respect to y) are continuous for t in $[a, b]$. Then the initial value problem given by Equation (8.3.19) is well-posed.

Using Taylor's theorem, numerical methods for solving the well-posed, first-order differential equation given by Equation (8.3.19) can be derived. Using equally-spaced *mesh points* $t_i = a + ih$ (for $i = 0, 1, 2, \ldots, N$) and w_i to denote an approximation to $y_i \equiv y(t_i)$, then methods generally use *difference equations* of the form

$$w_0 = \alpha, \qquad w_{i+1} = w_i + h\phi(t_i, w_i),$$

for each $i = 0, 1, 2, \ldots, N-1$. Here ϕ is a function depending on f. The difference method has *local truncation error* given by

$$\tau_{i+1}(h) = \frac{y_{i+1} - y_i}{h} - \phi(t_i, y_i),$$

for each $i = 0, 1, 2, \ldots, N-1$. The following formulas are called *Taylor methods*. Each has local truncation error

$$\frac{h^n}{(n+1)!} f^{(n)}(\xi_i, y(\xi_i)) = \frac{h^n}{(n+1)!} y^{(n+1)}(\xi_i, y(\xi_i))$$

for each $i = 0, 1, 2, \ldots, N-1$, where $\xi_i \in (t_i, t_{i+1})$. Thus, if $y \in C^{n+1}[a, b]$ the local truncation error is $O(h^n)$.

1. Euler's method ($n = 1$):

$$w_{i+1} = w_i + hf(t_i, w_i). \tag{8.3.20}$$

2. Taylor method of order n:

$$w_{i+1} = w_i + hT^{(n)}(t_i, w_i), \tag{8.3.21}$$

where $T^{(n)}(t_i, w_i) = f(t_i, w_i) + \frac{h}{2} f'(t_i, w_i) + \ldots + \frac{h^{n-1}}{n!} f^{(n-1)}(t_i, w_i)$.
The *Runge–Kutta methods* below are derived from the n^{th} degree Taylor polynomial in two variables.

3. Midpoint method:

$$w_{i+1} = w_i + h\left[f\left(t_i + \tfrac{h}{2}, w_i + \tfrac{h}{2} f(t_i, w_i)\right)\right]. \tag{8.3.22}$$

If all second-order partial derivatives of f are bounded, this method has local truncation error $O(h^2)$, as do the following two methods.

4. Modified Euler method:

$$w_{i+1} = w_i + \tfrac{h}{2}\left\{f(t_i, w_i) + f[t_{i+1}, w_i + hf(t_i, w_i)]\right\}. \tag{8.3.23}$$

5. Heun's method:

$$w_{i+1} = w_i + \frac{h}{4}\left\{f(t_i, w_i) + 3f\left[t_i + \frac{2}{3}h, w_i + \frac{2}{3}hf(t_i, w_i)\right]\right\}.$$

6. Runge–Kutta method of order four:

$$w_{i+1} = w_i + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4), \tag{8.3.24}$$

where

$$\begin{aligned}
k_1 &= hf\left(t_i, w_i\right), \\
k_2 &= hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_1\right), \\
k_3 &= hf\left(t_i + \frac{h}{2}, w_i + \frac{1}{2}k_2\right), \\
k_4 &= hf\left(t_{i+1}, w_i + k_3\right).
\end{aligned}$$

The local truncation error is $O(h^4)$ if the solution $y(t)$ has five continuous derivatives.

8.3.2.4 Multistep methods and predictor-corrector methods

A *multistep method* is a technique whose difference equation to compute w_{i+1} involves more prior values than just w_i. An *explicit method* is one in which the computation of w_{i+1} does not depend on $f(t_{i+1}, w_{i+1})$ whereas an *implicit method* does involve $f(t_{i+1}, w_{i+1})$. For each formula, $i = n-1, n, \dots, N-1$.

8.3.2.5 Adams–Bashforth n-step (explicit) methods

1. $(n = 2)$: $w_0 = \alpha, w_1 = \alpha_1, w_{i+1} = w_i + \frac{h}{2}[3f(t_i, w_i) - f(t_{i-1}, w_{i-1})]$.
   The local truncation error is $\tau_{i+1}(h) = \frac{5}{12}y^{(3)}(\mu_i)h^2$, for some $\mu_i \in (t_{i-1}, t_{i+1})$.
2. $(n = 3)$: $w_0 = \alpha, w_1 = \alpha_1, w_2 = \alpha_2$,
   $$w_{i+1} = w_i + \frac{h}{12}[23f(t_i, w_i) - 16f(t_{i-1}, w_{i-1}) + 5f(t_{i-2}, w_{i-2})].$$
   The local truncation error is $\tau_{i+1}(h) = \frac{3}{8}y^{(4)}(\mu_i)h^3$, for some $\mu_i \in (t_{i-2}, t_{i+1})$.
3. $(n = 4)$: $w_0 = \alpha$, $w_1 = \alpha_1$, $w_2 = \alpha_2$, $w_3 = \alpha_3$, $w_{i+1} = w_i + \frac{h}{24}[55f(t_i, w_i) - 59f(t_{i-1}, w_{i-1}) + 37f(t_{i-2}, w_{i-2}) - 9f(t_{i-3}, w_{i-3})]$.
   The local truncation error is $\tau_{i+1}(h) = \frac{251}{720}y^{(5)}(\mu_i)h^4$, for some $\mu_i \in (t_{i-3}, t_{i+1})$.

8.3.2.6 Adams–Moulton n-step (implicit) methods

1. $(n = 2)$: $w_0 = \alpha, w_1 = \alpha_1$,
   $$w_{i+1} = w_i + \frac{h}{12}[5f(t_{i+1}, w_{i+1}) + 8f(t_i, w_i) - f(t_{i-1}, w_{i-1})].$$
   The local truncation error is $\tau_{i+1}(h) = -\frac{1}{24}y^{(4)}(\mu_i)h^3$, for some $\mu_i \in (t_{i-1}, t_{i+1})$.
2. $(n = 3)$: $w_0 = \alpha, w_1 = \alpha_1, w_2 = \alpha_2, w_{i+1} = w_i + \frac{h}{24}[9f(t_{i+1}, w_{i+1}) + 19f(t_i, w_i) - 5f(t_{i-1}, w_{i-1}) + f(t_{i-2}, w_{i-2})]$.
   The local truncation error is $\tau_{i+1}(h) = -\frac{19}{720}y^{(5)}(\mu_i)h^4$, for some $\mu_i \in (t_{i-2}, t_{i+1})$.
3. $(n = 4)$: $w_0 = \alpha, w_1 = \alpha_1, w_2 = \alpha_2, w_3 = \alpha_3$,
   $$w_{i+1} = w_i + \frac{h}{720}[251f(t_{i+1}, w_{i+1}) + 646f(t_i, w_i) - 264f(t_{i-1}, w_{i-1}) + 106f(t_{i-2}, w_{i-2}) - 19f(t_{i-3}, w_{i-3})].$$
   The local truncation error is $\tau_{i+1}(h) = -\frac{3}{160}y^{(6)}(\mu_i)h^5$, for some $\mu_i \in (t_{i-3}, t_{i+1})$.

In practice, implicit methods are not used by themselves. They are used to improve approximations obtained by explicit methods. An explicit method *predicts* an approximation and the implicit method *corrects* this prediction. The combination is called a *predictor-corrector method.* For example, the Adams–Bashforth method with $n = 4$ might be used with the Adams–Moulton method with $n = 3$ since both have comparable errors. Initial values may be computed, say, by the Runge–Kutta method of order four, Equation (8.3.24).

8.3.2.7 Higher-order differential equations and systems

A system of m first-order initial value problems can be expressed in the form

$$\begin{aligned}\frac{du_1}{dt} &= f_1(t, u_1, u_2, \ldots, u_m), \qquad u_1(a) = \alpha_1,\\ \frac{du_2}{dt} &= f_2(t, u_1, u_2, \ldots, u_m), \qquad u_2(a) = \alpha_2,\\ &\vdots\\ \frac{du_m}{dt} &= f_m(t, u_1, u_2, \ldots, u_m), \qquad u_m(a) = \alpha_m.\end{aligned} \tag{8.3.25}$$

Generalizations of methods for solving first-order equations can be used to solve such systems. An example here uses the Runge-Kutta method of order four.

Partition $[a, b]$ as before, and let $w_{i,j}$ denote the approximation to $u_i(t_j)$ for $j = 0, 1, \ldots, N$ and $i = 1, 2, \ldots, m$. For the initial conditions, set $w_{1,0} = \alpha_1$, $w_{2,0} = \alpha_2$, ..., $w_{m,0} = \alpha_m$. From the values $\{w_{1,j}, w_{2,j}, \ldots, w_{m,j}\}$ previously computed, obtain $\{w_{1,j+1}, w_{2,j+1}, \ldots, w_{m,j+1}\}$ from

$$\begin{aligned}k_{1,i} &= hf_i\left(t_j, w_{1,j}, w_{2,j}, \ldots, w_{m,j}\right),\\ k_{2,i} &= hf_i\left(t_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{1,1}, w_{2,j} + \frac{1}{2}k_{1,2}, \ldots, w_{m,j} + \frac{1}{2}k_{1,m}\right),\\ k_{3,i} &= hf_i\left(t_j + \frac{h}{2}, w_{1,j} + \frac{1}{2}k_{2,1}, w_{2,j} + \frac{1}{2}k_{2,2}, \ldots, w_{m,j} + \frac{1}{2}k_{2,m}\right),\\ k_{4,i} &= hf_i\left(t_j + h, w_{1,j} + k_{3,1}, w_{2,j} + k_{3,2}, \ldots, w_{m,j} + k_{3,m}\right),\\ w_{i,j+1} &= w_{i,j} + \frac{1}{6}[k_{1,i} + 2k_{2,i} + 2k_{3,i} + k_{4,i}],\end{aligned} \tag{8.3.26}$$

where $i = 1, 2, \ldots, m$ for each of the above.

A differential equation of high order can be converted into a *system* of first-order equations. Suppose that a single differential equation has the form

$$y^{(m)} = f(t, y, y', y'', \ldots, y^{(m-1)}), \quad a \le t \le b \tag{8.3.27}$$

with initial conditions $y(a) = \alpha_1, y'(a) = \alpha_2, \ldots, y^{(m-1)}(a) = \alpha_m$. All derivatives are with respect to t. That is, $y^{(k)} = \frac{d^k y}{dt^k}$. Define $u_1(t) = y(t)$, $u_2(t) = y'(t)$, ..., $u_m(t) = y^{(m-1)}(t)$. This yields first-order equations

$$\frac{du_1}{dt} = u_2, \quad \frac{du_2}{dt} = u_3, \quad \ldots \quad \frac{du_{m-1}}{dt} = u_m, \quad \frac{du_m}{dt} = f(t, u_1, u_2, \ldots, u_m), \tag{8.3.28}$$

with initial conditions $u_1(a) = \alpha_1, \ldots, u_m(a) = \alpha_m$.

8.3.2.8 Partial differential equations

To develop difference equations for partial differential equations, one needs to estimate the partial derivatives of a function, say, $u(x, y)$. For example,

$$\frac{\partial u}{\partial x}(x,y) = \frac{u(x+h,y)-u(x,y)}{h} - \frac{h}{2}\frac{\partial^2 u(\xi,y)}{\partial x^2} \quad \text{for } \xi \in (x, x+h), \tag{8.3.29}$$

$$\frac{\partial^2 u}{\partial x^2}(x,y) = \frac{1}{h^2}[u(x+h,y)-2u(x,y)+u(x-h,y)] - \frac{h^2}{12}\frac{\partial^4 u(\xi,y)}{\partial x^4}, \quad \text{for } \xi \in (x-h, x+h). \tag{8.3.30}$$

Notes:

1. Equation (8.3.29) is simply Equation (8.3.12) applied to estimate the partial derivative. It is given here to emphasize its application for forming difference equations for partial differential equations. A similar formula applies for $\partial u/\partial y$, and others could follow from the formulas in Section 8.3.2.1.
2. An estimate of $\partial^2 u/\partial y^2$ is similar. A formula for $\partial^2 u/\partial x \partial y$ could be given. However, in practice, a change of variables is generally used to eliminate this mixed second partial derivative from the problem.

If a partial differential equation involves partial derivatives with respect to only one of the variables, the methods described for ordinary differential equations can be used. If, however, the equation involves partial derivatives with respect to both variables, the approximation of the partial derivatives require increments in both variables. The corresponding difference equations form a *system* of linear equations that must be solved.

Three specific forms of partial differential equations with popular methods of solution are given. The domains are assumed to be rectangular.

8.3.2.9 Wave equation

The wave equation is an example of a hyperbolic partial differential equation and has the form

$$\frac{\partial^2 u}{\partial t^2}(x,t) - \alpha^2 \frac{\partial^2 u}{\partial x^2}(x,t) = 0, \quad 0 < x < \ell, \quad t > 0. \tag{8.3.31}$$

(where α is a constant) subject to $u(0,t) = u(\ell,t) = 0$ for $t > 0$, and $u(x,0) = f(x)$ and $\dfrac{\partial u}{\partial t}(x,0) = g(x)$ for $0 \le x \le \ell$.

Select an integer $m > 0$, time-step size $k > 0$, and using $h = \ell/m$, mesh points (x_i, t_j) are defined by $x_i = ih$ and $t_j = jk$. Using $w_{i,j}$ to represent an approximation of $u(x_i, t_j)$ and $\lambda = \alpha k/h$, the difference equation becomes

$$w_{i,j+1} = 2(1-\lambda^2)w_{i,j} + \lambda^2(w_{i+1,j} + w_{i-1}, j) - w_{i,j-1},$$

with $w_{0,j} = w_{m,j} = 0$ and $w_{i,0} = f(x_i)$, for $i = 1, \dots, m-1$ and $j = 1, 2, \dots$. Also needed is an estimate for $w_{i,1}$, for each $i = 1, \dots, m-1$, which can be written

$$w_{i,1} = (1-\lambda^2)f(x_i) + \frac{\lambda^2}{2}f(x_{i+1}) + \frac{\lambda^2}{2}f(x_{i-1}) + kg(x_i).$$

The local truncation error of the method is $O(h^2 + k^2)$ but the method is extremely accurate if the true solution is infinitely differentiable. For the method to be stable, it is necessary that $\lambda \leq 1$.

8.3.2.10 Poisson equation

The Poisson equation is an elliptic partial differential equation that has the form

$$\nabla^2 u(x,y) = \frac{\partial^2 u}{\partial x^2}(x,y) + \frac{\partial^2 u}{\partial y^2}(x,y) = f(x,y) \tag{8.3.32}$$

for $(x,y) \in R = \{(x,y) \mid a < x < b, c < y < d\}$, with $u(x,y) = g(x,y)$ for $(x,y) \in S$, where $S = \partial R$. When the function $f(x,y) = 0$ the equation is called *Laplace's equation*.

To begin, partition $[a,b]$ and $[c,d]$ by choosing integers n and m, define step sizes $h = (b-a)/n$ and $k = (d-c)/m$, and set $x_i = a + ih$ for $i = 0, 1, \ldots, n$ and $y_j = c + jk$ for $j = 0, 1, \ldots, m$. The lines $x = x_i$, $y = y_j$, are called *grid lines* and their intersections are called *mesh points*. Estimates $w_{i,j}$ for $u(x_i, y_j)$ can be generated using Equation (8.3.30) to estimate $\frac{\partial^2 u}{\partial x^2}$ and $\frac{\partial^2 u}{\partial y^2}$. The method described here is called the *finite-difference method*.

Start with the values

$$w_{0,j} = g(x_0, y_j), \quad w_{n,j} = g(x_n, y_j), \quad w_{i,0} = g(x_i, y_0), \quad w_{i,m} = g(x_i, y_m), \tag{8.3.33}$$

and then solve the resulting system of linear algebraic equations

$$2\left[\left(\frac{h}{k}\right)^2 + 1\right] w_{i,j} - (w_{i+1,j} + w_{i-1,j}) - \left(\frac{h}{k}\right)^2 (w_{i,j+1} + w_{i,j-1}) = -h^2 f(x_i, y_j), \tag{8.3.34}$$

for $i = 1, 2, \ldots, n-1$ and $j = 1, 2, \ldots, m-1$. The local truncation error is $O(h^2 + k^2)$.

If the interior mesh points are labeled $P_\ell = (x_i, y_j)$ and $w_\ell = w_{i,j}$ where $\ell = i + (m-1-j)(n-1)$, for $i = 1, 2, \ldots, n-1$, and $j = 1, 2, \ldots, m-1$, then the two-dimensional array of values becomes a one-dimensional array. This results in a banded linear system. The case $n = m = 4$ yields $\ell = (n-1)(m-1) = 9$. Using the relabeled grid points, $f_\ell = f(P_\ell)$, the equations at the points P_i are

$$\begin{array}{lrcl}
P_1: & 4w_1 - w_2 - w_4 & = & w_{0,3} + w_{1,4} - h^2 f_1, \\
P_2: & 4w_2 - w_3 - w_1 - w_5 & = & w_{2,4} - h^2 f_2, \\
P_3: & 4w_3 - w_2 - w_6 & = & w_{4,3} + w_{3,4} - h^2 f_3, \\
P_4: & 4w_4 - w_5 - w_1 - w_7 & = & w_{0,2} - h^2 f_4, \\
P_5: & 4w_5 - w_6 - w_4 - w_2 - w_8 & = & 0 - h^2 f_5, \\
P_6: & 4w_6 - w_5 - w_3 - w_9 & = & w_{4,2} - h^2 f_6, \\
P_7: & 4w_7 - w_8 - w_4 & = & w_{0,1} + w_{1,0} - h^2 f_7, \\
P_8: & 4w_8 - w_9 - w_7 - w_5 & = & w_{2,0} - h^2 f_8, \\
P_9: & 4w_9 - w_8 - w_6 & = & w_{3,0} + w_{4,1} - h^2 f_9,
\end{array}$$

where the right-hand sides of the equations are obtained from the boundary conditions.

8.3.2.11 Heat or diffusion equation

The heat, or diffusion, equation is a parabolic partial differential equation of the form

$$\frac{\partial u}{\partial t}(x,t) = \alpha^2 \frac{\partial^2 u}{\partial x^2}(x,t), \quad 0 < x < \ell, \quad t > 0. \tag{8.3.35}$$

where $u(0,t) = 0 = u(\ell,t) = 0$, for $t > 0$, and $u(x,0) = f(x)$, for $0 \le x \le \ell$. An efficient method for solving this type of equation is the *Crank–Nicolson method*.

To apply the method, select an integer $m > 0$, set $h = \ell/m$, and select a time-step size k. Here $x_i = ih,\ i = 0,\dots,m$ and $t_j = jk,\ j = 0,\dots.$ The difference equation is given by:

$$\begin{aligned} &\frac{w_{i,j+1} - w_{i,j}}{k} \\ &\quad - \frac{\alpha^2}{2}\left[\frac{w_{i+1,j} - 2w_{i,j} + w_{i-1,j}}{h^2} + \frac{w_{i+1,j+1} - 2w_{i,j+1} + w_{i-1,j+1}}{h^2}\right] = 0. \end{aligned} \tag{8.3.36}$$

and has local truncation error $O(k^2 + h^2)$. The difference equations can be represented in the matrix form $A\mathbf{w}^{(j+1)} = B\mathbf{w}^{(j)}$, for each $j = 0, 1, 2, \dots,$ where $\lambda = \alpha^2 k/h^2$, $\mathbf{w}^{(j)} = (w_{1,j}, w_{2,j}, \dots, w_{m-1,j})^{\mathrm{T}}$, and the matrices A and B are

$$\begin{aligned} A &= \begin{bmatrix} (1+\lambda) & -\lambda/2 & 0 & 0 & \cdots & 0 & 0 \\ -\lambda/2 & (1+\lambda) & -\lambda/2 & 0 & \cdots & 0 & 0 \\ 0 & -\lambda/2 & (1+\lambda) & -\lambda/2 & & 0 & 0 \\ 0 & 0 & -\lambda/2 & (1+\lambda) & & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \\ 0 & 0 & 0 & 0 & & (1+\lambda) & -\lambda/2 \\ 0 & 0 & 0 & 0 & & -\lambda/2 & (1+\lambda) \end{bmatrix} \\ B &= \begin{bmatrix} (1-\lambda) & \lambda/2 & 0 & 0 & \cdots & 0 & 0 \\ \lambda/2 & (1-\lambda) & \lambda/2 & 0 & \cdots & 0 & 0 \\ 0 & \lambda/2 & (1-\lambda) & \lambda/2 & & 0 & 0 \\ 0 & 0 & \lambda/2 & (1-\lambda) & & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \\ 0 & 0 & 0 & 0 & & (1-\lambda) & \lambda/2 \\ 0 & 0 & 0 & 0 & & \lambda/2 & (1-\lambda) \end{bmatrix} \end{aligned} \tag{8.3.37}$$

8.3.3 NUMERICAL SUMMATION

A sum of the form $\sum_{j=0}^{n} f(x_0 + jh)$ (n may be infinite) can be approximated by the *Euler–MacLaurin sum formula*,

$$\sum_{j=0}^{n} f(x_0 + jh) = \frac{1}{h}\int_{x_0}^{x_0+nh} f(y)\,dy + \frac{1}{2}[f(x_0 + nh) + f(x_0)]$$
$$+ \sum_{k=1}^{m} \frac{B_{2k}}{(2k)!} h^{2k-1}[f^{(2k-1)}(x_0 + nh) - f^{(2k-1)}(x_0)] + E_m \quad (8.3.38)$$

where $E_m = \frac{nh^{2m+2}B_{2m+2}}{(2m+2)!} f^{(2m+2)}(\xi)$, with $x_0 < \xi < x_0 + nh$. The B_n here are *Bernoulli numbers* (see Section 1.3.5).

The above formula is useful even when n is infinite, although the error can no longer be expressed in this form. A useful error estimate (which also holds when n is finite) is that the error is less than the magnitude of the first neglected term in the summation on the second line of Equation (8.3.38) if $f^{(2m+2)}(x)$ and $f^{(2m+4)}(x)$ do not change sign and are of the same sign for $x_0 < x < x_0 + nh$. If just $f^{(2m+2)}(x)$ does not change sign in the interval, then the error is less than twice the first neglected term.

Quadrature formulas result from Equation (8.3.38) using estimates for the derivatives.

8.4 PROGRAMMING TECHNIQUES

Efficiency and accuracy are the goals when solving a problem numerically. Below are suggestions to consider when developing algorithms and computer programs.

1. *Every algorithm must have an effective stopping rule.* For example, loops should have an upper bound on the number of iterations.
2. *Avoid the use of arrays whenever possible*. For example, sometimes the same variables can be over-written.
3. *Avoid using formulae that may be highly susceptible to round off error*. For example, subtracting two numbers of similar size may reduce accuracy.
4. *Alter iteration formulae to improve accuracy*. For example, roundoff errors may be reduced if a “small correction” is added to an approximation.
5. *Use pivoting strategies*. These are recommended when solving linear systems to reduce round-off error.
6. *Eliminate unnecessary steps* that increase execution time or round-off error.
7. *Combine numerical methods for efficiency.* For example, use a weaker but reliable method to get near a solution, then use a stronger method to obtain rapid convergence.

Chapter 9

Mathematical Formulas from the Sciences

9.1 ACOUSTICS

Notation

- A total absorption (m^2)
- a side of rectangular waveguide
- b side of rectangular waveguide
- c sound speed ($c = \sqrt{\frac{\kappa}{\rho}}$)
- H height of a room
- L length of a room or organ pipe
- P pressure
- T sound duration
- $\mathbf{v}$ velocity ($\mathbf{v} = \nabla\phi$)
- V room volume (m^3)
- W width of a room
- κ bulk modulus
- ρ density
- ϕ potential
- $\gamma = \frac{c_p}{c_v} = \frac{\text{specific heat at constant pressure}}{\text{specific heat at constant volume}}$ (≈ 1.4 in air)

With the following assumptions

1. The medium is a Newtonian fluid, homogeneous, and at rest (that is, $\overline{\mathbf{v}} = \mathbf{0}$)
2. The medium is an ideal, adiabatic, and reversible gas (that is, $\dfrac{dP}{d\rho} = \dfrac{\gamma p}{\rho}$)
3. There are no body forces and no viscous forces
4. There are small disturbances (that is, the equations are about a steady state)
5. There is irrotational flow (that is, $\nabla \times \mathbf{v} = \mathbf{0}$)

The acoustic wave equation is any one of the following

$$\frac{\partial^2 \mathbf{v}}{\partial t^2} - c^2\nabla^2\mathbf{v} = \mathbf{0} \qquad \frac{\partial^2 P}{\partial t^2} - c^2\nabla^2 P = 0 \qquad \frac{\partial^2 \phi}{\partial t^2} - c^2\nabla^2\phi = 0$$

Equations

1. **Room modes** (where p, q, and r can be $0, 1, 2, \ldots$)

$$\text{supported frequency} = \frac{c}{2}\sqrt{\left(\frac{p}{L}\right)^2 + \left(\frac{q}{W}\right)^2 + \left(\frac{r}{H}\right)^2}$$

2. **Cutoff frequency** (rectangular waveguide, m and n are mode numbers)

$$\omega_c = c\sqrt{\left(\frac{n\pi}{a}\right)^2 + \left(\frac{m\pi}{b}\right)^2}$$

3. **Sabine's Reverberation Formula** $T = 0.161\dfrac{V}{A}$

4. **Organ pipes**
   (a) open at one end and closed at one end
      i. fundamental frequency $n_1 = c/4L$
      ii. first overtone (third harmonic) $n_3 = 3c/4L = 3n_1$
      iii. second overtone (fifth harmonic) $n_5 = 5c/4L = 5n_1$

   (b) open at both ends
      i. fundamental frequency $n_1 = c/2L$
      ii. first overtone (second harmonic) $n_2 = c/L = 2n_1$
      iii. second overtone (third harmonic) $n_3 = 3c/2L = 3n_1$

9.2 ASTROPHYSICS

Notation

- a semi-major axis
- AU astronomical unit
- b semi-minor axis
- b brightness ($\frac{\text{W}}{\text{m}^2}$)
- c speed of light ($\approx 3 \times 10^8$ m/sec)
- d distance to star in parsecs
- D proper distance
- E eccentric anomaly
- G gravitational constant
- h angular momentum
- H_0 Hubble's constant
- L luminosity
- m mass
- M mean anomaly
- p semi-latus rectum
- P orbital period
- r distance
- r_{min} perihelion
- r_{max} aphelion
- t time since perihelion
- v velocity
- ϵ eccentricity
- μ standard gravitational parameter
- θ true anomaly

Equations

1. **Escape velocity** $\quad v \geq \sqrt{\dfrac{2Gm}{r}}$

   Escape velocity from the Earth's surface is 11 km/s.
   Escape velocity from the solar system (from Earth) is 42 km/s.
2. **Hubble's law** $\quad v = H_0 D$
3. **Kepler's equation** $\quad M = E - \epsilon \sin E$
4. **Kepler's laws**

   (a) The orbit of every planet is an ellipse with the Sun at one of the two foci.
   $$r = \frac{p}{1 + \epsilon\, \cos\theta}$$
   (b) A line joining a planet and the Sun sweeps out equal areas during equal intervals of time. [Equivalently, angular momentum is conserved.]
   $$\frac{d}{dt}\left(\frac{1}{2}r^2\dot{\theta}\right) = 0 \qquad h^2 = G(m_1 + m_2)\,p$$
   (c) The square of the orbital period of a planet is directly proportional to the cube of the semi-major axis of its orbit.
   $$\frac{P^2_{\text{planet}}}{a^3_{\text{planet}}} = \text{constant} = \frac{4\pi^2}{Gm}$$
5. **Luminosity** $\quad L = 4\pi d^2 b$
6. **Magnitude of a star**
   $$\text{(apparent magnitude)} - \text{(absolute magnitude)} = 5\log_{10}\left(\frac{d}{10}\right)$$

7. **Orbital period**

one large mass $P = 2\pi\sqrt{\dfrac{a^3}{Gm}}$

two masses $P = 2\pi\sqrt{\dfrac{a^3}{G(m_1+m_2)}}$

For the solar system $P = \sqrt{a^3}$ years, when a is in astronomical units (AU).

8. **Relations**

$$a = \frac{p}{1-\epsilon^2} \qquad b = \frac{p}{\sqrt{1-\epsilon^2}} \qquad \mu = Gm$$

$$r_{\text{min}} = \frac{p}{1+\epsilon} \qquad r_{\text{max}} = \frac{p}{1-\epsilon}$$

$$M = \frac{2\pi t}{P} \qquad \tan\frac{\theta}{2} = \sqrt{\frac{1+\epsilon}{1-\epsilon}}\tan\left(\frac{E}{2}\right)$$

$$\epsilon = \frac{r_{\text{max}} - r_{\text{min}}}{r_{\text{max}} + r_{\text{min}}} \qquad \frac{1}{r_{\text{min}}} - \frac{1}{p} = \frac{1}{p} - \frac{1}{r_{\text{max}}}$$

9. **Rocket equation** $v_{\text{final}} = v_{\text{exhaust}} \log\left(\dfrac{m_{\text{initial}}}{m_{\text{final}}}\right)$

10. **Schwarzschild radius** $R_{\text{Sch}} = \dfrac{2Gm}{c^2}$

11. **Vis-viva equation** (elliptic orbits) $v^2 = G(m_1+m_2)\left(\dfrac{2}{r} - \dfrac{1}{a}\right)$

12. **Numerical parameters**

- AU 149,597,870.7 km or 92,955,807.27 miles
- G (gravitational constant) $6.67428 \times 10^{-11}\,\frac{\text{N m}^2}{\text{kg}^2}$
- Sun
  - mass 1.9889×10^{30} kg
  - mean radius 696,000 km
  - μ 132,712,440,018 km^3s^{-2}
- Earth
  - eccentricity 0.016710219
  - mass 5.9736×10^{24} kg
  - mean radius 6,367.454 km
  - orbital period 365.256366 days
  - semi-major axis 149,597,887.5 km
- Moon
  - eccentricity 0.0549
  - mass 7.3477×10^{22} kg
  - mean radius 1,737.10 km
  - orbital period 27.321582 days
  - semi-major axis 384,399 km

9.3 ATMOSPHERIC PHYSICS

Notation

- **a** acceleration
- D density (kg/m^3)
- DA density altitude (feet)
- R gas constant
- RH relative humidity
- T temperature (°K)
- T_C temperature (°C)
- E_s saturation pressure of water vapor (mb)
- P total air pressure ($P = P_d + P_v$)
- P_A station pressure (inches Hg)
- P_d pressure of dry air (partial pressure in Pascals)
- P_v pressure of water vapor (partial pressure in Pascals)
- R_d gas constant for dry air, 287.05 J/(kg °K)
- R_v gas constant for water vapor, 461.495 J/(kg °K)
- T_R temperature (°R)
- **v** velocity
- z altitude
- γ lapse rate
- Ω earth rotation rate

Equations

1. **Coriolis effect** $\mathbf{a} = -2\Omega \times \mathbf{v}$

2. **Density** (approximation)

$$D = \frac{P}{RT} = \left(\frac{P_d}{R_d T}\right) + \left(\frac{P_v}{R_v T}\right) = \left(\frac{P_d}{R_d T}\right)\left(1 - 0.378\frac{P_v}{P}\right)$$

3. **Density altitude** (approximation)

$$\text{DA} = 145442.16\left(1 - \left(\frac{17.326 P_A}{T_R}\right)^{0.235}\right)$$

4. **Lapse rate** $\gamma = -\dfrac{dT}{dz}$
   The dry adiabatic lapse rate (DALR) is 9.8°C/km

5. **Saturation pressure** (approximation)

$$E_s = 6.1078 \times 10^{7.5\, T_C/(237.3+T_C)}$$

6. **Vapor pressure**

   (a) $P_V = E_s$ at the dew point

   (b) $P_V = \text{RH}\, E_s$ otherwise

9.4 ATOMIC PHYSICS

Notation

- E energy
- e charge
- h Planck constant
- L angular momentum
- m mass
- n principal quantum number
- r radius
- v velocity
- λ wavelength
- ω frequency
- c speed of light ($\approx 3 \times 10^8$ m/sec)
- $\hbar$ reduced Planck constant ($\hbar = \frac{h}{2\pi} \approx 10^{-34}$ J·s)
- k Coloumb's constant $\approx 9 \times 10^9$ N m^2 C^{-2}
- R Rydberg constant for electron energy ($\frac{m(ke^2)^2}{2\hbar^2} \approx 13.6$ eV)
- R_λ Rydberg constant for reciprocal wavelength ($\frac{m(ke^2)^2}{4\pi c\hbar^3} \approx 10^7$ m^{-1})

1. **Bohr model**
   (a) angular momentum $L_n = n\hbar$
   (b) radiation/absorption of energy $\hbar\omega = E_n - E_m$
2. **Hydrogen atom**
   (a) energy $E_n = \dfrac{mv_n^2}{2} - \dfrac{ke^2}{r_n} = -\dfrac{ke^2}{2r_n} = -\dfrac{R}{n^2}$
   (b) radiation spectrum $\dfrac{1}{\lambda} = R_\lambda \left(\dfrac{1}{n_{\text{initial}}^2} - \dfrac{1}{n_{\text{final}}^2} \right)$
   for transitions between states. This is the Lyman series ($n_{\text{final}} = 1$), the Balmer series ($n_{\text{final}} = 2$), and the Paschen series ($n_{\text{final}} = 3$).
   (c) radius $r_n = \dfrac{n^2\hbar^2}{mke^2}$
   (d) Bohr radius $a_0 = r_1 = \dfrac{\hbar^2}{mke^2} = 0.53$ Å
   (e) velocity $v_n = \dfrac{ke^2}{n\hbar}$

9.4.1 RADIOACTIVITY

- A radioactive sample activity
- N number of intact nuclei
- t time
- $T_{1/2}$ half-life
- λ decay constant
- τ mean lifetime

1. Governing equation $\dfrac{dN}{N} = -\lambda\, dt$
2. Half-life $T_{1/2} = \dfrac{\ln 2}{\lambda} = \dfrac{0.693}{\lambda}$
3. Radioactive decay law $N = N_0 e^{-\lambda t}$
4. Mean lifetime $\tau = \displaystyle\int_0^\infty t\lambda e^{-\lambda t}\, dt = \frac{1}{\lambda} = \frac{T_{1/2}}{\ln 2}$
5. Radioactive sample activity $A = \left| \dfrac{dN}{dt} \right| = \lambda N = A_0 e^{-\lambda t}$

9.5 BASIC MECHANICS

Notation

- $a, \mathbf{a}$ acceleration
- A area
- d distance
- $F, \mathbf{F}$ force
- I moment of inertia
- k spring constant
- $\mathbf{L}$ angular momentum
- m mass
- $\mathbf{M}$ moment
- $p, \mathbf{p}$ momentum
- P pressure
- $r, \mathbf{r}$ distance
- s speed
- T period
- $v, \mathbf{v}$ velocity
- V volume
- V potential
- W work
- $\mathbf{x}$ position
- ω angular velocity
- ρ density
- τ torque

- c speed of light ($\approx 3 \times 10^8$ m/sec)
- g acceleration due to gravity ($\approx 9.8 \frac{\text{m}}{\text{s}^2}$)
- G gravitational constant ($\approx 6.67 \times 10^{-11} \frac{\text{N m}^2}{\text{kg}^2}$)
- α angular acceleration
- μ_k coefficient of kinetic friction
- μ_s coefficient of static friction
- θ angular displacement

Equations

1. **Angular momemtum** $\mathbf{L} = \mathbf{r} \times \mathbf{p}$ and $\frac{d\mathbf{L}}{dt} = \mathbf{M} = \mathbf{r} \times \mathbf{F}$
2. **Center of mass** $\mathbf{x}_{cm} = \dfrac{\sum_i m_i \mathbf{x}_i}{M}$ with $M = \sum_i m_i$
3. **Conservative force**
   (a) $F = \nabla V$
   (b) (kinetic energy) + (potential energy) = (constant)
4. **Density** $\rho = \dfrac{m}{V}$
5. **Friction**
   (a) static friction $F_{\text{friction}} \leq \mu_s F_{\text{normal}}$
   (b) kinetic friction $F_{\text{friction}} = \mu_k F_{\text{normal}}$
6. **Gravitational force** $F = G\dfrac{m_1 m_2}{r^2}$
7. **Hooke's law** (Δx = displacement from equilibrium) $F = -k\,\Delta x$
8. **Inclined plane** (θ is angle between inclined plane and horizontal)
   (a) $F_{\text{normal}} = mg\cos\theta$
   (b) $F_{\text{incline}} = mg\sin\theta$
9. **Kinetic energy**

$$\text{linear} = \frac{1}{2}mv^2$$

$$\text{rotational} = \frac{1}{2}I\omega^2$$

$$\text{rotational, without slipping} = \frac{1}{2}mv^2 + \frac{1}{2}I\omega^2$$

10. **Moment** $\mathbf{M} = \mathbf{r} \times \mathbf{F}$
11. **Momentum** $p = mv$ and $\mathbf{p} = m\mathbf{v}$
12. **Newton's laws**
    (a) first law: in equilibrium, $\sum \mathbf{F} = 0$ and $\tau_{\text{clockwise}} = \tau_{\text{counter-clockwise}}$
    (b) second law: $\mathbf{F} = m\mathbf{a}$
13. **Pendulum** (L = length of pendulum) $T = 2\pi\sqrt{\frac{L}{g}}$
14. **Period of simple harmonic motion** $T = 2\pi\sqrt{\frac{m}{k}}$
15. **Pressure** $P = \frac{F}{A}$
16. **Speed of a wave on a string** $v = \sqrt{\frac{\text{(length of string) (tension in string)}}{\text{(mass of string)}}}$
17. **Springs** $F(t) = k(x(t) - x_0)$
18. **Torque** (l = lever arm length) $\tau = F\,l$
19. **Work** $W = Fd$
20. **Equations of circular motion**
    (a) Angular speed $v = r\omega$
    (b) Centripetal acceleration and force (constant speeds v, ω, radius R)
    $$a = \frac{v^2}{R} = \omega^2 R \qquad F = m\frac{v^2}{R} = m\omega^2 R$$
    (c) Circular motion (constant acceleration)
    $$\omega = \omega_0 + \alpha t$$
    $$\theta = \theta_0 + \omega_0 t + \frac{1}{2}\alpha t^2$$
    $$\theta = \theta_0 + \omega t - \frac{1}{2}\alpha t^2$$
    $$\theta = \theta_0 + \frac{1}{2}(\omega + \omega_0)\,t$$
    $$\omega^2 = \omega_0^2 + 2\alpha(\theta - \theta_0)$$
21. **Equations of linear motion**
    (a) Average speed $v = \frac{d}{t}$
    (b) Free fall (from rest and a height of h) $v = \sqrt{2gh}$
    (c) Linear motion (constant acceleration)
    $$v = v_0 + at$$
    $$x = x_0 + v_0 t + \frac{1}{2}at^2$$
    $$v^2 = v_0^2 + a(x - x_0)$$

9.6 BEAM DYNAMICS

- E Young's modulus
- I moment of inertia
- K length factor
- L length of column
- M bending moment
- Q shear force
- q applied loading
- w beam displacement
- κ curvature
- ϕ slope of the beam
- σ bending stress

1. **Static one-dimension beam equation** $\dfrac{d^2}{dx^2}\left(EI\dfrac{d^2w}{dx^2}\right) = q(x)$
2. **Derivable quantities**

   Bending moment $M = EI\dfrac{d^2w}{dx^2}$

   Shear force $Q = -\dfrac{dM}{dx} = -\dfrac{d}{dx}\left(EI\dfrac{d^2w}{dx^2}\right)$

   Slope $\phi = \dfrac{dw}{dx}$

   Strain $\epsilon_x = -z\kappa = -z\dfrac{d^2w}{dx^2}$

   Bending stress (with symmetry) $\sigma = \dfrac{Mz}{I} = -zE\dfrac{d^2w}{dx^2}$
3. **Boundary conditions** The end $x = z$ of the beam ...
   - is clamped $w(z) =$ given and $w'(z) =$ given
   - is free $w''(z) = 0$ and $w'''(z) = 0$
   - is simply supported $w''(z) = 0$ and $w(z) =$ given
   - has a point force $w''(z) = 0$ and $w'''(z) =$ given
   - has a point torque $w'''(z) = 0$ and $w''(z) =$ given
4. **Boundary conditions** The internal point $x = z$ of the beam ...
   - is clamped $w(z-) = w(z+) =$ given, $w'(z-) = w'(z+) =$ given
   - is simply supported
     $w(z-) = w(z+) =$ given, $w'(z-) = w'(z+)$, $w''(z-) = w''(z+)$
   - has a point force $w(z-) = w(z+)$, $w'(z-) = w'(z+)$,
     $w''(z-) = w''(z+)$, $w'''(z-) = w'''(z+) =$ given
   - has a point torque $w(z-) = w(z+)$, $w'(z-) = w'(z+)$,
     $w''(z-) = w''(z+) =$ given, $w'''(z-) = w'''(z+)$
5. **Buckling of columns** critical buckling force $= \dfrac{\pi^2 EI}{(KL)^2}$
   - $K = 1$ if both ends pinned
   - $K = 0.5$ if both ends fixed
   - $K = 0.699$ if one end fixed, one end pinned
   - $K = 2$ if one end fixed, one end moves laterally

9.7 BIOLOGICAL MODELS

1. **Population growth – discrete** Let N_m be the number of individuals in generation m and q be the number of offspring per parent per generation. Then

$$N_m = N_0 q^m \tag{9.7.1}$$

2. **Population growth – continuous** Let $N(t)$ be the population at time t, r be the growth rate, and K be the maximum population size.

$$\begin{aligned} \frac{dN}{dt} &= rN && \text{exponential growth model} \\ \frac{dN}{dt} &= r\frac{N(K-N)}{K} && \text{logistic growth model} \end{aligned} \tag{9.7.2}$$

3. **Population growth – Renewal equations** Let $n(t,x)$ be the number of individuals at time t of age x and $B(x)$ be the birth rate from individuals of age x. The equations modeling population size are:

$$\begin{aligned} \frac{\partial n}{\partial t} + \frac{\partial n}{\partial x} &= 0 && \text{for } t \geq 0, x \geq 0 \\ n(t, x=0) &= \int_0^\infty B(y)n(t,y)\,dy \end{aligned} \tag{9.7.3}$$

4. **SIR Epidemic model**
   - $S(t)$ number of susceptible individuals at time t
   - $I(t)$ number of infected individuals at time t
   - $R(t)$ number of recovered (and immune) individuals at time t
   - $N = I + R + S$ is the total population
   - β is the contact rate
   - γ is the reciprocal of period that an individual is infectious

   The equations modeling population health are:

$$\begin{aligned} \frac{dS}{dt} &= -\beta\frac{IS}{N} \\ \frac{dI}{dt} &= \beta\frac{IS}{N} - \gamma I \\ \frac{dS}{dt} &= \gamma I \end{aligned} \tag{9.7.4}$$

   The expected number of new infections, from a single infected individual in a population of all susceptibles, is $R_0 = \beta/\gamma$ ("basic reproduction number"). If

   - $R_0 > N/S(0)$ then $I'(0) > 0$ and there will be an epidemic.
   - $R_0 < N/S(0)$ then $I'(0) < 0$ and there will not be an epidemic.

   There are variations of the SIR model that include, for example: birth and death rates, no immunity (e.g., common cold), and having incubation periods.

9.8 CHEMISTRY

See atomic physics (page 695), temperature (page 795), and thermodynamics (page 724).

- i van 't Hoff factor
- K_a equilibrium – weak acid
- K_b equilibrium – weak base
- K_p equilibrium – gas pressure
- K_c equilibrium – molar concentration
- K_b ebullioscopic constant (for water: 0.512 ^{o}K-kg/mol)
- K_F cryoscopic constant (for water: 1.853 ^{o}K-kg/mol)
- R ideal gas constant (8.31 J/mol-^{o}K)
- n amount of substance (in moles)
- P pressure
- T temperature
- V volume
- [ ] molar concentration

1. **Equilibrium**

$$K_c = \frac{[\text{products}]}{[\text{reactants}]} = \frac{[\text{C}]^c[\text{D}]^d}{[\text{A}]^a[\text{B}]^b} \qquad \text{where } a\text{A} + b\text{B} \rightleftarrows c\text{C} + d\text{D}$$

$$K_p = \frac{P_{\text{products}}}{P_{\text{reactants}}} = \frac{(P_C)^c\,(P_D)^d}{(P_A)^a\,(P_B)^b}$$

2. **Equilibrium – Acids and Bases**

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}} \qquad \text{pH} = -\log[\text{H}^+]$$

$$\text{pOH} = -\log[\text{OH}^-]$$

$$K_b = \frac{[\text{OH}^-][\text{HB}^+]}{[\text{B}]} \qquad \text{pH} + \text{pOH} = 14$$

3. **Gases**
   (a) Ideal gas law $PV = nRT$
   (b) Boyle's law $P \propto \frac{1}{V}$
   (c) Charles' law $V \propto T$
   (d) Avogadro's law $V \propto n$
   (e) Dalton's law $P_{\text{total}} = p_1 + \cdots + p_n$
   where $p_X = P_{\text{total}} \frac{\text{moles X}}{\text{total moles}}$ is the partial presure

4. **Solutions**

$$\text{Molarity}(M) = \frac{\text{moles of solute}}{\text{liters of solution}} \qquad \text{molality}(m) = \frac{\text{moles of solute}}{\text{kg of solvent}}$$

5. **Other**
   (a) Freezing point depression $\Delta T_F = K_F\, m\, i$
   (b) Boiling point elevation $\Delta T_b = K_b\, m\, i$
   (c) Kinetic energy per mole $\frac{3}{2}RT$
   (d) Rate equation – zero order $\frac{d[\text{A}]}{dt} = -k$ $\quad [\text{A}] = [\text{A}]_0 - kt$
   (e) Rate equation – first order $\frac{d[\text{A}]}{dt} = -k[\text{A}]$ $\quad \ln[\text{A}] = \ln[\text{A}]_0 - kt$

9.9 CLASSICAL MECHANICS

- H Hamiltonian
- J action variable
- K kinetic energy
- L Lagrangian
- $\mathbf{p}$ generalized momentum
- $\mathbf{q}$ generalized coordinate
- T temperature
- U potential energy
- θ_i action angle
- ρ probability density
- $\{,\}$ Poisson bracket

Equations

1. **Action angle evolution** (for periodic motion) $\dfrac{d\theta_i}{dt} = \dfrac{\partial H}{\partial J_i}$
2. **Action variable** (for periodic motion)
$$J_i = \frac{1}{2\pi}\oint \mathbf{p}_i \cdot d\mathbf{q}_i = \frac{1}{2\pi}\int_0^{2\pi}\left(\mathbf{p}_i \cdot \frac{\partial \mathbf{q}_i}{\partial \theta_i}\right) d\theta_i$$
3. **Generalized momentum** $p_i = \dfrac{\partial L}{\partial \dot{q}_i}$
4. **Gibb's distribution** $\rho(\mathbf{p}, \mathbf{q}) = \rho_0 e^{-H(\mathbf{p},\mathbf{q})/T}$
5. **Hamiltonian** $H(\mathbf{p}, \mathbf{q}) = K + U = -L + \sum_i p_i \dot{q}_i$
6. **Hamilton's equations of motion** $\dot{q}_i = \dfrac{\partial H}{\partial p_i}, \qquad \dot{p}_i = -\dfrac{\partial H}{\partial q_i}$
7. **Lagrangian** $L(\mathbf{q}, \dot{\mathbf{q}}) = K - U$
8. **Lagrange's equations** $\dfrac{d}{dt}\left(\dfrac{\partial L}{\partial \dot{q}_i}\right) - \dfrac{\partial L}{\partial q_i} = 0$
9. **Time evolution** (of any function $f(\mathbf{p}, \mathbf{q}, t)$)
$$\begin{aligned}\frac{df}{dt} &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial q_i}\dot{q}_i + \frac{\partial f}{\partial p_i}\dot{p}_i \\ &= \frac{\partial f}{\partial t} + \frac{\partial f}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial H}{\partial q_i} = \frac{\partial f}{\partial t} + \{f, H\}\end{aligned}$$
10. **Liouville's theorem**: For a probability distribution $\rho = \rho(\mathbf{p}, \mathbf{q})$
$$\frac{d\rho}{dt} = 0 = \frac{\partial \rho}{\partial t} + \{\rho, H\}$$

EXAMPLES

1. Consider a falling mass m. The kinetic energy is $K = \frac{1}{2}mv^2 = \frac{1}{2}m\dot{x}^2$. The potential energy is $U = -mgx$. The Lagrangian is $L = K - U = \frac{1}{2}m\dot{x}^2 - mgx$.
If $q = x$ then $\dot{q} = \dot{x}$ and $p = \frac{\partial L}{\partial \dot{q}} = \frac{\partial L}{\partial \dot{x}} = m\dot{x}$. Lagrangian's equations become $\dfrac{d}{dt}\left(\dfrac{\partial L}{\partial \dot{x}}\right) - \dfrac{\partial L}{\partial x} = \dfrac{d}{dt}(m\dot{x}) - mg = m\ddot{x} - mg = 0$ or $\ddot{x} = g$, as expected.
2. If two equal point masses (m) are connected by a spring (with spring constant k) of length ℓ, then $L = \frac{m}{2}\left(\dot{x}_1^2 + \dot{x}_2^2\right) - \frac{k}{2}\left(x_1 - x_2 - \ell\right)^2$.
3. If a pendulum (massless rigid stick of length ℓ) with a point mass (m) at the end is swinging due to gravity (g), and θ is the angle between the stick and the vertical, then $L = \frac{m}{2}\ell^2\dot{\theta}^2 + mg\ell\cos\theta$.

9.10 COORDINATE SYSTEMS – ASTRONOMICAL

Notation

- a altitude
- A azimuth
- b galactic latitude
- H hour angle ($H = \text{LST} - \alpha$)
- l galactic longitude
- LST local sidereal time
- α right ascension
- β ecliptic latitude
- δ declination
- ϵ Earth's axial tilt ($\epsilon \approx 23.4^\circ$)
- λ ecliptic longitude
- ϕ geographic latitude
- θ zenith ($\theta = 90^\circ - a$)

Relations

1. **Ecliptic coordinates to equatorial coordinates** (α, δ)

$$\begin{aligned}\sin\delta &= \sin\epsilon\sin\lambda\cos\beta + \cos\epsilon\sin\beta\\ \cos\alpha\cos\delta &= \cos\lambda\cos\beta\\ \sin\alpha\cos\delta &= \cos\epsilon\sin\lambda\cos\beta - \sin\epsilon\sin\beta\end{aligned}$$

2. **Equatorial coordinates to ecliptic coordinates** (β, λ)

$$\begin{aligned}\sin\beta &= \cos\epsilon\sin\delta - \sin\alpha\cos\delta\sin\epsilon\\ \cos\lambda\cos\beta &= \cos\alpha\cos\delta\\ \sin\lambda\cos\beta &= \sin\epsilon\sin\delta + \sin\alpha\cos\delta\cos\epsilon\end{aligned}$$

3. **Equatorial coordinates to galactic coordinates** (b, l)

$$\begin{aligned}b &= \sin^{-1}\left(\cos\delta\cos 27.4^\circ\cos(\alpha - 192.25^\circ) + \sin\delta\sin 27.4^\circ\right)\\ l &= \tan^{-1}\left(\sin\delta - \frac{\sin b\sin 27.4^\circ}{\cos\delta\sin(\alpha - 192.25^\circ)\cos 27.4^\circ}\right) + 33^\circ\end{aligned}$$

4. **Equatorial coordinates to horizontal coordinates** (a, A)

$$\begin{aligned}\sin a &= \sin\phi\sin\delta + \cos\phi\cos\delta\cos H = \cos\theta\\ \cos A &= \frac{\cos\phi\sin\delta - \sin\phi\cos\delta\cos H}{\cos a}\end{aligned}$$

5. **Galactic coordinates to equatorial coordinates** (α, δ)

$$\begin{aligned}\delta &= \sin^{-1}\left(\cos b\cos 27.4^\circ\sin(l - 33^\circ) + \sin b\sin 27.4^\circ\right)\\ \alpha &= \tan^{-1}\left(\frac{\cos b\cos(l - 33^\circ)}{\sin b\cos 27.4^\circ - \cos b\sin 27.4^\circ - \sin(l - 33^\circ)}\right) + 192.5^\circ\end{aligned}$$

6. **Horizontal coordinates to equatorial coordinates** (α, δ)

$$\begin{aligned}\sin\delta &= \quad\sin\phi\sin a - \cos\phi\cos a\cos A\\ \cos\delta\cos H &= \quad\cos\phi\sin a + \sin\phi\cos a\cos A\\ \cos\delta\sin H &= -\sin A\cos a\end{aligned}$$

9.11 COORDINATE SYSTEMS – TERRESTRIAL

Notation

- a semi-major axis
- b semi-minor axis
- e eccentricity
- e' second eccentricity
- f flattening
- h height
- N radius of curvature in the prime vertical
- α angular eccentricity
- λ longitude
- ϕ geodetic latitude
- ϕ' geocentric latitude

- ENU local (East, North, Up) coordinates
- ECEF Earth-centered Earth-fixed coordinates
- ECF same as ECEF
- $\{X_r, Y_r, Z_r\}$ reference point for ENU (in ECEF)

Relations

1. $b = a(1-f)$
2. $e^2 = \dfrac{a^2-b^2}{a^2} = f(2-f)$
3. $e'^2 = \dfrac{a^2-b^2}{b^2} = \dfrac{f(2-f)}{(1-f)^2}$
4. $f = 1 - \cos\alpha = \dfrac{a-b}{a}$
5. $N(\phi) = \dfrac{a}{\sqrt{1-\sin^2\phi\sin^2\alpha}}$
6. $\alpha = \cos^{-1}\left(\dfrac{b}{a}\right)$
7. $\chi = \sqrt{1-e^2\sin^2\phi}.$
8. $\tan\phi' = \dfrac{\frac{a}{\chi}(1-f)^2+h}{\frac{a}{\chi}+h}\tan\phi$

Transformations

1. **ECEF coordinates to ENU coordinates** (x, y, z)

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} -\sin\lambda & \cos\lambda & 0 \\ -\sin\phi\cos\lambda & -\sin\phi\sin\lambda & \cos\phi \\ \cos\phi\cos\lambda & \cos\phi\sin\lambda & \sin\phi \end{bmatrix} \begin{bmatrix} X - X_r \\ Y - Y_r \\ Z - Z_r \end{bmatrix}$$

2. **ENU coordinates to ECEF coordinates** (X, Y, Z)

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} -\sin\lambda & -\sin\phi\cos\lambda & \cos\phi\cos\lambda \\ \cos\lambda & -\sin\phi\sin\lambda & \cos\phi\sin\lambda \\ 0 & \cos\phi & \sin\phi \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} X_r \\ Y_r \\ Z_r \end{bmatrix}$$

3. **Geodetic coordinates to ECEF coordinates** (X, Y, Z)

$$X = \left(\frac{a}{\chi}+h\right)\cos\phi\cos\lambda \qquad Y = \left(\frac{a}{\chi}+h\right)\cos\phi\sin\lambda$$

$$Z = \left(\frac{a(1-e^2)}{\chi}+h\right)\sin\phi$$

4. **Geodetic coordinates to geocentric rectangular coordinates** (X, Y, Z)

$$X = (N+h)\cos\phi\cos\lambda \quad Y = (N+h)\cos\phi\sin\lambda \quad Z = \left(N\cos^2 + h\right)\sin\phi$$

9.12 EARTHQUAKE ENGINEERING

- A maximum excursion of a Wood–Anderson seismograph
- A_0 empirical function
- E_S radiated seismic energy
- M_0 magnitude of the seismic moment (in dyne-centimeters)
- M_{e} energy magnitude
- M_{L} local magnitude (Richter magnitude)
- M_{W} moment magnitude
- v velocity
- δ distance of seismograph to earthquake
- λ first Lamé parameter
- μ shear modulus
- ρ density

Equations

1. **Magnitude 0 event**: earthquake showing a maximum combined horizontal displacement of 1μm (0.00004 in) on a Wood–Anderson torsion seismometer at a distance of 100 km from the earthquake epicenter.
2. $E_S = 1.6 \times 10^{-5} M_0$
3. $M_{\text{e}} = \frac{2}{3} \log_{10} E_S - 2.9$
4. $M_{\text{L}} = \log_{10} A - \log_{10} A_0(\delta)$
5. $M_{\text{W}} = \frac{2}{3} \log_{10} M_0 - 10.7$
6. **Seismic moment** $=$ (Rigidity) $\times$ (Fault Area) $\times$ (Slip Length Area)
7. **Speed of seismic waves**

$$v_{\text{S-wave}} = \sqrt{\frac{\lambda + 2\mu}{\rho}} \qquad v_{\text{P-wave}} = \sqrt{\frac{\mu}{\rho}}$$

8. **Energy in earthquakes**

Richter Magnitude M_L	Amount of TNT	Joule equivalent
0	15 g	63 kJ
1	474 g	2 MJ
2	15 kg	63 MJ
3	474 kg	2 GJ
4	15 metric tons	63 GJ
5	474 metric tons	2 TJ
6	15 kilotons	63 TJ
7	474 kilotons	2 PJ
8	15 megatons	63 PJ
9	474 megatons	2 EJ
10	15 gigatons	63 EJ

9.13 ECONOMICS (MACRO)

- C = Consumption
- E = Number of employed
- G = Government Spending
- I = Investment
- L = Labor force
- M = Imports
- P = Price
- POP = Adult population
- Q = Quantity
- S = Savings
- T = Taxes
- U = Number of Unemployed
- X = Exports
- Y = Total Income
- f = Rate of Job Finding
- s = Rate of Job Separation

1. *Consumption*
   (a) *Marginal Propensity to Consume* $\text{MPC} = \dfrac{\text{Change in Consumption}}{\text{Change in Income}}$
   (b) *Marginal Propensity to Save* $\text{MPS} = 1 - \text{MPC} = \dfrac{\text{Change in Savings}}{\text{Change in Income}}$
   (c) *Price elasticity of demand* (PED)
   $$\text{PED} = \frac{\%\ \Delta\ \text{quantity demanded}}{\%\ \Delta\ \text{price}} = \frac{\text{dQ/Q}}{\text{dP/P}} = \frac{\text{P}}{\text{Q}}\frac{\text{dQ}}{\text{dP}}$$
   (d) *Spending multiplier* $\text{Spending multiplier} = 1/\text{MPS}$
   (e) *Tax multiplier* $\text{Tax multiplier} = \text{MPC}/\text{MPS}$

2. *Employment*
   (a) *Labor force* $\text{L} = \text{E} + \text{U}$
   (b) *Labor force participation* $\text{Labor force participation} = \text{L}/\text{POP}$
   (c) *Steady state labor market* $sE = fU$
   (d) *Unemployment rate* $\text{Unemployment rate} = \dfrac{\text{U}}{\text{L}} = \dfrac{\text{s}}{\text{s} + \text{f}}$

3. *Gross Domestic Product* (GDP)
   (a) *Nominal GDP* (NGDP)
      i. Income Approach $\text{NGDP} = \text{C} + \text{S} + \text{T}$
      ii. Spending Approach $\text{NGDP} = \text{C} + \text{I} + \text{G} + \underbrace{(\text{X} - \text{M})}_{\text{net exports}}$
   (b) *Real GDP* (RGDP) $\text{RGDP} = P_1Q_1 + P_2Q_2 + P_3Q_3 + \ldots$
   (c) *GDP deflator* $\text{GDP deflator} = \frac{\text{NGDP}}{\text{RGDP}} \times 100$

4. *Inflation*
   (a) *Consumer Price Index* (CPI) $\text{CPI} = \dfrac{\text{price in specific year}}{\text{price in base year}}$
   (b) *Inflation rate* $\text{Inflation rate} = \dfrac{\text{CPI}_{\text{new}} - \text{CPI}_{\text{old}}}{\text{CPI}_{\text{old}}} \times 100$

5. *Interest rate*

   (a) Correct $$(\text{Real interest rate}) = \frac{1 + (\text{Nominal interest rate})}{1 + (\text{Inflation rate})} - 1$$

   (b) Fisher equation (approximation)
   $$(\text{Real interest rate}) = (\text{Nominal interest rate}) - (\text{Inflation rate})$$

6. *Profit Maximizing point* Marginal Cost = Marginal Revenue

7. *Saving*

   (a) *National saving* National saving = Y − C − G
   (b) *Private saving* Private saving = Y − T − C
   (c) *Public saving* Public saving = T − G
   (d) *Open economy* Saving = I + X − M
   (e) *Closed economy (no trade)* Saving = I

8. *Time value of money* Future value = Present value × (1 + Interest rate)

9. *Wages* (Change in real wage) = (Change in nominal wage) − Inflation

9.13.1 DEPRECIATION

There are many different, yet standard, ways to determine depreciation. For example, for a $V = \$100$ asset with a lifetime of $N = 3$ or $N = 5$ years, the yearly depreciation (D), rounded to the nearest dollar, is:

Lifetime of 3 years

Year	SL	DD	SYD	MACRS
1	$33	$67	$50	$33
2	$33	$22	$33	$44
3	$33	$11	$17	$15
4				$7

Lifetime of 5 years

Year	SL	DD	SYD	MACRS
1	$20	$40	$33	$20
2	$20	$24	$24	$32
3	$20	$14	$20	$19
4	$20	$9	$13	$12
5	$20	$13	$7	$12
6				$6

where the depreciation methods are:

- *Straight Line Depreciation* (SL) $D = V\frac{1}{N}$ for years $y = 1, 2, \ldots, N$
- *Double Declining Depreciation* (DD); $D = V\frac{2}{N}\left(\frac{N-2}{N}\right)^{y-1}$ for years $y = 1, 2, \ldots, N-1$, with a remainder in year N
- *Sum of Year's Digits* (SYD); $D = V\frac{2(N+1-y)}{N(N+1)}$ for years $y = 1, 2, \ldots, N$
- *Modified Accelerated Cost Recovery System* (MACRS); depreciation is given in the US Internal Revenue Service's (IRS) publication 946.

9.14 ELECTROMAGNETIC TRANSMISSION

9.14.1 ANTENNAS

- A area of physical aperture
- BW antenna beamwidth in radians
- BW_{az} azimuth beamwidth
- BW_{el} elevation beamwidth
- d antenna diameter
- G antenna gain
- L length of rectangular aperture
- W width of rectangular aperture
- η efficiency
- λ wavelength

Equations

1. Parabolic antenna beamwidth (in degrees) $BW = \frac{70\,\lambda}{d}$
2. Antenna gain
   (a) idealized uniform antenna pattern
   $$G = \frac{\text{area of sphere}}{\text{area of antenna pattern}} = \frac{4\pi}{BW_{az}BW_{el}}$$
   (b) real antenna pattern $G = \dfrac{4\pi\eta A}{\lambda^2}$
   (c) rectangular X-band aperture (L and W are in cm) $G = 1.4\,L\,W$
   (d) circular X-band aperture (d is in cm) $G = d^2\eta$

9.14.2 WAVEGUIDES

- a dimension of waveguide ($a > b$)
- b dimension of waveguide
- c speed of light in medium
- r waveguide radius
- ϵ permittivity of medium
- μ permeability of medium
- f_c cutoff frequency (transmitted above and attenuated below this frequency)
- f_c^{mn} cutoff frequency for TE_{mn} mode
- TE transverse electric mode (E-field orthogonal to axis of waveguide)
- TE_{ij} TE mode with i and j wave oscillations in a and b directions $(i, j \geq 0)$
- TM transverse magnetic mode (H-field orthogonal to axis of waveguide)
- TM_{ij} TM mode with i and j wave oscillations in a and b directions $(i, j \geq 1)$

Equations

1. Circular waveguide (TM_{01} mode) $$f_c = \frac{2.4048}{r\sqrt{\mu\epsilon}} = \frac{2.4048\,c}{r}$$
2. Rectangular waveguide
   $$f_c = \frac{c\,\pi}{2a} \qquad f_c^{mn} = \frac{c\,\pi}{2}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2} = \frac{\pi}{2\sqrt{\mu\epsilon}}\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}$$

EXAMPLE

Rectangular waveguide ($a = 2$ cm, $b = 1$ cm) filled with deionized water ($\mu = 1$, $\epsilon = 81$) operating at 3 GHz. Propagating modes and cutoff frequencies are:

Mode	f_c^{mn}
TE_{10}	0.833 GHz
TE_{01}, TE_{20}	1.667 GHz
TE_{11}, TM_{11}	1.863 GHz
TE_{21}, TM_{21}	2.357 GHz
TE_{30}	2.5 GHz

9.15 ELECTROSTATICS AND MAGNETISM

- $B, \mathbf{B}$ magnetic field
- c speed of light
- $\mathbf{D} = \epsilon\mathbf{E}$ displacement vector
- $E, \mathbf{E}$ electric field
- $F, \mathbf{F}$ force
- $\mathbf{H} = \mathbf{B}/\mu_0 - \mathbf{M}$ magnetic field
- I current
- $\mathbf{J}$ current density
- $k = \frac{1}{4\pi\epsilon_0} \approx 9 \times 10^9 \frac{\mathrm{N\,m^2}}{\mathrm{C^2}}$
- q charge
- r separation
- $\mathbf{S}$ Poynting vector
- U potential energy
- ϵ dielectric constant
- ρ charge density
- $v, \mathbf{v}$ velocity
- $\mu_0 = 4\pi \times 10^{-7} \frac{\mathrm{T\,m^2}}{\mathrm{A}}$

Equations

1. **Ampere's law** $\nabla \times \mathbf{B} = \mu_o \mathbf{J}$
2. **Biot–Savart law**
   (a) for a segment of wire $d\mathbf{s}$ $$d\mathbf{B} = \frac{\mu_0 I}{4\pi}\frac{d\mathbf{s} \times \mathbf{r}}{r^3}$$
   (b) at distance r from an infinitely long wire $B = \mu_0 I/2\pi r$
   (c) at the center of a loop of radius R $B = \mu_0 I/2R$
3. **Coulomb law** $$F = k\frac{q_1 q_2}{r^2}$$
4. **Densities**
   (a) electric field energy density $(\mathbf{E} \cdot \mathbf{D})/2$
   (b) magnetic field energy density $(\mathbf{B} \cdot \mathbf{H})/2$
   (c) momentum density $(\mathbf{E} \times \mathbf{B})/c^2$
5. **Electric field due to a point charge** $$E = k\frac{q}{r^2}$$
6. **Electric potential energy** $$U = k\frac{q_1 q_2}{r}$$
7. **Force**
   (a) in a constant electric field $\mathbf{F} = q\mathbf{E}$
   (b) on a charge moving in a magnetic field $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$
   (c) Lorentz force $\mathbf{F} = q(\mathbf{e} + \mathbf{v} \times \mathbf{B})$
8. **Gauss' law** $\nabla \cdot \mathbf{E} = \rho/\epsilon_0$
9. **Larmor formula** (energy loss of an accelerated charge) $$\frac{dE}{dt} = \frac{q^2 \dot{v}^2}{6\pi c^3 \epsilon_0}$$
10. **Maxwell's equations** (differential form)
    (a) $\nabla \cdot \mathbf{B} = 0$
    (b) $\nabla \times \mathbf{E} = -\dfrac{\partial \mathbf{B}}{\partial t}$
    (c) $\nabla \cdot \mathbf{D} = \rho_{\text{free}}$
    (d) $\nabla \times \mathbf{H} = \mathbf{J}_{\text{free}} + \dfrac{\partial \mathbf{D}}{\partial t}$
11. **Poynting vector** $\mathbf{S} = \mathbf{E} \times \mathbf{H}$

9.16 ELECTROMAGNETIC FIELD EQUATIONS

Notation

- $\mathbf{A}$ (vector) magnetic potential
- $B, \mathbf{B}$ magnetic field
- c speed of light
- $E, \mathbf{E}$ electric field
- F stress-energy tensor
- $\mathbf{J}$ (vector) current density
- $\mu_0 = 4\pi \times 10^{-7} \frac{\mathrm{T\,m}^2}{\mathrm{A}}$
- ϕ (scalar) electric potential
- ρ (scalar) charge density

We have $\mathbf{B} = \nabla \times \mathbf{A}$ and $\mathbf{E} = -\nabla\phi - \frac{1}{c}\frac{\partial \mathbf{A}}{\partial t}$. The electromagnetic four-potential is the one-form $A^\nu = \begin{bmatrix} \frac{\phi}{c} & \mathbf{A} \end{bmatrix}$ and the electromagnetic stress-energy tensor F

1. is a covariant antisymmetric tensor
2. is the exterior derivative $F = dA$
3. is, in component form, $F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu$ or $F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu$

where $\partial^\nu = \frac{\partial}{\partial x_\nu} = \begin{bmatrix} \frac{1}{c}\frac{\partial}{\partial t} & -\nabla \end{bmatrix}$. The contravariant form ($F^{\mu\nu}$) and the covariant form ($F_{\mu\nu} = \eta_{\mu\alpha} F^{\alpha\beta} \eta_{\beta\nu}$) for the metric tensor $\eta_{\mu,\alpha} = \operatorname{diag}(1, -1, -1, -1)$ are

$$F^{\mu\nu} = \begin{bmatrix} 0 & -\frac{E_x}{c} & -\frac{E_y}{c} & -\frac{E_z}{c} \\ \frac{E_x}{c} & 0 & -B_z & B_y \\ \frac{E_y}{c} & B_z & 0 & -B_x \\ \frac{E_z}{c} & -B_y & B_x & 0 \end{bmatrix} \qquad F_{\mu\nu} = \begin{bmatrix} 0 & \frac{E_x}{c} & \frac{E_y}{c} & \frac{E_z}{c} \\ -\frac{E_x}{c} & 0 & -B_z & B_y \\ -\frac{E_y}{c} & B_z & 0 & -B_x \\ -\frac{E_z}{c} & -B_y & B_x & 0 \end{bmatrix} \tag{9.16.1}$$

The electromagnetic four-current is $J^\nu = \begin{bmatrix} c\rho & \mathbf{J} \end{bmatrix}$ (note that $\partial_\alpha J^\alpha = 0$) and Maxwell's equations are

$$\begin{aligned} \partial_\alpha F^{\alpha\nu} &= \mu_0 J^\nu \\ \partial^\lambda F^{\mu\nu} + \partial^\mu F^{\nu\lambda} + \partial^\nu F^{\lambda\mu} &= 0 \end{aligned} \tag{9.16.2}$$

Notes

1. The meaning of "gauge invariance" is this: if $\mathbf{A}$ is changed by the gradient of a scalar field Λ (that is, $\mathbf{A} \to \mathbf{A} + \nabla\Lambda$) then Equation (9.16.2) is unchanged.
2. The Lorenz gauge condition is $\partial_\alpha A^\alpha = 0$.
3. The electromagnetic force on a particle with charge q is:

   $$\text{Force} = q\left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) = qF^{\mu\nu}u^\mu$$

   where the four vector velocity is $u^\mu = \begin{bmatrix} c & u_x & u_y & u_z \end{bmatrix}$.
4. Transforming the field strength tensor to a different inertial frame is accomplished by: $\mathbf{F}' = F'^{\mu\nu} = \mathbf{L}\mathbf{F}\mathbf{L}^{\mathrm{T}} = L^{\mu\gamma} L^{\nu\delta} F^{\gamma\delta}$
   where $\mathbf{L}$ is the Lorentz transformation tensor:

   $$\mathbf{L} = \begin{bmatrix} \gamma & -\gamma\beta_x & -\gamma\beta_y & -\gamma\beta_z \\ -\gamma\beta_x & 1 + \frac{(\gamma-1)\beta_x^2}{\beta^2} & \frac{(\gamma-1)\beta_x\beta_y}{\beta^2} & \frac{(\gamma-1)\beta_x\beta_z}{\beta^2} \\ -\gamma\beta_y & \frac{(\gamma-1)\beta_x\beta_y}{\beta^2} & 1 + \frac{(\gamma-1)\beta_y^2}{\beta^2} & \frac{(\gamma-1)\beta_y\beta_z}{\beta^2} \\ -\gamma\beta_z & \frac{(\gamma-1)\beta_x\beta_z}{\beta^2} & \frac{(\gamma-1)\beta_y\beta_z}{\beta^2} & 1 + \frac{(\gamma-1)\beta_z^2}{\beta^2} \end{bmatrix}$$

   and $\beta = \mathbf{v}/c$, $\beta^2 = |\beta|^2$, and $\gamma = 1/\sqrt{1 - \beta \cdot \beta}$

9.17 ELECTRONIC CIRCUITS

- A area
- C capacitance
- d distance
- I current
- L inductance
- P power
- Q charge
- R resistance
- V voltage
- Z impedance
- ω frequency
- ρ resistivity
- $\epsilon_0 \approx 8.85 \times 10^{-12} \frac{\text{F}}{\text{m}}$
- κ dielectric constant

Equations

1. **Ohm's law** $V = IR$
2. **Electric power** $P = IV = I^2R = \dfrac{V^2}{R}$, $P_{\text{average}} = I_{\text{rms}}^2 R = \dfrac{V_{\text{rms}}^2}{R}$
3. **Kirchhoff's laws**
   (a) Loop rule $\sum_{\text{around any loop}} \Delta V_i = 0$
   (b) Node rule $\sum_{\text{at any node}} I_i = 0$
4. **RMS values** (for AC circuits) $V_{\text{rms}} = \dfrac{V_{\text{peak}}}{\sqrt{2}}$, $I_{\text{rms}} = \dfrac{I_{\text{peak}}}{\sqrt{2}}$
5. **Resistors**
   (a) Adding in series $R_S = R_1 + R_2 + R_3 + \dots$
   (b) Adding in parallel $\dfrac{1}{R_P} = \dfrac{1}{R_1} + \dfrac{1}{R_2} + \dfrac{1}{R_3} + \dots$
   (c) Resistance of a wire of length L and cross-sectional area A $R = \dfrac{\rho L}{A}$
6. **Capacitors**
   (a) Adding in series $\dfrac{1}{C_S} = \dfrac{1}{C_1} + \dfrac{1}{C_2} + \dfrac{1}{C_3} + \dots$
   (b) Adding in parallel $C_P = C_1 + C_2 + C_3 + \dots$
   (c) Capacitance $C = \dfrac{Q}{V}$
   (d) Capacitance of a parallel plate capacitor $C = \dfrac{\kappa \epsilon_0 A}{d}$
   (e) Current across a capacitor $I(t) = C\dfrac{dV(t)}{dt}$
   (f) Energy stored in a capacitor $= \dfrac{1}{2}CV^2 = \dfrac{1}{2}QV = \dfrac{1}{2}\dfrac{Q^2}{C}$
7. **Inductor** (inductors add with the same formulas as that for resistors)
   (a) Voltage across an inductor $V(t) = L\dfrac{dI(t)}{dt}$
   (b) Energy stored in an inductor $= \dfrac{1}{2}LI^2$
8. **Impedance** (impedances add with the same formulas as that for resistors)
   (a) Ideal resistor $Z_R = R$
   (b) Ideal inductor $Z_L = i\omega L$
   (c) Ideal capacitor $Z_C = \dfrac{1}{i\omega C}$

9.18 EPIDEMIOLOGY

Notation

- FN False negative
- FP False positive
- LR likelihood ratio
- PV predictive value
- TN True negative
- TP True positive

Equations

1. **Operating characteristics**

		Disease: Yes	Disease: No
Meets case definition	Yes	True positive (TP)	False positive (FP)
	No	False negative (FN)	True negative (TN)

(a) Sensitivity $= \text{Prob}\,(T+ \mid D+) = \dfrac{TP}{TP + FN}$

(b) Specificity $= \text{Prob}\,(T- \mid D-) = \dfrac{TN}{TN + FP}$

(c) False positive rate $= \text{Prob}\,(T+ \mid D-) = \dfrac{FP}{TN + FP}$

(d) False negative rate $= \text{Prob}\,(T- \mid D+) = \dfrac{FN}{TP + FN}$

(e) Positive predictive value

$$PV+ = \frac{(\text{Sensitivity})(\text{Prior Probability})}{(\text{Sensitivity})(\text{Prior Probability}) + (1 - \text{Sensitivity}(1 - \text{Prior Probability}))}$$

(f) Likelihood ratios

$$LR+ = \frac{\text{Sensitivity}}{(1 - \text{Specificity})} \qquad LR- = \frac{(1 - \text{Sensitivity})}{\text{Specificity}}$$

2. **Cohort analysis for binomial data**

		Disease: Yes	Disease: No
Exposure	Yes	a	b
	No	c	d

$$\text{risk ratio} = RR = \frac{a/(a+b)}{c/(c+d)}$$

$$\text{odds ratio} = OR = \frac{a/b}{c/d}$$

9.19 FLUID MECHANICS

- A area
- g gravity
- h height
- M molar mass
- n number of moles of gas
- N number of molecules
- P pressure
- Q volume flow rate
- R gas law constant $\approx 8.31 \frac{\text{J}}{\text{K mole}}$
- k_B Boltzmann's constant $\approx 1.38 \times 10^{-23} \frac{\text{J}}{\text{K}}$
- T temperature
- v velocity
- V volume
- μ fluid viscosity
- μ_m mass of molecule
- ν kinematic viscosity
- ω vorticity
- ρ density

Equations

1. **Average molecular kinetic energy** $K_{\text{avg}} = \frac{3}{2} k_B T$
2. **Basic relations** $\omega = \nabla \times \mathbf{v}, \qquad \nu = \frac{\mu}{\rho}$
3. **Bernoulli's equation** $P + \rho g h + \frac{1}{2}\rho v^2 = \text{constant}$
4. **Boyle's law** $PV = \text{constant}$
5. **Continuity equation** $\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$
6. **Euler equation** (incompressibility constraint and the following)
$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P$$
7. **Hydrostatic pressure** $P = P_0 + \rho g h$
8. **Ideal gas law** $PV = nRT = Nk_BT$
9. **Incompressibility constraint** (constant density) $\nabla \cdot (\mathbf{v}) = 0$
10. **Navier–Stokes equations**
$$\rho \left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} \right) = -\nabla P + \mu \nabla^2 \mathbf{v} + (\text{body force})$$
Sometimes the "body force" is $-\rho g h$ due to hydrostatics.
11. **Root mean square velocity** $v_{\text{rms}} = \sqrt{\frac{3RT}{M}} = \sqrt{\frac{3k_BT}{\mu_m}}$
12. **Reynolds number** $\text{Re} = \frac{\text{inertial forces}}{\text{viscous forces}} = \frac{\rho v^2/L}{\mu v/L^2} = \frac{\rho v L}{\mu}$
where L is a characteristic length scale.
13. **Volume flow rate** $Q = Av$

9.20 HUMAN BODY

Notation

- a age (years)
- BMI body mass index
- BSA body surface area (m^2)
- BMR basal metabolic rate ($\frac{\text{kcal}}{\text{day}}$)
- LBM lean body mass (kg)
- MET metabolic equivalent of task
- w weight (w_{kg} for kilograms, w_{lb} for pounds)
- h height (h_{cm} for centimeters, h_{m} for meters, h_{in} for inches)

Equations

1. **Basal metabolic rate**
   - Using LBM $\quad \text{BMR} = 500 + 22\,\text{LBM}$
   - Mifflin–St. Jeor equations
     - Men
       $$\begin{aligned}\text{BMR} &= 10\,w_{\text{kg}} + 6.25\,h_{\text{cm}} - 5a + 5\\ &= 22\,w_{\text{lb}} + 2.46\,h_{\text{in}} - 5a + 5\end{aligned}$$
     - Women
       $$\begin{aligned}\text{BMR} &= 10\,w_{\text{kg}} + 6.25\,h_{\text{cm}} - 5a - 161\\ &= 22\,w_{\text{lb}} + 2.46\,h_{\text{in}} - 5a - 161\end{aligned}$$
2. **Body mass index**
   $$\begin{aligned}\text{BMI} &= \frac{w_{\text{kg}}}{h_{\text{m}}^2} && \text{(metric units)}\\ &= 703\frac{w_{\text{lb}}}{h_{\text{in}}^2} && \text{(English units)}\end{aligned}$$
   An "optimal" BMI value may be between 18.5 to 25.
3. **Body surface area**
   - Dubois & Dubois $\quad \text{BSA} = 0.20247\,h_{\text{m}}^{0.725}\,w_{\text{kg}}^{0.425}$
   - Gehan & George $\quad \text{BSA} = 0.235\,h_{\text{cm}}^{0.42246}\,w_{\text{kg}}^{0.51456}$
   - Mosteller $\quad \text{BSA} = \dfrac{\sqrt{h_{\text{cm}}\,w_{\text{kg}}}}{60}$
4. **Caloric needs** $=$ BMR $\times$ (lifestyle factor) (Harris Benedict formula)
   - (lifestyle factor) = 1.2 for *sedentary* lifestyle
   - (lifestyle factor) = 1.375 for *lightly active* lifestyle
   - (lifestyle factor) = 1.55 for *moderately active* lifestyle
   - (lifestyle factor) = 1.725 for *very active* lifestyle
   - (lifestyle factor) = 1.9 for *extra active* lifestyle
5. **Calories burned** $=$ (duration in minutes) $\times$ MET $\times \frac{3.5}{200} \times w_{\text{kg}}$

   Sample MET values:
   - sleeping 0.9
   - watching television 1.0
   - walking at 1.7 mph 2.3
   - walking at 3.4 mph 3.6
   - bicycling at 10 mph 4.0
   - jogging 7.0
   - calisthenics 8.0
   - rope jumping 10

9.21 MODELING PHYSICAL SYSTEMS

For many physical systems there are "across variables" A (whose value is the difference between two measurements) and "through variables" T. For these systems:

1. There is a through-type energy storage device; the energy stored scales as the square of the T variable.
2. There is an across-type energy storage device; the energy stored scales as the square of the A variable.
3. When using a graph (whose edges represent components) to model these systems there are conservation laws analogous to Kirkhoff's laws in electricity:
   (a) Conservation of through variables at each node: $\sum_{i\in\text{node}} T_i = 0$.
   (b) Conservation of across variables along closed loops: $\sum_{j\in\text{loop}} A_j = 0$.

Domain	**across variable** (A)	**through variable** (T)
electrical	voltage V	current I
fluid	pressure P	volume flow rate Q
mechanical (rotation)	angular velocity ω	torque T
mechanical (translation)	velocity v	force F

Let E_A be the energy stored in an A-type energy storage device, E_T be the energy stored in a T-type energy storage device, and let P represent dissipated power.

1. **Electrical** $C =$ capacitance, $L =$ inductance, $R =$ resistance

A variable	$E_A = \frac{1}{2}CV^2$	$I = C\frac{dV}{dt}$
T variable	$E_T = \frac{1}{2}LI^2$	$V = L\frac{dI}{dt}$
Power & constitutive	$P = IV = \frac{1}{R}V^2$	$I = \frac{1}{R}V$

2. **Fluid** $I =$ fluid inertance, $C_f =$ fluid capacitance, $R_f =$ fluid resistance

A variable	$E_A = \frac{1}{2}C_f P^2$	$Q = C_f\frac{dP}{dt}$
T variable	$E_T = \frac{1}{2}IQ^2$	$P = I\frac{dQ}{dt}$
Power & constitutive	$P = QP = \frac{1}{R_f}P^2$	$Q = \frac{1}{R_f}P$

3. **Mechanical (rotational)**
   $B =$ rotational damping, $K =$ rotational stiffness, $J =$ moment of inertia

A variable	$E_A = \frac{1}{2}J\omega^2$	$T = J\frac{d\omega}{dt}$
T variable	$E_T = \frac{1}{2}\frac{1}{K}F^2$	$\omega = \frac{1}{K}\frac{dT}{dt}$
Power & constitutive	$P = T\omega = B\omega^2$	$T = B\omega$

4. **Mechanical (translational)** $m =$ mass, $k =$ stiffness, $b =$ damping

A variable	$E_A = \frac{1}{2}mv^2$	$F = m\frac{dv}{dt}$
T variable	$E_T = \frac{1}{2}\frac{1}{k}F^2$	$v = \frac{1}{k}\frac{dF}{dt}$
Power & constitutive	$P = Fv = bv^2$	$F = bv$

9.22 OPTICS

- f = focal length
- M = magnification of a lens
- n = refractive index
- s = distance (s_o lens to object; s_i lens to image)
- P = power of a lens
- r = radius of curvature of a lens
- v = speed of light

Equations

1. A **converging lens** is "thicker in the center" than on the edges and has $f < 0$, a **diverging lens** is "thinner in the center" than on the edges and has $f > 0$.
2. **Focal length**
   (a) For thin lenses in contact: $\dfrac{1}{f} = \dfrac{1}{f_1} + \dfrac{1}{f_2}$
   (b) For thin lenses separated by d (with $d < f_1$): $\dfrac{1}{f} = \dfrac{1}{f_1} + \dfrac{1}{f_2} - \dfrac{d}{f_1 f_2}$
   (c) **Thin lens formula**: $\dfrac{1}{f} = \dfrac{1}{s_i} + \dfrac{1}{s_o}$.
3. **Lensmaker's equation** (d = thickness of a lens)
   (a) For a lens in air: $\dfrac{1}{f} = (n_{\text{lens}} - 1)\left(\dfrac{1}{R_1} - \dfrac{1}{R_2} + \dfrac{(n_{\text{lens}} - 1)d}{n_{\text{lens}} R_1 R_2}\right)$
   (b) For a thin lens: $\dfrac{1}{f} = \left(\dfrac{n_{\text{lens}} - n}{n}\right)\left(\dfrac{1}{R_1} - \dfrac{1}{R_2}\right)$
4. **Magnification** of a thin lens: $M = \dfrac{s_i}{s_o} = \dfrac{\text{image height}}{\text{object height}}$
5. **Lens power**
   (a) **Power of a thin lens**: $P = \dfrac{1}{f}$
   (b) **Power of many thin lenses in contact**: $P = \dfrac{1}{f_1} + \dfrac{1}{f_2} + \dfrac{1}{f_3} + \dfrac{1}{f_4} + \dots$
   (c) Power is measured in **diopter** units; the inverse of the focal length in meters (i.e., 1 meter$/f$). For example, a lens with a focal length of 500mm has $\frac{1\text{ meter}}{500\text{ mm}} = +2$ diopters. For lenses in contact, diopters can be added: a -2 diopter lens and a $+4$ diopter lens give a $+2$ diopter lens.
6. **Snell's law**: $n_1 \sin\theta_1 = n_2 \sin\theta_2$ where θ_i is the angle of the light ray.
7. Matrices in optics
   (a) **Propagation in free space**: $\begin{bmatrix} 1 & d \\ 0 & 1 \end{bmatrix}$ where d is distance
   (b) **Refraction at curved surface**: $\begin{bmatrix} 1 & 0 \\ \frac{n_1 - n_2}{R n_2} & \frac{n_1}{n_2} \end{bmatrix}$ ($R = \infty$ for flat surface)
   (c) **Reflection from curved mirror**: $\begin{bmatrix} 1 & 0 \\ -\frac{2}{R} & 1 \end{bmatrix}$ ($R = \infty$ for flat mirror)
   (d) **Thin lens**: $\begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}$

9.23 POPULATION GENETICS

Notation

- N_e effective population size
- N_f number of females
- N_m number of males
- p frequency of allele A
- q frequency of allele a
- f inbreeding coefficient (probability two alleles in a diploid zygote are identical by descent)
- s selection coefficient ($s = 1 - w$)
- u mutation rate $A \to a$
- v mutation rate $a \to A$
- w fitness ($w = 1 - s$)

Information

1. **Effective population size** (diploid individuals with separate sexes)
$$N_e = \frac{4N_f N_m}{N_f + N_m}$$
2. **Equilibrium: balanced polymorphism**
$$\left\{\hat{p} = \frac{s_{AA}}{s_{AA} + s_{aa}}, \quad \hat{q} = \frac{s_{aa}}{s_{AA} + s_{aa}}\right\}$$
3. **Equilibrium: recurrent mutation and selection** $\quad \hat{q} = \sqrt{\dfrac{u}{s_a}}$
4. **Equilibrium: mutation** $\quad \left\{\hat{p} = \dfrac{v}{u+v}, \quad \hat{q} = \dfrac{u}{u+v}\right\}$
5. **Fitness** $\quad \overline{w} = p^2 w_{AA} + 2pqw_{Aa} + q^2 w_{aa}$
6. **Genetic drift** (in one generation)
$$\text{variation} = \frac{pq}{N_e}$$
7. **Hardy–Weinberg equation** $\quad p^2 + 2pq + q^2 = 1$
8. **Inbreeding**
   - frequency of AA $\quad p^2 + pqf$
   - frequency of Aa $\quad p^2 - 2pqf$
   - frequency of aa $\quad q^2 + pqf$
9. **Mutation** (generation $(t-1)$ to generation t)
   - generational change $\quad \Delta = (up_{t-1} - vq_{t-1})$
   - frequency of A $\quad q_t = q_{t-1} + \Delta$
   - frequency of a $\quad p_t = p_{t-1} - \Delta$

9.24 QUANTUM MECHANICS

Notation

- A, B Hermitian operators
- c speed of light ($\approx 3 \times 10^8$ m/sec)
- E energy
- f frequency
- H Hamiltonian
- $\hbar = h/2\pi$
- J_* angular momentum
- m mass
- p, $\mathbf{p}$ momentum
- λ wavelength
- Ψ wavefunction
- σ standard deviation
- $[,]$ commutator
- $\langle , \rangle$ bracket

Equations

1. **Correspondences** $E \mapsto i\hbar\frac{\partial}{\partial t}, \quad p_x \mapsto -i\hbar\frac{\partial}{\partial x}, \quad \mathbf{p} \mapsto -i\hbar\nabla$
2. **Dirac matrices**
$$\gamma^0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \gamma^1 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \gamma^2 = \begin{bmatrix} 0 & 0 & 0 & -i \\ 0 & 0 & i & 0 \\ 0 & i & 0 & 0 \\ -i & 0 & 0 & 0 \end{bmatrix}, \gamma^3 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$
3. **Energy**
   (a) blackbody radiation (n=1,2,...) $E = n\hbar f$
   (b) photon (Plank's law) $E = \hbar f$
   (c) particle $E = pc = mc^2$
   (d) released by nuclear fission or fusion $E = \Delta m\, c^2$
4. **Heisenberg uncertainty principle**
$$\langle A^2 \rangle \, \langle B^2 \rangle \geq \frac{1}{4} \left| \langle [A, B] \rangle \right|^2 \qquad \text{or} \qquad \sigma_A \, \sigma_B \geq \frac{1}{2} \left| \langle [A, B] \rangle \right|$$
   Examples:
$$\sigma_x \, \sigma_{p_x} = \sigma_{\text{position}} \, \sigma_{\text{momentum}} \geq \frac{\hbar}{2} \qquad \sigma_{\text{position}} \, \sigma_{\text{kinetic energy}} \geq \frac{\hbar}{2m} \left| \langle p_x \rangle \right|$$
$$\sigma_{\text{energy}} \sigma_{\text{time}} \geq \frac{\hbar}{2} \qquad \sigma_{J_i} \, \sigma_{J_j} \geq \frac{\hbar}{2} \left| \langle J_k \rangle \right|$$
5. **Probability density** $= |\Psi|^2 = \Psi^* \Psi$
6. **Schrödinger equation**
$$i\hbar\frac{\partial \Psi}{\partial t} = H\Psi = \left(\frac{p^2}{2m} + V \right) \Psi = \left(-\frac{\hbar^2}{2m} \nabla^2 + V \right) \Psi$$
7. **Time evolution** $\frac{d}{dt} \langle A \rangle = \left\langle \frac{dA}{dt} \right\rangle + \frac{i}{\hbar} \langle [H, A] \rangle$
8. **Wavelength** $\lambda = \frac{h}{p}$

9.24.1 DIRAC BRAKET NOTATION

Bra-ket notation is used for vectors and linear functionals; its use is widespread in quantum mechanics. The elements are: "bra" $\langle\phi|$, "ket" $|\psi\rangle$, and "bra-ket" $\langle\phi|\psi\rangle$.

1. A **ket** is conceptually a column vector: $|\psi\rangle = \begin{bmatrix} c_0 & c_1 & c_2 & \dots \end{bmatrix}^{\mathrm{T}}$.
2. A **bra** is the conjugate transpose of the ket and vice versa.
3. The **bra** corresponding to the above ket $|\psi\rangle$ is conceptually the row vector $\langle\psi| = \begin{bmatrix} c_0^* & c_1^* & c_2^* & \dots \end{bmatrix}$.
4. The combination of a bra with a ket to form a complex number is called a bra-ket or braket.
5. If A is a linear operator, then
   (a) Applying the operator A to the ket $|\psi\rangle$ results in the ket $(A\,|\psi\rangle)$ which may be written $|A\psi\rangle$.
   (b) Composing the bra $\langle\phi|$ with the operator A results in the bra $(\langle\phi|\,A)$. This is a linear functional defined by $(\langle\phi|\,A)\,|\psi\rangle = \langle\phi|\,(A\,|\psi\rangle) = \langle\phi|A|\psi\rangle = \langle\phi|A\psi\rangle$
   (c) The **outer product** $|\phi\rangle\,\langle\psi|$ is a rank-one linear operator that maps the ket $|\rho\rangle$ to the ket $|\phi\rangle\,\langle\psi|\rho\rangle$; note that $\langle\psi|\rho\rangle$ is a scalar.
6. The bra and ket operators are linear. If the $\{c_i\}$ are complex numbers, then
   $\langle\phi|\,(c_1\,|\psi_1\rangle + c_2\,|\psi_2\rangle) = c_1\,\langle\phi|\psi_1\rangle + c_2\,\langle\phi|\psi_2\rangle$
   $(c_1\,\langle\phi_1| + c_2\,\langle\phi_2|)\,|\psi\rangle = c_1\,\langle\phi_1|\psi\rangle + c_2\,\langle\phi_2|\psi\rangle$
7. The bra and ket operators are associative:
   $\langle\phi|\,(A\,|\psi\rangle) = (\langle\phi|\,A)\,|\psi\rangle$
   $(A\,|\psi\rangle)\,\langle\phi| = A(|\psi\rangle\,\langle\phi|)$

In quantum mechanics:

1. The expectation of an observable represented by the linear operator A in the state $|\psi\rangle$ is $\langle\psi|A|\psi\rangle$.
2. $\langle\phi|\psi\rangle$ is the probability that state ψ collapses into the state ϕ.
3. The Schrödinger equation is $\hat{H}\,|\Psi\rangle = E\,|\Psi\rangle$
4. Quantum mechanical operators are easily represented:

$$\begin{aligned} \hat{p} &= \int \frac{dp}{2\pi}\ |p\rangle\ p\ \langle p| = \int \frac{dp}{2\pi}\ |p\rangle\ \left(-i\hbar\frac{\partial}{\partial x}\right)\ \langle p| \\ \hat{x} &= \int dx\ |x\rangle\ x\ \langle x| = \int dx\ |x\rangle\ \left(\frac{i}{\hbar}\frac{\partial}{\partial p}\right)\ \langle x| \end{aligned} \tag{9.24.1}$$

5. A quantum wave function is represented as $\Psi(\mathbf{x}) = \langle\mathbf{x}|\Psi\rangle$. A linear operator acting on this wavefunction is understood to operate on the underlying by kets by $A\psi(\mathbf{x}) = \langle\mathbf{x}|A|\psi\rangle$. Likewise: $\Psi_p = \langle p|\Psi\rangle$, and $\Psi_k = \langle k|\Psi\rangle$,

9.25 QUATERNIONS

- **p** point in space
- **q** quaternion
- **u** unit vector
- θ rotation angle

Equations

1. Quaternions are four-vectors that have the representations

$$\mathbf{q} = w + x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = [w \quad x \quad y \quad z] = (s, \mathbf{v})$$

   where $s = w$, $\mathbf{v} = [x \quad y \quad z]$, and $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ represent unit vectors that satisfy:

$$\mathbf{ii} = \mathbf{jj} = \mathbf{kk} = -1, \quad \mathbf{ij} = -\mathbf{ji} = \mathbf{k}, \quad \mathbf{jk} = -\mathbf{kj} = \mathbf{i}, \quad \mathbf{ki} = -\mathbf{ik} = \mathbf{j}$$

2. The *multiplication* of two quaternions $\mathbf{q}_1 = (s_1, \mathbf{v}_1)$ and $\mathbf{q}_2 = (s_2, \mathbf{v}_2)$ is another quaternion $\mathbf{q}_1\mathbf{q}_2 = (s_1 s_2 - \mathbf{v}_1 \cdot \mathbf{v}_2, s_1\mathbf{v}_2 + s_2\mathbf{v}_1 + \mathbf{v}_1 \times \mathbf{v}_2)$.
3. The *inner product* of two quaternions $\mathbf{q}_1 = (s_1, \mathbf{v}_1)$ and $\mathbf{q}_2 = (s_2, \mathbf{v}_2)$ is a scalar $\mathbf{q}_1 \cdot \mathbf{q}_2 = \mathbf{q}_2 \cdot \mathbf{q}_1 = s_1 s_2 + \mathbf{v}_1 \cdot \mathbf{v}_2$.
4. The *conjugate* of $\mathbf{q}$ is $\mathbf{q}' = w - x\mathbf{i} - y\mathbf{j} - z\mathbf{k} = (s, -\mathbf{v})$
5. The *magnitude* of $\mathbf{q}$ is $||\mathbf{q}|| = \sqrt{\mathbf{q}\mathbf{q}'} = \sqrt{w^2 + x^2 + y^2 + z^2} = \sqrt{s^2 + ||\mathbf{v}||^2}$
6. A *unit quaternion* has $||\mathbf{q}|| = 1$. For a unit quaternion $\mathbf{q}^{-1} = \mathbf{q}'$.
7. Quaternion multiplication is associative: $(\mathbf{q}_1\mathbf{q}_2)\mathbf{q}_3 = \mathbf{q}_1(\mathbf{q}_2\mathbf{q}_3)$
8. Quaternion multiplication is not commutative: $\mathbf{q}_1\mathbf{q}_2 \neq \mathbf{q}_2\mathbf{q}_1$ in general.
9. Rotations
   (a) For the unit vector $\mathbf{u}$, the unit quaternion $\mathbf{q} = (s, \mathbf{v})$ with $s = \cos\frac{\theta}{2}$ and $\mathbf{v} = \mathbf{u}\sin\frac{\theta}{2}$ represents a rotation about $\mathbf{u}$ by the angle θ. (The rotation is clockwise if our line of sight points in the direction pointed to by $\mathbf{u}$.)
   (b) A point in space $\mathbf{p}$ can be represented by the quaternion $\mathbf{P} = (0, \mathbf{p})$. This point, when rotated, has the representation $\mathbf{P}_{\text{rotated}} = \mathbf{q}\mathbf{P}\mathbf{q}^{-1} = \mathbf{q}\mathbf{P}\mathbf{q}'$.
   (c) To rotate $\mathbf{p}$ by the unit quaternions $\{\mathbf{q}_i\}_{i=1}^n$ (first by $\mathbf{q}_1$, then by $\mathbf{q}_2$, ...) form $\mathbf{P}_{\text{rotated by }\{\mathbf{q}_i\}} = (\mathbf{q}_n \cdots \mathbf{q}_2\mathbf{q}_1)\mathbf{P}(\mathbf{q}_n \cdots \mathbf{q}_2\mathbf{q}_1)'$.
   (d) The orthogonal matrix corresponding to a rotation by the unit quaternion $\mathbf{q} = a + b\mathbf{i} + c\mathbf{j} + d\mathbf{k}$ is

$$\begin{bmatrix} a^2 + b^2 - c^2 - d^2 & 2bc - 2ad & 2bd + 2ac \\ 2bc + 2ad & a^2 - b^2 + c^2 - d^2 & 2cd - 2ab \\ 2bd - 2ac & 2cd + 2ab & a^2 - b^2 - c^2 + d^2 \end{bmatrix}$$

EXAMPLE Consider rotation about the axis $\mathbf{w} = \mathbf{i} + \mathbf{j} + \mathbf{k}$ by an angle of $120^\circ = \frac{2\pi}{3}$ radians. The appropriate unit quaternion is

$$\mathbf{q} = (s, \mathbf{v}) = \left(\cos\frac{\theta}{2}, \frac{\mathbf{w}}{||\mathbf{w}||}\sin\frac{\theta}{2}\right) = \left(\cos\frac{\pi}{3}, \frac{\mathbf{i}+\mathbf{j}+\mathbf{k}}{\sqrt{3}}\sin\frac{\pi}{3}\right) = \left(\frac{1}{2}, \frac{\mathbf{i}+\mathbf{j}+\mathbf{k}}{2}\right)$$

If a point has the representation $\mathbf{p} = a\mathbf{i} + b\mathbf{j} + c\mathbf{k}$ then

$$\mathbf{P}_{\text{rotated}} = \mathbf{q}\ (0, \mathbf{p})\ \mathbf{q}' = \frac{(1, \mathbf{i}+\mathbf{j}+\mathbf{k})}{2}(0, a\mathbf{i} + b\mathbf{j} + c\mathbf{k})\frac{(1, -\mathbf{i}-\mathbf{j}-\mathbf{k})}{2} = c\mathbf{i} + a\mathbf{j} + b\mathbf{k}$$

as expected. (The rotation corresponds to a cube, held fixed at one point, being rotated about the long diagonal through the fixed point. This permutes the axes cyclically.)

9.26 RADAR

- c speed of light ($\approx 3 \times 10^8$ m/sec)
- F Receiver noise factor
- G_r Receiver gain
- G_t Transmitter gain
- L General loss factor
- P_r Received power (Watts)
- P_t Transmit power (Watts)
- R_r Range - receiver to target (m)
- R_t Range - transmitter to target (m)
- T effective noise temperature (°K)
- k_B Boltzmann's constant $k_B \approx 1.38 \times 10^{-23} \frac{\text{J}}{\text{K}}$
- λ Radar wavelength (m)
- σ Target radar cross section (m^2)
- τ pulse duration (s)

If the radar is monostatic then $G_r = G_t$ and $R_r = R_t$.

1. **Radar equation**
$$P_r = \frac{P_t G_t G_r \lambda^2 \sigma}{(4\pi)^3 R_t^2 R_r^2 L}$$
2. **Noise** (assuming the white noise power spectral density (PSD) is kT)
$$N = \frac{kTF}{\tau}$$
3. **Signal to Noise ratio** (SNR) usually expressed in dB
$$\text{SNR} = \frac{P_r}{N} \qquad \text{SNR in dB} = 10 \log_{10}\left(\frac{P_r}{N}\right)$$
4. Pulse repetition frequency (PRF)$= 1/\tau$
5. The practical (unambiguous) range$= c$(PRI)$/2$
6. Radar cross section for different geometries (when λ is sufficiently small)
   (a) corner reflectors (a is a characteristic length)
   dihedral: $\sigma = \dfrac{8\pi a^4}{\lambda^2}$ trihedral: $\sigma = \dfrac{12\pi a^4}{\lambda^2}$
   (b) planar surface of area A: $\sigma = \dfrac{4\pi A^2}{\lambda^2}$
   (c) right circular cylinder (radius a, height b): $\sigma = \dfrac{2\pi a b^2}{\lambda}$
   (d) sphere of radius a: $\sigma = \pi a^2$

Radar cross sections (for a cm radar)	
Insect	0.00001 m^2
Bird	0.01 m^2
Stealth aircraft	<0.1 m^2
Surface-to-air-missile	≈0.1 m^2
Human	1 m^2
Combat aircraft	2–6 m^2
Cargo aircraft	up to 100 m^2
Corner reflector with 1.5 m edge	20,000 m^2
Container ship (200 m)	10,000–80,000 m^2

Standard Radar Frequencies		
Band	Frequency	Wavelength
L	1–2 GHz	30.0–15.0 cm
S	2–4 GHz	15–7.5 cm
C	4–8 GHz	7.5–3.8 cm
X	8–12 GHz	3.8–2.5 cm
Ku	12–18 GHz	2.5–1.7 cm
K	18–27 GHz	1.7–1.1 cm
Ka	27–40 GHz	1.1–0.75 cm
V	40–75 GHz	0.75–0.40 cm
W	75–110 GHz	0.40–0.27 cm

9.27 RELATIVISTIC MECHANICS

- c speed of light ($\approx 3 \times 10^8$ m/sec)
- m mass
- t time
- x distance
- v velocity
- z redshift
- β Lorentz transformation factor

Equations

1. **Factor** $\beta = \sqrt{1 - \frac{v^2}{c^2}}$
2. **Redshift** $z = \frac{v}{c}$
3. **Relativistic length contraction** $\Delta x = \beta\, \Delta x_0$
4. **Relativistic mass increase** $m = \frac{m_0}{\beta}$
5. **Relativistic time dilation** $\Delta t = \frac{\Delta t_0}{\beta}$

9.27.1 LORENTZ TRANSFORMATION

- c speed of light ($\approx 3 \times 10^8$ m/sec)
- $x^0 = ct$
- $\{x^1, x^2, x^3\}$ spatial dimensions
- $\mathbf{x}^{\mathrm{T}} = \begin{bmatrix} x^0 & x^1 & x^2 & x^3 \end{bmatrix}$
- t time

The Lorentz group of transformations, **A**, leaves the length of the 4-vector **x**, invariant under the flat space-time metric: $g = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$. Proper Lorentz transformations have $|A| = 1$ and are continuous with the identity transformation.

The transformation A can be written as $A = \exp\left(-\sum_{i=1}^{3} (\theta_i S_i + \zeta_i K_i)\right)$ where θ and ζ are constant 3-vectors; their components are the six parameters of the transformation. The $\{S_i\}$ matrices generate rotations in three spatial dimensions:

$$S_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \qquad S_2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{bmatrix}, \qquad S_3 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{9.27.1}$$

The $\{K_i\}$ matrices produce boosts:

$$K_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad K_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad K_3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix} \tag{9.27.2}$$

Powers of these matrices may be produced from the relations:

$$\begin{gathered} S_1^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \qquad S_2^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}, \qquad S_3^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \\ K_1^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad K_2^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad K_3^2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \end{gathered} \tag{9.27.3}$$

along with $S_i^3 = -S_i$, $K_i^3 = K_i$ and the commutator relations $[S_i, S_j] = \epsilon_{ijk} S_k$, $[S_i, K_j] = \epsilon_{ijk} K_k$, and $[K_i, K_j] = -\epsilon_{ijk} K_k$, where ϵ_{ijk} is the permutation symbol.

9.28 SOLID MECHANICS

Notation

- A area
- C stiffness (tensor)
- E Young's modulus
- F, $\mathbf{F}$ force
- G shear modulus ($G = \mu$)
- I identity matrix
- K bulk modulus
- P pressure
- $\mathbf{u}$ displacement
- v velocity
- V volume
- ϵ strain (tensor)
- λ first Lamé parameter
- μ shear modulus ($\mu = G$)
- ν Poisson's ratio
- ρ density
- σ stress (tensor)
- $\mathrm{tr}(\,)$ trace function

Equations

1. **Basic relations**

$$\sigma = \frac{F}{A} \qquad P = -\frac{1}{3}\,\mathrm{tr}(\sigma) \qquad K = -V\frac{\partial P}{\partial V}$$

$$\epsilon = \frac{1}{2}\left(\nabla\mathbf{u} + (\nabla\mathbf{u})^{\mathrm{T}}\right) \quad \text{or} \quad \epsilon_{ij} = \frac{1}{2}\left(u_{i,j} + u_{j,i}\right) = \frac{1}{2}\left(\partial_i u_j + \partial_j u_i\right)$$

$$\sigma_{ij} = \sigma_{ji} \qquad \epsilon_{ij} = \epsilon_{ji} \qquad C_{ijkl} = C_{klij} = C_{jikl} = C_{ijlk}$$

2. **Constants**: **Engineering** $\{G, Y, \nu\}$ **Lamé** $\{\lambda, \mu\}$
3. **Relations between constants** (homogeneous isotropic materials)

$$E = \frac{\mu(3\lambda + 2\mu)}{\lambda + \mu} = 2G(1+\nu) = 3K(1-2\nu)$$

$$\nu = \frac{\lambda}{2(\lambda+\mu)} = \frac{E}{2G} - 1$$

$$\mu = G = \frac{\lambda(1-2\nu)}{2\nu} = \frac{E}{2(1+\nu)}$$

$$\lambda = \frac{E\nu}{2(1+\nu)(1-2\nu)} = K - \frac{2G}{3}$$

4. **Equations of motion** $\rho\ddot{\mathbf{u}} = \nabla\cdot\sigma + \mathbf{F}$ or $\rho\partial_{tt}u_i = \sigma_{ji,j} + F_i$
5. **Hooke's Law** $\sigma_{ij} = C_{ijkl}\epsilon_{kl}$
6. **Navier equations** (steady state) $(\lambda + 2\mu)\nabla(\nabla\cdot\mathbf{u}) - \mu\nabla\times(\nabla\times\mathbf{u}) = 0$
7. **Saint-Venant's conditions** $\epsilon_{ij,km} + \epsilon_{km,ij} - \epsilon_{ik,jm} - \epsilon_{jm,ik} = 0$
8. **Speed of sound** $v = \sqrt{\dfrac{K}{\rho}}$
9. **Stress–strain relation** (homogeneous isotropic material)

$$\sigma = 2\mu e + \lambda\,\mathrm{tr}(\epsilon)I \quad \text{or} \quad \sigma_{ij} = 2\mu\varepsilon_{ij} + \lambda\delta_{ij}\epsilon_{kk}$$

$$\epsilon_{ij} = \frac{1}{E}\left[(1+\nu)\sigma_{ij} - \nu\delta_{ij}\sigma_{kk}\right]$$

9.29 STATISTICAL MECHANICS

Notation

- E_i energy
- m mass
- T temperature
- $Z(T)$ partition function
- $\mathbf{v}$ velocity
- μ chemical potential
- ρ probability density
- g_i degeneracy (number of states with energy E_i)
- k_B Boltzmann's constant $\approx 1.38 \times 10^{-23}\,\frac{\text{J}}{\text{K}}$
- N_i number of particles with energy E_i

Equations

1. **Boltzmann distribution** $$\frac{N_i}{N} = \frac{g_i e^{-E_i/(k_B T)}}{Z(T)}$$
   with $N = \sum_i N_i$ and $Z(T) = \sum_i g_i e^{-E_i/(k_B T)}$

2. **Energy distributions** [$f(E)$ is the probability that a particle has energy E]
   (a) Bose–Einstein $$f(E) = \frac{1}{e^{E/kT} - 1}$$
   (b) Fermi–Dirac $$f(E) = \frac{1}{e^{E/kT} + 1}$$
   (c) Maxwell–Boltzmann $$f(E) = \frac{1}{e^{E/kT}}$$

3. **Characteristic speeds**
   (a) Average speed $$v_{\text{average}} = \frac{2}{\sqrt{\pi}} v_{\max}$$
   (b) Maximum speed $$v_{\max} = \sqrt{\frac{2T}{m}}$$
   (c) Root mean square speed $$v_{\text{rms}} = \frac{3}{2} v_{\max}$$

4. **Maxwellian distribution** $$\rho(\mathbf{v}) = n_0 \left(\frac{m}{2\pi T}\right)^{3/2} e^{-m|\mathbf{v}|^2/2T}$$

5. **Particle thermal energy**
   (a) One dimensional $$\frac{m\overline{v_x^2}}{2} = \frac{T}{2}$$
   (b) Three dimensional $$\frac{m\overline{v^2}}{2} = \frac{3T}{2}$$

9.30 THERMODYNAMICS

- State variables
  - N particle number
  - P pressure
  - Q heat
  - S entropy
  - T temperature
  - V volume
  - μ chemical potential
  - ν_s frequency
- Thermodynamic potentials
  - A Helmholtz free energy
  - G Gibbs free energy
  - H Enthalpy
  - U Internal energy
- σ Stefan's constant ($\approx 5.67 \times 10^{-8} \frac{\mathrm{W}}{\mathrm{m^2K^4}}$)

1. **Laws of thermodynamics**
   (a) Energy cannot be created, or destroyed, only modified in form.
   $dU = \delta Q - \delta W$ where
      i. dU increase in internal energy of the system
      ii. δW infinitesimal amount of Work (W)
      iii. δQ infinitesimal amount of Heat (Q)
      iv. δ is an "inexact differential" (i.e., path-dependent)

   (b) A system operating in a cycle cannot produce a positive heat flow from a colder body to a hotter body. $\displaystyle\int \frac{\delta Q}{T} = \int dS \geq 0$
   (c) All processes cease as temperature approaches zero.
   If T goes to zero then S becomes constant.

   The zeroth law states that if two systems are in equilibrium with a third system, then the two systems are in equilibrium with each other.
2. **Thermodynamic potentials**
$$dA\,(T, V, N_i) = -S\,dT - P\,dv + \sum_i \mu_i dN_i$$
$$dG\,(T, P, N_i) = -S\,dT + V\,dP + \sum_i \mu_i dN_i$$
$$dH\,(S, P, N_i) = \quad T\,dS + V\,dP + \sum_i \mu_i dN_i$$
$$dU\,(S, V, N_i) = \quad T\,dS - P\,dV + \sum_i \mu_i dN_i$$
3. **Entropy change at constant** T (for phase changes) $\Delta S = Q/T$
4. **Planck's law** (electromagnetic radiation at all wavelengths emitted from a black body in a cavity in thermodynamic equilibrium) $\dfrac{2h\nu^3}{c^2}\dfrac{1}{e^{h\nu/kT} - 1}$
5. **Stefan's law** rate of energy radiated $= \sigma T^4$
6. **Work done on/by a gas** $W = P\,\Delta V$

Chapter 10

Miscellaneous

10.1 CALENDAR COMPUTATIONS

10.1.1 DAY OF WEEK FOR ANY GIVEN DAY

The following formula gives the day of the week for the Gregorian calendar (i.e., for any date after 1582):

$$W \equiv \left(d + \lfloor 2.6m - 0.2 \rfloor + Y + \left\lfloor \frac{Y}{4} \right\rfloor + \left\lfloor \frac{C}{4} \right\rfloor - 2C\right) \pmod 7 \qquad (10.1.1)$$

where

- W is the day of the week (Sunday $\Rightarrow 0, \ldots,$ Saturday $\Rightarrow 6$).
- d is the day of the month (1 to 31).
- m is the month where January and February are treated as months of the preceding year: March $\Rightarrow$ 1, April $\Rightarrow$ 2, ..., December $\Rightarrow$ 10, January $\Rightarrow$ 11, February $\Rightarrow$ 12.
- C is the century minus one (1997 has $C = 19$ while 2025 has $C = 20$).
- Y is the year (1997 has $Y = 97$ except $Y = 96$ for January and February).
- $\lfloor \cdot \rfloor$ denotes the integer floor function.
- The "mod" function returns a non-negative value.

EXAMPLE Consider the date 16 March 2017 for which $d = 16$, $m = 1$, $C = 20$, and $Y = 17$. From Equation (10.1.1), we compute

$$W \equiv 16 + \lfloor 2.4 \rfloor + 17 + \left\lfloor \frac{17}{4} \right\rfloor + \left\lfloor \frac{20}{4} \right\rfloor - 40 \pmod 7$$
$$\equiv 2 + 2 + 3 + 4 + 5 - 5 \pmod 7 \equiv 4 \pmod 7$$

So this date was a Thursday.

Notes:

1. In any given year the following dates fall on the same day of the week: 4/4, 6/6, 8/8, 10/10, 12/12, 9/5, 5/9, 7/11, 11/7, and the last day of February.
2. Because 7 does not divide 400, January 1 occurs more frequently on some days of the week than on others! In a cycle of 400 years, January 1 and March 1 occur on the following days with the following frequencies:

	Sun	Mon	Tue	Wed	Thu	Fri	Sat
January 1	58	56	58	57	57	58	56
March 1	58	56	58	56	58	57	57

3. The 13th of a month is more likely to be a Friday than any other day.
4. Excel uses a sequential integer for each day, with day 1 being 1 January 1900.

10.1.2 LEAP YEARS

If a year is divisible by 4, then it will be a leap year, unless the year is divisible by 100 (when it will not be a leap year), unless the year is divisible by 400 (when it will be a leap year). Hence the list of leap years includes 1896, 1904, 1908, 1992, 1996, 2000, 2004, 2008 and the list of non-leap years includes 1900, 1998, 1999, 2001.

10.2 CELLULAR AUTOMATA

In a cellular automata there is a grid of *cells* with each cell containing a *state*. The cellular automata evolves according to rules. For a specific cell, the current state of that cell and the cell's *neighbors* control the state of that cell in the next iteration.

In the simplest case the cells are in a rectangular grid and the states are either 0 or 1 (indicated graphically as white and black). We assume that initially all cells, except some specific ones, are white.

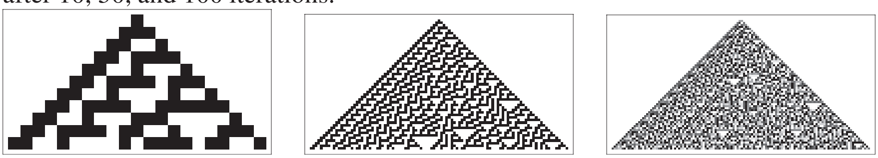
In a one-dimensional cellular automata there is a row of cells whose contents are replaced each iteration using the rules. Graphically, the sequence of states is shown as a two-dimensional figure, with each new set of states appearing below the previous states. If a cell is updated based on its current value and the value of its $2k$ nearest neighbors (k on each side), then the next state is based on the current state of $2k+1$ of cells. There are $2^{\left(2^{2k+1}\right)}$ possible sets of rules in this case.

In the case of $k = 1$ there are $2^{\left(2^3\right)} = 256$ sets of rules. A specific rule in this case is a set of triplet mappings, denoted textually $\{(111 \Rightarrow 0), (110 \Rightarrow 0), \ldots, (001 \Rightarrow 1), (000 \Rightarrow 0)\}$ or graphically:

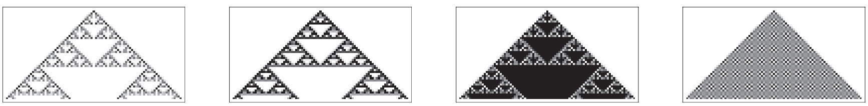
111	110	101	100	011	010	001	000
■■■	■■□	■□■	■□□	□■■	□■□	□□■	□□□
⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓
□ , 0	□ , 0	□ , 0	■ , 1	■ , 1	■ , 1	■ , 1	□ , 0

The sequence of 0's and 1's that the rule produces can be interpreted as the number of the rule. Hence, the above rule is rule number 30 (since $30 = 00011110_2$).

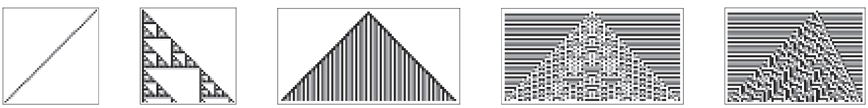
Starting from a single black square, the following shows the evolution of rule 30 after 10, 50, and 100 iterations.



Even in the simple case of a one-dimensional cellular automata, with two types of states, and only nearest neighbor interactions, there is a wide variety of possible behavior. Below are the results of several different rules after 50 iterations:



26 126 182 122



138 60 94 73 89

- The cellular automata with rule 30 is chaotic.
- Conway's *Game of Life* is a two-dimensional cellular automata.

10.3 COMMUNICATION THEORY

10.3.1 INFORMATION THEORY

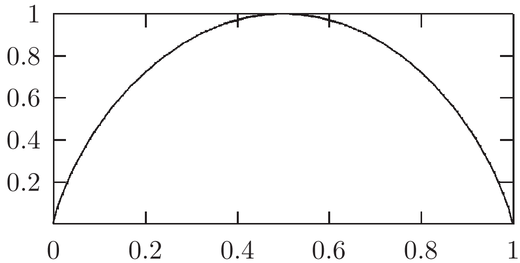
Let $\mathbf{p}_X = (p_{x_1}, p_{x_2}, \ldots, p_{x_n})$ be the probability distribution of the discrete random variable X with Prob $(X = x_i) = p_{x_i}$. The *entropy* of the distribution is

$$H(\mathbf{p}_X) = -\sum_{x_i} p_{x_i} \log_2 p_{x_i}. \tag{10.3.1}$$

The units for entropy are *bits*. Entropy measures how much information is gained from learning the value of X. When X takes only two values, $\mathbf{p} = (p, 1-p)$, then

$$H(\mathbf{p}_X) = H(p, 1-p) = -p \log_2 p - (1-p) \log_2(1-p). \tag{10.3.2}$$

This is also denoted $H(p)$. The range of $H(p)$ is from 0 to 1 with a maximum at $p = 0.5$. Below is a plot of p versus $H(p)$. The maximum of $H(\mathbf{p}_X)$ is $\log_2 n$ and is obtained when X is uniformly distributed, taking n values.



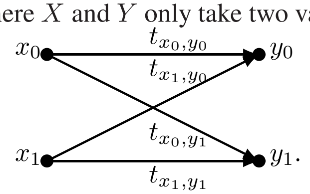
Given two discrete random variables X and Y, $\mathbf{p}_{X \times Y}$ is the joint distribution of X and Y. The *mutual information* of X and Y is defined by

$$I(X, Y) = H(\mathbf{p}_X) + H(\mathbf{p}_Y) - H(\mathbf{p}_{X \times Y}). \tag{10.3.3}$$

Note that (a) $I(X, Y) = I(Y, X)$; (b) $I(X, Y) \geq 0$; and (c) $I(X, Y) = 0$ if and only if X and Y are independent. Mutual information gives the amount of information that learning a value of X says about the value of Y (and vice versa).

10.3.1.1 Channel capacity

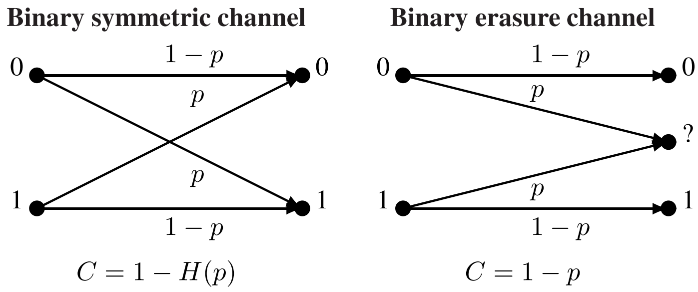
The *transition probabilities* are defined by $t_{x,y} = \text{Prob}\,(Y = y \mid X = x)$. The distribution $\mathbf{p}_X$ determines $\mathbf{p}_Y$ by $p_y = \sum t_{x,y} p_x$. The matrix $T = (t_{x,y})$ is the *transition matrix*. The matrix T defines a *channel* given by a transition diagram (input is X, output is Y). For example (here X and Y only take two values),



The *capacity* of the channel is defined as

$$C = \max_{\mathbf{p}_X} I(X, Y). \tag{10.3.4}$$

A channel is *symmetric* if each row is a permutation of the first row and each column is a permutation of the first column. The capacity of a symmetric channel is $C = \log_2 n - H(\mathbf{p})$, where $\mathbf{p}$ is the first row; the capacity is achieved when $\mathbf{p}_X$ represents equally likely inputs. The channel shown on the left is symmetric; both channels achieve capacity with equally likely inputs.



10.3.1.2 Shannon's theorem

Let both X and Y be discrete random variables with values in an alphabet A. A *code* is a set of *codewords* (n-tuples with entries from A) that is in one-to-one correspondence with a set of M messages. The *rate* R of the code is defined as $\frac{1}{n}\log_2 M$. Assume that the codeword is sent via a channel with transition matrix T by sending each vector element independently. Define

$$e = \max_{\text{all codewords}} \text{Prob}\left((\text{codeword incorrectly decoded})\right). \tag{10.3.5}$$

Shannon's coding theorem states:

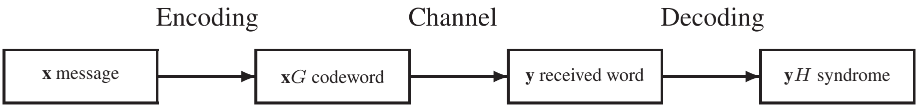
1. If $R < C$, then there is a sequence of codes with rate R and $n \to \infty$ such that $e \to 0$.
2. If $R \geq C$, then e is always bounded away from 0.

10.3.2 BLOCK CODING

10.3.2.1 Definitions

A *code* C over an alphabet A is a set of vectors of a fixed length n with entries from A. Let A be the finite field GF(q) (see Section 2.5.8.1). If C is a vector space over A, then C is a *linear code*; the *dimension* k of a linear code is its dimension as a vector space. The *Hamming distance* $d_{\text{H}}(\mathbf{u}, \mathbf{v})$ between two vectors, $\mathbf{u}$ and $\mathbf{v}$, is the number of places in which they differ. For a vector $\mathbf{u}$ over GF(q), define the *weight*, wt($\mathbf{u}$), as the number of non-zero components. Then $d_{\text{H}}(\mathbf{u}, \mathbf{v}) = \text{wt}(\mathbf{u} - \mathbf{v})$. The minimum Hamming distance between two distinct vectors in a code C is called the *minimum distance* d. A code can detect e errors if $e < d$. A code can correct t errors if $2t + 1 \leq d$.

10.3.2.2 Coding diagram for linear codes



1. A *message* $\mathbf{x}$ consists of k information symbols.
2. The message is encoded as $\mathbf{x}G \in C$, where G is a $k \times n$ matrix called the generating matrix.
3. After transmission over a channel, a (possibly corrupted) vector $\mathbf{y}$ is received.
4. There exists a *parity check matrix* H such that $\mathbf{c} \in C$ if and only if $\mathbf{c}H = \mathbf{0}$. Thus the *syndrome* $\mathbf{z} = \mathbf{y}H$ can be used to try to decode $\mathbf{y}$.
5. If G has the form $[I\ A]$, where I is the $k \times k$ identity matrix, then $H = \begin{bmatrix} -A \\ I \end{bmatrix}$

10.3.2.3 Cyclic codes

A linear code C of length n is *cyclic* if $(a_0, a_1, \ldots, a_{n-1}) \in C$ implies $(a_{n-1}, a_0, \ldots, a_{n-2}) \in C$. To each codeword $(a_0, a_1, \ldots, a_{n-1}) \in C$ is associated the polynomial $a(x) = \sum_{i=0}^{n-1} a_i x^i$. Every cyclic code has a *generating polynomial* $g(x)$ such that $a(x)$ corresponds to a codeword if and only if $a(x) \equiv d(x)g(x) \pmod{x^n - 1}$ for some $d(x)$. The *roots* of a cyclic code are roots of $g(x)$ in some extension field GF(q') with primitive element α.

1. *BCH Bound:* If a cyclic code C has roots $\alpha^i, \alpha^{i+1}, \ldots, \alpha^{i+d-2}$, then the minimum distance of C is at least d.
2. *Binary BCH codes* (BCH stands for Bose, Ray-Chaudhuri, and Hocquenghem): Fix m, define $n = 2^m - 1$, and let α be a primitive element in GF(2^m). Define $f_i(x)$ as the minimum binary polynomial of α^i. Then
$$g(x) = \text{LCM}\,(f_1(x), \ldots, f_{2e}(x)) \tag{10.3.6}$$
defines a generating polynomial for a binary BCH code of length n and minimum distance at least $\delta = 2e + 1$ (δ is called the *designed distance*). The code dimension is at least $n - me$.
3. *Dual code:* Given a code C, the dual code is $C^\perp = \{\mathbf{a} \mid \mathbf{a} \cdot \mathbf{x} = 0 \text{ for all } \mathbf{x} \in C\}$. The code $C^\perp$ is an $(n, n-k)$ linear code over the same field. A code is *self-dual* if $C = C^\perp$.
4. *MDS codes:* A linear code that meets the Singleton bound, $n + 1 = k + d$, is called *MDS* (for *maximum distance separable*). Any k columns of a generating matrix of an MDS code are linearly independent.
5. *Reed–Solomon codes:* Let α be a primitive element for GF(q) and $n = q - 1$. The generating polynomial $g(x) = (x - \alpha)(x - \alpha^2)\cdots(x - \alpha^{d-1})$ defines a cyclic MDS code with distance d and dimension $k = n - d + 1$.
6. *Perfect codes:* A linear code is *perfect* if it satisfies the Hamming bound, $q^{n-k} = \sum_{i=0}^{e} \binom{n}{i}(q-1)^i$. The binary Hamming codes and Golay codes are perfect.

7. *Binary Hamming codes:* These codes have parameters $\{n = 2^m - 1, k = 2^m - 1 - m, d = 3\}$. The parity check matrix is the $2^m - 1 \times m$ matrix whose rows are all of the binary m-tuples in a fixed order. The generating and parity check matrices for the $(7, 4)$ Hamming code are

$$G = \begin{bmatrix} I & A \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}, \qquad H = \begin{bmatrix} -A \\ I \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

8. *Ternary Golay code:* This has the parameters $\{n = 12, k = 6, d = 6\}$. The generating matrix is

$$G = \begin{bmatrix} I & A \end{bmatrix} = \begin{bmatrix} 1\,0\,0\,0\,0\,0 & 0\,1\,1\,1\,1\,1 \\ 0\,1\,0\,0\,0\,0 & 1\,0\,1\,2\,2\,1 \\ 0\,0\,1\,0\,0\,0 & 1\,1\,0\,1\,2\,2 \\ 0\,0\,0\,1\,0\,0 & 1\,2\,1\,0\,1\,2 \\ 0\,0\,0\,0\,1\,0 & 1\,2\,2\,1\,0\,1 \\ 0\,0\,0\,0\,0\,1 & 1\,1\,2\,2\,1\,0 \end{bmatrix}. \tag{10.3.7}$$

10.3.2.4 Table of best binary codes

Let $A(n, d)$ be the number of codewords[1] in the largest binary code of length n and minimum distance d. Note that $A(n - 1, d - 1) = A(n, d)$ if d is odd and $A(n, 2) = 2^{n-1}$ (given, e.g., by even weight words).

n	$d = 4$	$d = 6$	$d = 8$	$d = 10$
6	4	2	1	1
7	8	2	1	1
8	16	2	2	1
9	20	4	2	1
10	40	6	2	2
11	72–79	12	2	2
12	144–158	24	4	2
13	256	32	4	2
14	512	64	8	2
15	1024	128	16	4
16	2048	256	32	4
17	2720–3276	256–340	36–37	6
18	5248–6552	512–680	64–74	10
19	10496–13104	1024–1288	128–144	20
20	20480–26208	2048–2372	256–279	40
21	36864–43690	2560–4096	512	42–48
22	73728–87380	4096–6942	1024	68–88
23	147456–173784	8192–13774	2048	64–150
24	294912–344636	16384–24106	4096	128-280

[1]Data from *Sphere Packing, Lattices and Groups* by J. H. Conway and N. J. A. Sloane, 2nd ed., Springer-Verlag, New York, 1993.

10.3.2.5 Bounds

Bounds for block codes investigate the trade-offs between the length n, the number of codewords M, the minimum distance d, and the alphabet size q. The number of errors that can be corrected is e with $2e+1 \leq d$. If the code is linear, then the bounds concern the dimension k with $M = q^k$.

1. *Hamming* or *sphere-packing bound*: $M \leq q^n \Big/ \sum_{i=0}^{e} \binom{n}{i}(q-1)^i$.
2. *Plotkin bound*: Suppose that $d > n(q-1)/q$. Then $M \leq \frac{qd}{qd-n(q-1)}$.
3. *Singleton bound*: For any code, $M \leq q^{n-d+1}$; if the code is linear, then $k + d \leq n + 1$.
4. *Varsharmov–Gilbert bound*: There is a block code with minimum distance at least d and $M \geq q^n \Big/ \sum_{i=0}^{d-1} \binom{n}{i}(q-1)^i$.

10.3.3 SOURCE CODING FOR ENGLISH TEXT

English text has, on average, 4.08 bits/character.

Letter	Probability	Huffman code	Alphabetical code
Space	0.1859	000	00
A	0.0642	0100	0100
B	0.0127	011111	010100
C	0.0218	11111	010101
D	0.0317	01011	01011
E	0.1031	101	0110
F	0.0208	001100	011100
G	0.0152	011101	011101
H	0.0467	1110	01111
I	0.0575	1000	1000
J	0.0008	0111001110	1001000
K	0.0049	01110010	1001001
L	0.0321	01010	100101
M	0.0198	001101	10011
N	0.0574	1001	1010
O	0.0632	0110	1011
P	0.0152	011110	110000
Q	0.0008	0111001101	110001
R	0.0484	1101	11001
S	0.0514	0010	1101
T	0.0796	0010	1110
U	0.0228	11110	111100
V	0.0083	0111000	111101
W	0.0175	001110	111110
X	0.0013	0111001100	1111110
Y	0.0164	001111	11111110
Z	0.0005	0111001111	11111111
Cost	4.0799	4.1195	4.1978

10.4 CONTROL THEORY

General terminology

1. If the control is bounded above and below (say $u_i^- < u_i < u_i^+$), then a *bang–bang* control is one for which $u_i = u_i^-$ or $u_i = u_i^+$. (That is, for every t, either $u_i(t) = u_i^-(t)$ or $u_i(t) = u_i^+(t)$; switches are possible.) A *bang–off–bang* control is one for which $u_i = 0$, $u_i = u_i^-$, or $u_i = u_i^+$.
2. If $u(t)$ is the control signal sent to a system, $y(t)$ is the measured output, $r(t)$ is the measured output, and the error is $r(t) = r(t) - y(t)$, then the Proportional-Integral-Differential (PID) controller has the form

$$u(t) = K_P\, e(t) + K_I \int e(t)\, dt + K_D \frac{de(t)}{dt} \tag{10.4.1}$$

3. In H_∞ (i.e., "H-infinity") methods a controller is selected to minimize the H_∞ norm. The H_∞ norm is the maximum singular value of a matrix-valued function; over the space of matrix-valued functions which are analytic and bounded in the open right-half complex plane.

10.4.1 CONTINUOUS LINEAR TIME-INVARIANT SYSTEMS

Let $\mathbf{x}$ be a state vector, let $\mathbf{y}$ be an observation vector, and let $\mathbf{u}$ be the control. If a system evolves as:

$$\begin{aligned} \dot{\mathbf{x}} &= A\mathbf{x} + B\mathbf{u} \\ \mathbf{y} &= C\mathbf{x} + D\mathbf{u} \end{aligned} \tag{10.4.2}$$

1. Taking Laplace transforms and solving the algebraic equations results in $\tilde{\mathbf{y}} = G(s)\tilde{\mathbf{u}}$ where $G(s)$ is the transfer function $G(s) = C(sI - A)^{-1}B + D$.
2. A system is said to be *controllable* if and only if for any times $\{t_0, t_1\}$ and any states $\{\mathbf{x}_0, \mathbf{x}_1\}$ there exists a control $\mathbf{u}(t)$ such that $\mathbf{x}(t_0) = \mathbf{x}_0$ and $\mathbf{x}(t_1) = \mathbf{x}_1$. The system is controllable if and only if

$$\operatorname{rank}\begin{bmatrix} B & AB & A^2B & \dots & A^{n-1}B \end{bmatrix} = n. \tag{10.4.3}$$

   For the discrete linear time-invariant system $\mathbf{x}(k+1) = A\mathbf{x}(k) + B\mathbf{u}(k)$ the system is controllable if and only if Equation (10.4.3) is satisfied.
3. If, given $\mathbf{u}(t)$ and $\mathbf{y}(t)$ on some interval $t_0 < t < t_1$, the value of $\mathbf{x}(t)$ can be deduced on that interval, then the system is said to be *observable*. Observability is equivalent to the condition

$$\operatorname{rank}\begin{bmatrix} C^{\mathrm{T}} & A^{\mathrm{T}}C^{\mathrm{T}} & \dots & \left(A^{(n-1)}\right)^{\mathrm{T}} C^{\mathrm{T}} \end{bmatrix} = n. \tag{10.4.4}$$

10.4.2 PONTRYAGIN'S MAXIMUM PRINCIPLE

Consider a system evolving as

$$\frac{d\mathbf{x}}{dt} = \mathbf{b}(t, \mathbf{x}, \mathbf{u}) \qquad \mathbf{x}(0) = \text{given} \qquad t \geq 0 \tag{10.4.5}$$

where $\mathbf{u}$ is the control. The goal is to minimize the total cost, defined to be

$$V(\mathbf{x}(0)) = \int_0^\tau c(t, \mathbf{x}, \mathbf{u})\, dt + C(\tau, \mathbf{x}(\tau)) \tag{10.4.6}$$

Define the Hamiltonian

$$H(t, \mathbf{x}, \mathbf{u}, \lambda) = \lambda^{\mathrm{T}} \mathbf{b}(t, \mathbf{x}, \mathbf{u}) - c(t, \mathbf{x}, \mathbf{u}) \tag{10.4.7}$$

Pontryagin's maximum principle states that, if $(\mathbf{x}_*, \mathbf{u}_*)$ is optimal then, for all $t \leq \tau$, there are *adjoint paths* (λ) and μ such that

1. $H(t, \mathbf{x}_*, \mathbf{u}, \lambda) + \mu$ has a maximum value of 0, achieved at $\mathbf{u} = \mathbf{u}_*$
2. $\dfrac{d\lambda^{\mathrm{T}}}{dt} = -\lambda^{\mathrm{T}} \nabla \mathbf{b}(t, \mathbf{x}_*, \mathbf{u}_*) + \nabla \mathbf{c}(t, \mathbf{x}_*, \mathbf{u}_*)$
3. $\dfrac{d\mu}{dt} = -\lambda^{\mathrm{T}} \dfrac{d\, \mathbf{b}(t, \mathbf{x}_*, \mathbf{u}_*)}{dt} + \dfrac{d\, c(t, \mathbf{x}_*, \mathbf{u}_*)}{dt}$
4. $\dfrac{d\mathbf{x}_*}{dt} = \mathbf{b}(t, \mathbf{x}_*, \mathbf{u}_*)$
5. *Transversality conditions*. Either ("time-constrained" means τ is fixed):
   (a) time un-constrained: $\mu + \dfrac{dC}{dt} = 0$
   (b) time-constrained: $\left(\lambda^{\mathrm{T}} + \nabla C(\tau, \mathbf{x}(\tau))\right) \sigma = 0$ for all $\sigma \in \Sigma$

Note that if b, c, and C do not depend explicitly on t, then $\mu = 0$.

The first four statements can be written in terms of the Hamiltonian as

$$\mathbf{0} = \frac{\partial H}{\partial \mathbf{u}}, \qquad \frac{d\lambda}{dt} = -\frac{\partial H}{\partial \mathbf{x}}, \qquad \frac{d\mu}{dt} = -\frac{\partial H}{\partial t}, \qquad \frac{d\mathbf{x}}{dt} = \frac{\partial H}{\partial \lambda}. \tag{10.4.8}$$

Example Consider a car on a straight road with initial position p_0 and velocity q_0. The goal is to bring the car to rest, at position 0, in minimal time. The control is the acceleration u with $|u| \leq 1$ (i.e., $u = 1$ is full throttle and $u = -1$ is full reverse). The dynamics are:

$$\frac{d}{dt}\begin{bmatrix} p \\ q \end{bmatrix} = \frac{d\mathbf{x}}{dt} = \mathbf{b}(t, \mathbf{x}, \mathbf{u}) = \begin{bmatrix} q \\ u \end{bmatrix} \qquad \mathbf{x}(0) = \begin{bmatrix} p_0 \\ q_0 \end{bmatrix}, \quad \mathbf{x}(T) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

The goal is to minimize $\int_0^T 1\, dt$; where T is the first time for which $\mathbf{x}(T) = (0, 0)$; hence $c = 1$ and $C = 0$. Since time un-constrained, μ need not be considered.

The Hamiltonian is $H(t, \mathbf{x}, \mathbf{u}, \lambda) = \lambda^{\mathrm{T}} \mathbf{b} - c = \lambda_1 q + \lambda_2 u - 1$.

1. To maximize H, u must be an extremal value. Since $|u| \leq 1$, we have $u_* = \operatorname{sgn}(\lambda_2)$. Hence, $H = \lambda_1 q + |\lambda_2| - 1$.
2. When the car is stopped at $t = T$, we have $q(T) = 0$; hence (since $H = 0$ for the optimal control) we have $\lambda_2(T) = \pm 1$.
3. The adjoint equations are: $\dfrac{d\lambda_1}{dt} = -\dfrac{\partial H}{\partial p} = 0$ and $\dfrac{d\lambda_2}{dt} = -\dfrac{\partial H}{\partial q} = -\lambda_1$. So $\lambda_1(t) = \lambda_1$ is a constant and $\lambda_2(t) = \lambda_2(0) - \lambda_1 t$.
4. Note that $\dfrac{dp}{dt} = \dfrac{\partial H}{\partial \lambda_1} = q$ and $\dfrac{dq}{dt} = \dfrac{\partial H}{\partial \lambda_2} = u = \operatorname{sgn}(\lambda_2)$.
5. There are now cases to consider:
   (a) If $\lambda_2(T) = 1$ and $\lambda_1 \geq 0$, then $\lambda_2(t) \geq 0$. In this case $q(t) = t$ and $p(t) = \frac{t^2}{2} = \frac{q^2}{2}$.
   (b) If $\lambda_2(T) = 1$ and $\lambda_1 < 0$, then $\lambda_2(t) \geq 0$ (and $p = \frac{q^2}{2}$) only for $t \leq t_0 = \frac{1}{|\lambda_1|}$. For $t > t_0$ we have $\lambda_2(t) < 0$, so $u = -1$, $q(t) = t - 2t_0$ and $p(t) = 2tt_0 - \frac{t^2}{2} - t_0^2$.

There is a similar analysis for $\lambda_2(T) = -1$. The result is that there is a *switching locus* given by $p = -\operatorname{sgn}(q)\frac{q^2}{2}$. An initial state (p_0, q_0) that is above the locus lies on a parabola with $p = -\frac{q^2}{2} + d$ with $d > 0$. The optimal control is to initially move around the parabola until the switching locus is reached. Then the acceleration changes sign and the car is brought to rest at the origin by moving along the locus.

10.5 COMPUTER LANGUAGES

Common computer languages used by scientists and engineers.

1. Freely available
   - Numerical C++ www.gnu.org/software/commoncpp
   - Numerical Octave www.gnu.org/software/octave
   - Numerical SciPy www.scipy.org/
   - Statistical R www.r-project.org
   - Symbolic Maxima maxima.sourceforge.net
   - Symbolic Sage www.sagemath.org
2. Commercial
   - Numerical MathCad www.ptc.com/engineering-math-software/mathcad
   - Numerical MATLAB www.mathworks.com
   - Optimization AMPL www.ampl.com
   - Statistical Minitab www.minitab.com
   - Statistical SPSS www.ibm.com/analytics/us/en/technology/spss/
   - Symbolic Maple www.maplesoft.com
   - Symbolic Mathematica www.wolfram.com

10.6 COMPRESSIVE SENSING

Compressive sensing is a way to combine compression with sensing. In this context, compression means dimension reduction. Define the following:

1. A vector $\mathbf{x} \in C^n$ is k-sparse if only k elements are non-zero and $k < n$.
2. $\mathbf{x}$ is "compressible" if only a small number of its elements are significantly non-zero.

In many signal processing applications an n-dimensional signal $\mathbf{x} \in C^n$ is to be found from the linear system

$$A\mathbf{x} = \mathbf{b} \tag{10.6.1}$$

where the measurement matrix $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_n \end{bmatrix} \in C^{m \times n}$ and the m-dimensional measurement vector $\mathbf{b} \in C^m$ are known. Ignoring degenerate cases:

1. If $m = n$ then this is a square system; there is a unique solution to Equation (10.6.1). We have the solution $\hat{\mathbf{x}} = A^{-1}\mathbf{b}$.
2. If $m > n$ then this is an overdetermined system; there is no solution to Equation (10.6.1). We sometimes use $\hat{\mathbf{x}} = \left(A^{\mathrm{H}}A\right)^{-1} A^{\mathrm{H}}\mathbf{b}$ which minimizes the residual $||A\mathbf{x} - \mathbf{b}||_2$ (i.e., minimal least squares).
3. If $m < n$ then this is an underdetermined system; there are infinitely many solutions to Equation (10.6.1). We sometimes use $\hat{\mathbf{x}} = A^{+}\mathbf{b}$, where A^{+} is the pseudo-inverse, which minimizes $||\mathbf{x}||_2$.

Compressive sensing is used to solve Equation (10.6.1) when the data is vastly undersampled ($m \ll n$) and the solution $\mathbf{x}$ is compressible ($k \ll m$). One way to find the sparsest solution $\mathbf{x}$, the one with the least number of non-zero terms, is to solve

$$\min \left(||\mathbf{x}||_0 \mid A\mathbf{x} = \mathbf{b},\ \mathbf{x} \in C^n \right) \tag{10.6.2}$$

where $||\mathbf{x}||_0$ is the number of nonzero terms of $\mathbf{x}$. Unfortunately, the solution to (10.6.2) is generally not unique and solving (10.6.2) is NP-hard (i.e., computationally difficult). So, instead, consider the new problem

$$\min \left(||\mathbf{x}||_1 \mid A\mathbf{x} = \mathbf{b},\ \mathbf{x} \in C^n \right) \tag{10.6.3}$$

This problem can be solved using linear programming

$$\begin{aligned} \min \quad & \mathbf{1}^{\mathrm{T}}\mathbf{x} \\ \text{with} \quad & A\mathbf{z} = \mathbf{b} \\ & -\mathbf{x} \leq \mathbf{z} \leq \mathbf{x} \end{aligned} \tag{10.6.4}$$

There are theorems indicating that, often, the solution to (10.6.3) works:

1. Suppose the problem has been scaled so that $||\mathbf{a}_i||_2 = 1$. Define $M(A) = \max_{i\neq j} \left|\mathbf{a}_i^{\mathrm{T}}\mathbf{a}_j\right|$. If $\mathbf{x}^*$ is the sparsest solution of (10.6.2) and

$$||\mathbf{x}^*||_0 < \frac{\sqrt{2}-\frac{1}{2}}{M(A)} \tag{10.6.5}$$

then the solution of (10.6.3) is equal to the solution of (10.6.2).

2. Let δ_k be the smallest number such that

$$(1-\delta_k)\,||\mathbf{x}||_2^2 \leq ||A\mathbf{x}||_2^2 \leq (1+\delta_k)\,||\mathbf{x}||_2^2$$

for all k-sparse vectors $\mathbf{x}$. If $\delta_{2k} < \sqrt{2}-1$ for all k-sparse vectors $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{b}$ then the solution of (10.6.3) is equal to the solution of (10.6.2).

Notes

- Compressive sensing often works since many natural signals are sparse in the sense that they have concise representations when expressed in the proper basis.
- Often the observations $\mathbf{b}$ are noisy. In this case the optimization problem in (10.6.3) may be modified to be

$$\min\left(\lambda\,||\mathbf{x}||_1 + ||A\mathbf{x}-\mathbf{b}||_2^2\right) \tag{10.6.6}$$

which contains an L_1 term promoting sparsity and an L_2 term promoting a better fit to the data. The λ represents the tradeoff between the two.

10.7 CONSTRAINED LEAST SQUARES

The solution to the following constrained least squares problem

$$\min_{\mathbf{x}} ||A\mathbf{x}-\mathbf{b}||^2 \qquad \text{subject to } C\mathbf{x} = \mathbf{d} \tag{10.7.1}$$

is given by (assuming non-degeneracy)

$$\mathbf{x} = \left(A^{\mathrm{T}}A\right)^{-1}\left(A^{\mathrm{T}}\mathbf{b} - C^{\mathrm{T}}\left(C\left(A^{\mathrm{T}}A\right)^{-1}C^{\mathrm{T}}\right)^{-1}\left(C\left(A^{\mathrm{T}}A\right)^{-1}A^{\mathrm{T}}\mathbf{b}-\mathbf{d}\right)\right) \tag{10.7.2}$$

Special cases include:

1. Using $\mathbf{d} = \mathbf{0}$ and $A = I$ in (10.7.1):

$$\begin{array}{rl} \text{the solution of} & \min ||\mathbf{x}-\mathbf{b}||^2 \qquad \text{subject to } C\mathbf{x} = \mathbf{0} \\ \text{is} & \mathbf{x} = \mathbf{b} - C^{\mathrm{T}}\left(CC^{\mathrm{T}}\right)^{-1}C\mathbf{b} \end{array}$$

2. Using $\mathbf{b} = \mathbf{0}$ in (10.7.1):

$$\begin{array}{rl} \text{the solution of} & \min ||A\mathbf{x}||^2 \qquad \text{subject to } C\mathbf{x} = \mathbf{d} \\ \text{is} & \mathbf{x} = \left(A^{\mathrm{T}}A\right)^{-1}C^{\mathrm{T}}\left(C\left(A^{\mathrm{T}}A\right)^{-1}C^{\mathrm{T}}\right)^{-1}\mathbf{d} \end{array}$$

3. Using $\mathbf{b} = \mathbf{0}$ and $A = I$ in (10.7.1):

$$\begin{array}{rl} \text{the solution of} & \min ||\mathbf{x}||^2 \qquad \text{subject to } C\mathbf{x} = \mathbf{d} \\ \text{is} & \mathbf{x} = C^{\mathrm{T}}\left(CC^{\mathrm{T}}\right)^{-1}\mathbf{d} \end{array}$$

10.8 CRYPTOGRAPHY

Cryptography is the practice and study of hiding information. A *cipher* is a pair of algorithms for encrypting and decrypting the message.

Suppose Alice wants to send a message (the *plaintext*) M to Bob. Using a *key* Alice encrypts M to create the *ciphertext* E, and sends E to Bob. Bob converts E back to M, also using a *key*. The goal is to make it difficult to convert E to M if some information is unknown. Common cipher classes include:

- Secret key cryptography uses one key for encryption and decryption. Examples include:
  - Data Encryption Standard (DES) is a block-cipher that is out of use. Triple-DES with 168-bit keys is currently used.
  - Advanced Encryption Standard (AES) uses a block-cipher with 128-bit keys and blocks.
- Public key cryptography uses one key for encryption and another key for decryption. These depend upon *one-way functions*; mathematical functions that are easy to compute but difficult to invert. (The existence of one-way functions has never been proven. If P = NP, then they do not exist.)
  For example, multiplying two 1,000-digit numbers is easy, factoring a 2,000-digit number to obtain its 1,000-digit factors is hard. Also, given two number a and b it is easy to compute a^b. Given a number N that is of this form, it is difficult to determine a and b.
  Examples include:
  - Elliptic curve cryptography (ECC); see Section 10.10. ECC encryption exploits the difficulty of the logarithm problem. ECC methods provide security equivalent to RSA while using fewer bits.
  - RSA encryption (named after Rivest, Shamir, and Adleman) uses a variable size encryption block and a variable size key. RSA encryption exploits the difficulty of factoring large numbers.

The RSA process is:

1. **Initialization**
   (a) Bob randomly selects two large primes p and q, with $p \neq q$.
   (b) The values $n = pq$ and $\phi \equiv \phi(n) = (p-1)(q-1)$ are computed, where ϕ is Euler's totient function.
   (c) Bob selects an integer e with $1 < e < \phi$ and $\text{GCD}(e, \phi) = 1$.
   (d) Bob computes the integer d such that $1 < d < \phi$ and $ed \equiv 1 \pmod{\phi}$. (This can be achieved with the Euclidean algorithm.)
   (e) Bob publishes the values (n, e) and keeps the value of d secret.
2. **Use**
   (a) Alice wants to send the plaintext number M to Bob.
   (b) Alice computes the ciphertext $E = M^e \pmod{n}$ and sends it to Bob.
   (c) Bob accepts E and computes $E^d \pmod{n} = M^{de} \pmod{n} = M \pmod{n}$; thus recovering the plaintext message.

10.9 DISCRETE DYNAMICAL SYSTEMS AND CHAOS

A dynamical system described by a function $f : M \to M$ is *chaotic* if

1. f is *transitive*—that is, for any pair of non-empty open sets U and V in M there exists a positive constant k such that $f^k(U) \cap V$ is not empty (here $f^k = \underbrace{f \circ f \circ \cdots \circ f}_{k \text{ times}}$); and
2. The *periodic points* (the points that map to themselves after a finite number of iterations) of f are dense in M; and
3. f has a sensitive dependence on initial conditions—that is, there is a positive number δ (depending only on f and M) such that in every non-empty open subset of M there is a pair of points whose eventual iterates under f are separated by a distance of at least δ.

Some systems depend on a parameter and become chaotic for some values of that parameter. There are various routes to chaos, one of them is via period doubling bifurcations. Let the distance between successive bifurcations of a process be d_k. The limiting ratio $\delta = \lim_{k\to\infty} d_k/d_{k+1}$ is constant in many situations and is equal to Feigenbaum's constant $\delta \approx 4.6692016091029$.

Some chaotic one-dimensional maps are:

1. Logistic map: $x_{n+1} = 4x_n(1 - x_n)$ with $x_0 \in [0, 1]$.
   Solution is $x_n = \frac{1}{2} - \frac{1}{2}\cos[2^n \cos^{-1}(1 - 2x_0)]$.
2. Tent map: $x_{n+1} = 1 - 2\left|x_n - \frac{1}{2}\right|$ with $x_0 \in [0, 1]$.
   Solution is $x_n = \frac{1}{\pi}\cos^{-1}[\cos(2^n \pi x_0)]$.
3. Baker transformation: $x_{n+1} = 2x_n \pmod 1$ with $x_0 \in [0, 1]$.
   Solution is $x_n = \frac{1}{\pi}\cot^{-1}[\cot(2^n \pi x_0)]$.

Chaotic differential equations include: $\ddot{x} + a\dot{x} - \dot{x}^2 + x = 0$ for $2.0168 < a < 2.0577$ and $\ddot{x} + x^3 = \sin \Omega t$ for most of the range $0 < \Omega < 2.8$.

10.9.1 ERGODIC HIERARCHY

The ergodic hierarchy is a classification of dynamical systems:

$$\text{Ergodic} \supset \text{Weak Mixing} \supset \text{Strong Mixing} \supset \text{Kolmogorov} \supset \text{Bernoulli} \quad (10.9.1)$$

with a precise technical definition for each of the five levels. The higher up levels are "more random," with a Bernoulli System being "completely random." It may be that Strong Mixing is a necessary condition for a system to be chaotic, and being a Kolmogorov System is a sufficient condition for a system to be chaotic.

10.9.2 JULIA SETS AND THE MANDELBROT SET

Consider the complex points $\{z_n\}$ formed by iterating the map $z_{n+1} = z_n^2 + c$ with $z_0 = z$. For each z, either the iterates remain bounded (z is in the *prisoner set*) or they escape to infinity (z is in the *escape set*). The *Julia set* J_c is the boundary between these two sets. Using lighter colors to indicate a "faster" escape to infinity, Figure 10.1 shows two Julia sets. One of these Julia sets is connected, the other is disconnected. The Mandelbrot set, M, is the set of those complex values c for which J_c is a connected set (see Figure 10.2). Alternately, the *Mandelbrot* set consists of all points c for which the discrete dynamical system, $z_{n+1} = z_n^2 + c$ with $z_0 = 0$, converges.

The boundary of the Mandelbrot set is a *fractal*. There is no universal agreement on the definition of "fractal." One definition is that it is a set whose fractal dimension differs from its topological dimension.

FIGURE 10.1
Connected Julia set for $c = -0.5i$ *(left). Disconnected Julia set for* $c = -\frac{3}{4}(1+i)$ *(right). (Julia sets are the black objects.)*

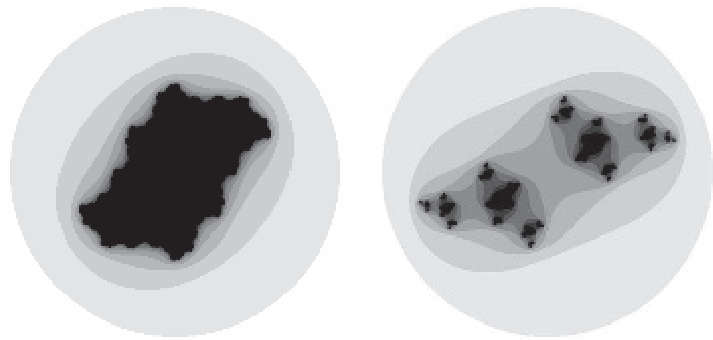
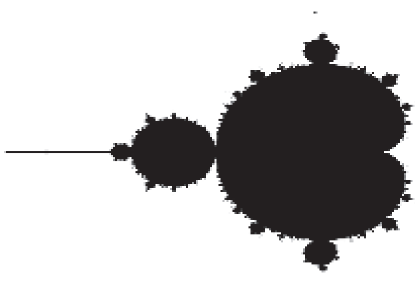


FIGURE 10.2
The Mandelbrot set. The leftmost point has the coordinates $(-2, 0)$.

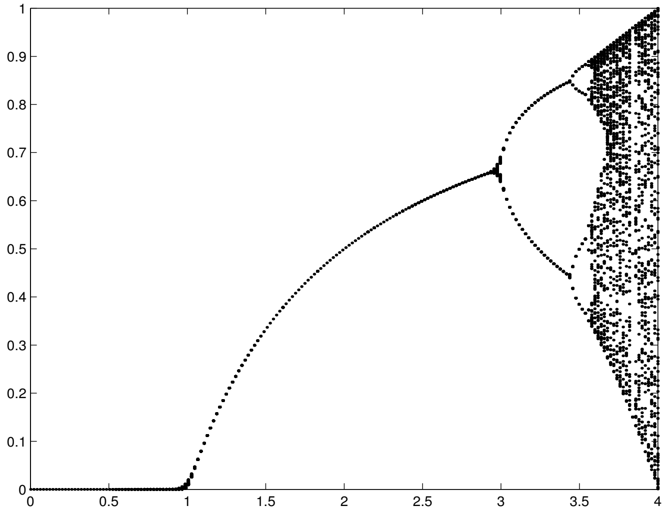


10.9.3 LOGISTIC MAP

Consider $u_{n+1} = f(u_n) = au_n(1 - u_n)$ with $a \in [0, 4]$. Note that if $0 \leq u_0 \leq 1$ then $0 \leq u_n \leq 1$. The fixed points satisfy $u = f(u) = au(1 - u)$; they are $u = 0$ and $u = (a - 1)/a$.

1. If $a = 0$ then $u_n = 0$.
2. If $0 < a \leq 1$ then $u_n \to 0$.
3. If $1 < a < 3$ then $u_n \to (a - 1)/a$.
4. If $3 \leq a < 3.449490\ldots$ then u_n oscillates between the two roots of $u = f(f(u))$ which are not roots of $u = f(u)$, that is, $u_\pm = (a + 1 \pm \sqrt{a^2 - 2a - 3})/2a$.

The location of the final state is summarized by the following diagram (the horizontal axis is the a value).



Minimal values of a at which a cycle with a given number of points appears:

n	2^n points in a cycle	minimal a
1	2	3
2	4	3.449490...
3	8	3.544090...
4	16	3.564407...
5	32	3.568750...
6	64	3.56969...
7	128	3.56989...
8	256	3.569934...
9	512	3.569943...
10	1024	3.5699451...
11	2048	3.569945557...

10.10 ELLIPTIC CURVES

Consider the general cubic curve $Ax^3 + Bx^2y + Cxy^2 + Dy^3 + Ex^2 + Fxy + Gy^2 + Hx + Iy + J = 0$ over a field F. By an appropriate change of variables the curve can be simplified:

- If F has characteristic 2, to $y^2 + \alpha y = x^3 + \beta x^2 + cxy + dx + e$. A commonly considered special case is (here $b \neq 0$)

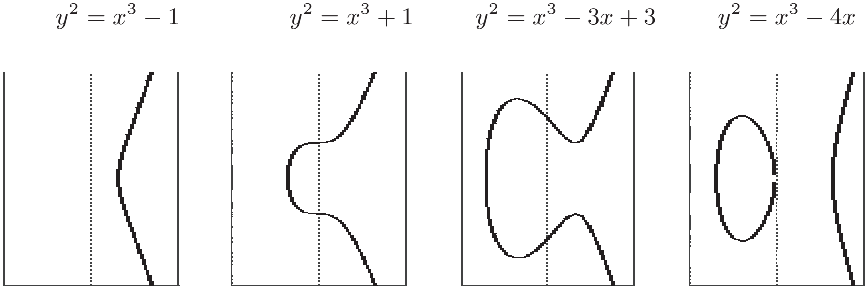
$$y^2 + xy = x^3 + ax^2 + b \tag{10.10.1}$$

- If F has characteristic 3, to $y^2 = x^3 + ax^2 + bx + c$.
- If F has characteristic not equal to 2 or 3 (e.g., F is $\mathbb{R}$ or $\mathbb{C}$) to

$$y^2 = x^3 + ax + b \tag{10.10.2}$$

The graph of Equation (10.10.2) is symmetric about the x-axis, see Figure 10.3 for different values of a and b.

FIGURE 10.3
The elliptic curve $y^2 = x^3 + ax + b$ for different values of a and b.



A curve $F(x, y) = 0$ is *singular* if there are any points with $\frac{\partial F}{\partial x} = 0 = \frac{\partial F}{\partial y}$. Equation (10.10.2) will be non-singular if the curve's *discriminant*, $-16(4a^3+27b^2)$, is non-zero. In this case the curve does not have repeated roots.

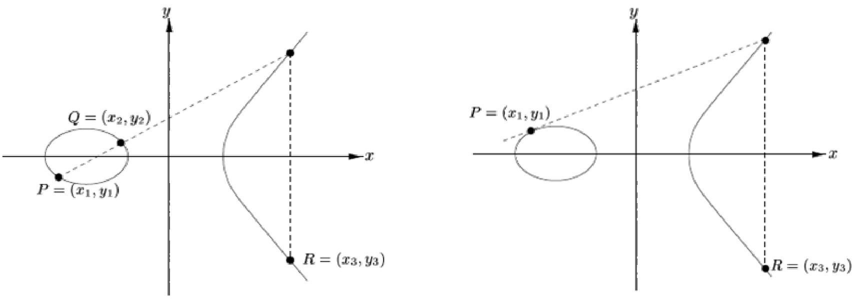
The points on an elliptic curve can form an Abelian group if the curve is non-singular. There also needs to be a *point at infinity* indicated by ∞; it is the group identity. The group's binary operation is called addition, denoted "+." Two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ can be added, $P_1 + P_2 = P_3 = (x_3, y_3)$, as follows:

- If $P_1 = \infty$ then $P_3 = \infty + P_2 = P_2$.
- If $P_1 = -P_2$ then $P_3 = \infty$.
- Otherwise, for Equation (10.10.2)
  - If $x_1 \neq x_2$ then $\lambda = \frac{y_1 - y_2}{x_1 - x_2}$, $x_3 = \lambda^2 - x_1 - x_2$, and $y_3 = \lambda(x_1 - x_3) - y_1$.
  - If $P_1 = P_2$ and $y_1 = 0$ then $P_3 = \infty$.
  - If $P_1 = P_2$ and $y_1 \neq 0$ then $\lambda = \frac{3x_1^2 + a}{2y_1}$, $x_3 = \lambda^2 - 2x_1$, and $y_3 = \lambda(x_1 - x_3) - y_1$.

- Otherwise, for Equation (10.10.1)
  - If $x_1 \neq x_2$ then $\lambda = \frac{y_1+y_2}{x_1+x_2}$, $x_3 = \lambda^2 + \lambda + x_1 + x_2 + a$, and $y_3 = \lambda(x_1 + x_3) + x_3 + y_1$.
  - If $P_1 = P_2$ then $\lambda = x_1 + \frac{y_1}{x_1}$, $x_3 = \lambda^2 + \lambda + a$, and $y_3 = x_1^2 + (\lambda + 1)x_3$.

Addition can be understood geometrically: a line is drawn through P_1 and P_2. Either this line does not intersect the curve and the result is ∞, or it does intersect the curve. If the line does intersect the curve, then reflect the point of intersection about the x-axis (i.e., change the sign of the y term); this results in P_3. See Figure 10.4.

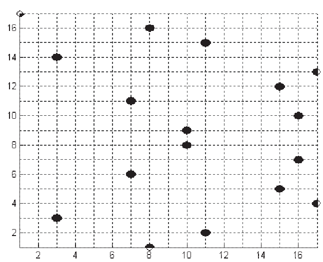
FIGURE 10.4
Addition of two points on an elliptic curve. Distinct points (left, $R = P + Q$) and the same point (right, $R = P + P = 2P$).



Now assume that the field F is not the real numbers (as implicitly assumed above) but of finite order; for example, F_p for a prime p. If the determinant of the curve is non-zero (modulo p) then the points form a group, as before. (The computation of λ in the addition formulas is then taken modulo p.)

EXAMPLE

Consider the points on the curve $y^2 = x^3 - 1$ over the field F_{17}. There are 17 such points: $\{\ldots, (3, 3), (3, 14), (7, 6), (7, 11), (8, 1), (8, 16), \ldots\}$. Note that the points are symmetric about the line $y = 8.5$.



EXAMPLE

Use the field F_{2^m} with 2^m elements and use the elliptic curve in Equation (10.10.1). For example, consider $y^2 + xy = x^3 + g^4x^2 + 1$ over the field F_{2^4}, where g is a multiplicative generator of F_{2^4} satisfying $g^4 = g + 1$. The 15 points on this elliptic curve are: $(1, g^6)$, $(1, g^{13})$, (g^3, g^{13}), (g^3, g^8), (g^5, g^{11}), (g^5, g^3), (g^6, g^{14}), (g^6, g^8), (g^9, g^{10}), (g^9, g^{13}), (g^{10}, g^8), (g^{10}, g), (g^{12}, g^{12}), (g^{12}, ∞), and $(\infty, 1)$.

Scalar "multiplication" of an integer k by the elliptic curve point P, to obtain a new elliptic curve point, is achieved by the "double and add" algorithm. That is: $2P = P + P$, $3P = (2P) + P$, $4P = (2P) + (2P)$, $5P = (4P) + P$, etc. The *discrete logarithm for elliptic curves* is to determine k when given P and kP. Elliptic curve cryptography (ECC) schemes exploit the difficulty of the logarithm problem.

- **Key exchange** is a process by which Alice and Bob secretly determine a key that others cannot determine. The process is:

  1. Alice and Bob agree on an elliptic curve to use, a finite field F, an initial point P_0, and a process for creating a secret key from an elliptic curve point (e.g., given $P = (x, y)$ the key could be x).
  2. Alice selects a secret number A, creates $P_A = AP_0$, sends P_A to Bob.
  3. Bob selects a secret number B, creates $P_B = BP_0$, sends P_B to Alice.
  4. Alice accepts P_B and computes $P_{AB} = AP_B = ABP_0$.
  5. Bob accepts P_A and computes $P_{AB} = BP_A = ABP_0$.
  6. Alice and Bob create the same secret key from P_{AB}.

  Due to the difficulty of the logarithm problem, an observer who knows the elliptic curve used and $\{F, P_0, P_A, P_B\}$ will not be able to easily determine P_{AB} and the resulting secret key.

  The key exchange algorithm above is the basic Diffie–Hellman key exchange done over an Elliptic Curve group. The more common approach is to use Diffie–Hellman key exchange over a multiplicative group. In this method two numbers are made public: p (a prime) and g (a primitive root modulo p). Alice selects a secret number A, creates $\alpha = g^A \pmod{p}$, and sends this to Bob. Bob selects a secret number B, creates $\beta = g^B \pmod{p}$, and sends this to Alice. The secret key is then K, computed by Alice as $K = \beta^A \pmod{p} = g^{AB} \pmod{p}$ and computed by Bob as $K = \alpha^B \pmod{p} = g^{AB} \pmod{p}$.
- **Secure communication** (El-Gamal process)
  Alice and Bob agree on an elliptic curve to use, a finite field F, an initial point P_0, and a process for converting between a number and an elliptic curve point (e.g., given $P = (x, y)$ the value could be x).

  1. **Initialization**
     Bob selects a secret number B, creates $P_B = BP_0$, sends P_B to Alice. This can be used for many messages.
  2. **Encryption**
     Alice wants to send a message z, which maps to the point P_z on the elliptic curve, to Bob. Alice selects a secret number A and creates the pair of values $(z_1, z_2) = (AP_0, P_z + AP_B)$ and sends them to Bob.
  3. **Decryption**
     Bob accepts the pair of values (z_1, z_2) and determines the point sent via $z_2 - Bz_1 = (P_z + AP_B) - B(AP_0) = P_z + (ABP_0 - BAP_0) = P_z$. From P_z the message z can be determined.

10.11 FINANCIAL FORMULAS

10.11.1 BORROWING AND LENDING

Notation

A	amount that P is worth, after n time periods, with i percent interest per period
B	total amount borrowed
P	principal to be invested (equivalently, present value)
a	future value multiplier after one time period
i	percent interest per time period (expressed as a decimal)
m	amount to be paid each time period
n	number of time periods

Note that the units of A, B, P, and m must all be the same, for example, dollars.

10.11.1.1 Formulas connecting financial terms

1. **Interest**: Let the principal amount P be invested at an interest rate of $i\%$ per time period (expressed as a decimal), for n time periods. Let A be the amount that this is worth after n time periods. Then

   (a) *Simple interest*:

$$A = P(1+ni) \quad \text{and} \quad P = \frac{A}{(1+ni)} \quad \text{and} \quad i = \frac{1}{n}\left(\frac{A}{P} - 1\right). \tag{10.11.1}$$

   (b) *Compound interest*

$$A = P(1+i)^n \quad \text{and} \quad P = \frac{A}{(1+i)^n} \quad \text{and} \quad i = \left(\frac{A}{P}\right)^{1/n} - 1. \tag{10.11.2}$$

   When interest is compounded q times per time period for n time periods, it is equivalent to an interest rate of $(i/q)\%$ per time period for nq time periods.

$$\begin{aligned} A &= P\left(1+\frac{i}{q}\right)^{nq}, \\ P &= A\left(1+\frac{i}{q}\right)^{-nq}, \\ i &= q\left[\left(\frac{A}{P}\right)^{1/nq} - 1\right]. \end{aligned} \tag{10.11.3}$$

   Continuous compounding occurs when the interest is compounded infinitely often in each time period (i.e., $q \to \infty$). In this case: $A = Pe^{in}$.

2. **Present value**: If A is to be received after n time periods of $i\%$ interest per time period, then the present value P of such an investment is given by (from Equation (10.11.2)) $P = A(1+i)^{-n}$.

3. **Annuities**: Suppose that the amount B (in dollars) is borrowed, at a rate of $i\%$ per time period, to be repaid at a rate of m (in dollars) per time period, for a total of n time periods.

$$m = Bi\frac{(1+i)^n}{(1+i)^n - 1}, \tag{10.11.4}$$

$$B = \frac{m}{i}\left(1 - \frac{1}{(1+i)^n}\right). \tag{10.11.5}$$

Using $a = (1+i)$, these equations can be written more compactly as

$$m = Bi\frac{a^n}{a^n - 1} \qquad \text{and} \qquad B = \frac{m}{i}\left(1 - \frac{1}{a^n}\right). \tag{10.11.6}$$

10.11.1.2 Examples

1. **Question**: If \$100 is invested at 5% per year, compounded annually for 10 years, what is the resulting amount?
   - **Analysis**: Using Equation (10.11.2), we identify
     (a) Principal invested, $P = 100$ (the units are dollars)
     (b) Time period, 1 year
     (c) Interest rate per time period, $i = 5\% = 0.05$
     (d) Number of time periods, $n = 10$
   - **Answer**: $A = P(1+i)^n$ or $A = 100(1+0.05)^{10} = \$162.89$.

2. **Question**: If \$100 is invested at 5% per year and the interest is compounded quarterly (4 times a year) for 10 years, what is the final amount?
   - **Analysis**: Using Equation (10.11.3) we identify
     (a) Principal invested, $P = 100$ (the units are dollars)
     (b) Time period, 1 year
     (c) Interest rate per time period, $i = 5\% = 0.05$
     (d) Number of time periods, $n = 10$
     (e) Number of compounding time periods, $q = 4$
   - **Answer**: $A = P\left(1 + \frac{i}{q}\right)^{nq}$ or $A = 100(1 + \frac{0.05}{4})^{4\cdot 10} = 100(1.0125)^{40}$ $= \$164.36$.
   - **Alternate analysis**: Using Equation (10.11.2), we identify
     (a) Principal invested, $P = 100$ (the units are dollars)
     (b) Time period, quarter of a year
     (c) Interest rate per time period, $i = \frac{5\%}{4} = \frac{0.05}{4} = 0.0125$
     (d) Number of time periods, $n = 10 \cdot 4 = 40$
   - **Alternate answer**: $A = P(1+i)^n$ or $A = 100(1.0125)^{40} = \$164.36$.

3. **Question**: If \$100 is invested now, and we wish to have \$200 at the end of 10 years, what yearly compound interest rate must we receive?
   - **Analysis**: Using Equation (10.11.2), we identify
     (a) Principle invested, $P = 100$ (the units are dollars)
     (b) Final amount, $A = 200$
     (c) Time period, 1 year
     (d) Number of time periods, $n = 10$
   - **Answer**: $i = \left(\frac{A}{P}\right)^{1/n} - 1$ or $i = \left(\frac{200}{100}\right)^{1/10} - 1 = 0.0718$. Hence, we must receive an annual interest rate of 7.2%.

4. **Question**: An investment returns \$10,000 in 10 years. If the interest rate will be 10% per year, what is the present value? (That is, how much money would have to be invested now to obtain this amount in 10 years?)
   - **Analysis**: Using Equation (10.11.2), we identify
     (a) Final amount, $A =$10,000 (the units are dollars)
     (b) Time period, 10 years
     (c) Interest rate per time period, $i = 10\% = 0.1$
     (d) Number of time periods, $n = 10$
   - **Answer**: $P = A(1 + i)^{-n} = 10{,}000(1.1)^{-10} = 3855.43$; the present value of this investment is \$3,855.43.

5. **Question**: A mortgage of \$100,000 is obtained with which to buy a house. The mortgage will be repaid at an interest rate of 6% per year, compounded monthly, for 30 years. What is the monthly payment?
   - **Analysis**: Using Equation (10.11.6), we identify
     (a) Amount borrowed, $B = 100,000$ (the units are dollars)
     (b) Time period, 1 month
     (c) Interest rate per time period, $i = 0.06/12 = 0.005$
     (d) Number of time periods, $n = 30 \cdot 12 = 360$
   - **Answer**: $a = 1 + i = 1.005$ and $m = Bi\frac{a^n}{a^n - 1} = (100,000)(.005)$ $\times\frac{(1.005)^{360}}{(1.005)^{360} - 1} = 599.55$. The monthly payment is \$599.55.

6. **Question**: Suppose that interest rates on 15-year mortgages are currently 6%, compounded monthly. By spending \$800 per month, what is the largest mortgage obtainable?
   - **Analysis**: Using Equation (10.11.6), we identify
     (a) Time period, 1 month
     (b) Payment amount, $m = 800$ (the units are dollars)
     (c) Interest rate per time period, $i = 0.06/12 = 0.005$
     (d) Number of time periods, $n = 15 \cdot 12 = 180$
   - **Answer**:
     $a = 1 + i = 1.005$ and $B = m\left(1 - 1/a^n\right)/i = \frac{800}{0.005}\left(1 - \frac{1}{1.005^{180}}\right)$ $= 94802.81$. The largest mortgage amount obtainable is \$94,802.81.

10.11.2 OPTIONS

Notation

- S asset (spot) price
- V value of an option
- r risk free interest rate
- σ volatility of the asset
- τ time to asset maturity

First order effects

1. Delta $\Delta = \dfrac{\partial V}{\partial S}$ = change in value with respect to price
2. Lambda $\lambda = \dfrac{\partial V}{\partial S} \times \dfrac{S}{V}$ = change in value (%) with respect to price (%)
3. Rho $\rho = \dfrac{\partial V}{\partial r}$ = change in value with respect to interest rate
4. Theta $\Theta = -\dfrac{\partial V}{\partial \tau}$ = change in value as time changes
5. Vega $\nu = \dfrac{\partial V}{\partial \sigma}$ = change in value with respect to volatility

Second order effects

1. Charm $-\dfrac{\partial \Delta}{\partial \tau} = -\dfrac{\partial \Theta}{\partial S} = -\dfrac{\partial^2 V}{\partial S\, \partial \tau}$
2. Gamma $\Gamma = \dfrac{\partial \Delta}{\partial S} = \dfrac{\partial^2 V}{\partial S^2} =$
3. Vanna $\dfrac{\partial \Delta}{\partial \sigma} = \dfrac{\partial \nu}{\partial S} = \dfrac{\partial^2 V}{\partial S\, \partial \sigma}$
4. Vera $\dfrac{\partial \rho}{\partial \sigma} = \dfrac{\partial^2 V}{\partial \sigma\, \partial r}$
5. Veta $\dfrac{\partial \nu}{\partial \tau} = \dfrac{\partial^2 V}{\partial \sigma\, \partial \tau}$
6. Vomma $\dfrac{\partial \nu}{\partial \sigma} = \dfrac{\partial^2 V}{\partial \sigma^2}$

Third order effects

1. Color $\dfrac{\partial \Gamma}{\partial \tau} = \dfrac{\partial^3 V}{\partial S^2\, \partial \tau}$
2. Speed $\dfrac{\partial \Gamma}{\partial S} = \dfrac{\partial^3 V}{\partial S^3}$
3. Ultima $\dfrac{\partial(\text{vomma})}{\partial \sigma} = \dfrac{\partial^3 V}{\partial \sigma^3}$
4. Zomma $\dfrac{\partial \Gamma}{\partial \sigma} = \dfrac{\partial(\text{vanna})}{\partial S} = \dfrac{\partial^3 V}{\partial S^2\, \partial \sigma}$

10.11.2.1 Values for a European option

Here

- $\phi(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the standard normal density function
- $\Phi(x) = \int_{-\infty}^{x} \phi(z)\, dz$ is the standard normal cumulative distribution function
- $d_1 = \frac{\ln(S/K)+(r-q+\sigma^2/2)\tau}{\sigma\sqrt{\tau}}$
- $d_2 = \frac{\ln(S/K)+(r-q-\sigma^2/2)\tau}{\sigma\sqrt{\tau}} = d_1 - \sigma\sqrt{\tau}$
- K is the strike price
- q is the annual dividend yield

Call values

value	$Se^{-q\tau}\Phi(d_1) - e^{-r\tau}K\Phi(d_2)$
First order effects	
delta	$e^{-q\tau}\Phi(d_1)$
rho	$K\tau e^{-r\tau}\Phi(d_2)$
theta	$-e^{-q\tau}\frac{S\phi(d_1)\sigma}{2\sqrt{\tau}} - rKe^{-r\tau}\Phi(d_2) + qSe^{-q\tau}\Phi(d_1)$
vega	$Se^{-q\tau}\phi(d_1)\sqrt{\tau} = Ke^{-r\tau}\phi(d_2)\sqrt{\tau}$
Second order effects	
charm	$qe^{-q\tau}\Phi(d_1) - e^{-q\tau}\phi(d_1)\frac{2(r-q)\tau - d_2\sigma\sqrt{\tau}}{2\tau\sigma\sqrt{\tau}}$
gamma	$e^{-q\tau}\frac{\phi(d_1)}{S\sigma\sqrt{\tau}} = Ke^{-r\tau}\frac{\phi(d_2)}{S^2\sigma\sqrt{\tau}}$
vanna	$-e^{-q\tau}\phi(d_1)\frac{d_2}{\sigma} = \frac{\nu}{S}\left[1 - \frac{d_1}{\sigma\sqrt{\tau}}\right]$
veta	$Se^{-q\tau}\phi(d_1)\sqrt{\tau}\left[q + \frac{(r-q)\,d_1}{\sigma\sqrt{\tau}} - \frac{1+d_1d_2}{2\tau}\right]$
vomma	$Se^{-q\tau}\phi(d_1)\sqrt{\tau}\frac{d_1d_2}{\sigma} = \nu\frac{d_1d_2}{\sigma}$
Third order effects	
color	$-e^{-q\tau}\frac{\phi(d_1)}{2S\tau\sigma\sqrt{\tau}}\left[2q\tau + 1 + \frac{2(r-q)\tau - d_2\sigma\sqrt{\tau}}{\sigma\sqrt{\tau}}d_1\right]$
speed	$-e^{-q\tau}\frac{\phi(d_1)}{S^2\sigma\sqrt{\tau}}\left(\frac{d_1}{\sigma\sqrt{\tau}} + 1\right) = -\frac{\Gamma}{S}\left(\frac{d_1}{\sigma\sqrt{\tau}} + 1\right)$
ultima	$-\frac{\nu}{\sigma^2}\left[d_1d_2(1-d_1d_2) + d_1^2 + d_2^2\right]$
zomma	$e^{-q\tau}\frac{\phi(d_1)\,(d_1d_2-1)}{S\sigma^2\sqrt{\tau}} = \Gamma\cdot\left(\frac{d_1d_2-1}{\sigma}\right)$

Put values

The Put values are the same as the Call values for: vega, gamma, vanna, veta, vomma, speed, zomma, color, and ultima.

value	$e^{-r\tau}K\Phi(-d_2) - Se^{-q\tau}\Phi(-d_1)$
First order effects	
delta	$-e^{-q\tau}\Phi(-d_1)$
rho	$-K\tau e^{-r\tau}\Phi(-d_2)$
theta	$-e^{-q\tau}\frac{S\phi(d_1)\sigma}{2\sqrt{\tau}} - rKe^{-r\tau}\Phi(d_2) + qSe^{-q\tau}\Phi(d_1)$
Second order effects	
charm	$-qe^{-q\tau}\Phi(-d_1) - e^{-q\tau}\phi(d_1)\frac{2(r-q)\tau - d_2\sigma\sqrt{\tau}}{2\tau\sigma\sqrt{\tau}}$

10.11.3 MODERN PORTFOLIO THEORY

There are assets $\{1, 2, \dots\}$; a collection of them will be purchased for a portfolio.

- Asset i has a random return A_i with mean r_i and variance σ_i^2.
- The covariance of the random returns, for assets i and j, is $\sigma_{ij} = \sigma_i \sigma_j \rho_{ij}$. Here ρ_{ij} is the correlation coefficient and $\rho_{ii} = 1$.

A portfolio contains w_i of asset i. Define the return vector $\mathbf{r} = \begin{bmatrix} r_i & r_2 & \dots \end{bmatrix}^{\mathrm{T}}$, the covariance matrix $\Sigma = (\sigma_{ij}^2)$, and the weight vector $\mathbf{w} = \begin{bmatrix} w_i & w_2 & \dots \end{bmatrix}^{\mathrm{T}}$.

1. The random portfolio return is $R_p = \sum_i w_i A_i$
2. The portfolio's expected return of $\mathrm{E}\left[R_p\right] = \sum_i w_i \mathrm{E}\left[A_i\right] = \sum_i w_i r_i = \mathbf{w}^{\mathrm{T}}\mathbf{r}$
3. The portfolio's variance is $\mathrm{Var}\left[R_p\right] = \sigma_p^2 = \sum_i \sum_j w_i w_j \sigma_{ij} = \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w}$
4. The portfolio has a volatility of $\sigma_p = \sqrt{\sigma_p^2}$.

Define the risk tolerance $q \geq 0$; if $q = 0$ then no risk is acceptable, if q is large then a large amount of risk is acceptable. The portfolio optimization problem is to solve the quadratic programming problem:

$$\begin{array}{lll} \text{minimize}_{\mathbf{w}} & \mathbf{w}^{\mathrm{T}}\Sigma\mathbf{w} - q\,\mathbf{w}^{\mathrm{T}}\mathbf{r} & \\ \text{subject to} & \displaystyle\sum_i w_i = 1 & \text{(the initial portfolio value is fixed)} \end{array}$$

If the solution has any $w_k < 0$ it means that asset k has been shorted. To avoid shorting assets, include the constraint $w_i \geq 0$ for $i = 1, 2, \dots$ in the problem statement.

10.11.4 MERTON'S PORTFOLIO PROBLEM

You need to determine how much of your wealth to consume today, and how much to keep for investment. The goal is to maximize the value of the your utility, out to infinite time. Assume

1. $w(t)$ is your wealth at time t; which evolves stochastically
2. $c(t)$ is the amount consumed at time t (where $c(t) \leq w(t)$)
3. $u(c)$ is your utility function; when consuming c
4. ρ is your subjective discount rate (consuming now is better than later)
5. (μ, σ^2) are the mean and variance for stocks with a random return
6. $\pi(t)$ is the fraction of your wealth in stocks; the rest is in risk-free assets
7. r is the return for a risk-free investment

The problem statement is to maximize $\mathrm{E}\left[\int_0^\infty e^{-\rho t} u(c(t))\, dt\right]$. If your utility function is $u(x) = \dfrac{x^{1-\gamma}}{1-\gamma}$, which represents constant relative risk aversion parametrized by γ, then $\pi(t) = \dfrac{\mu - r}{\sigma^2 \gamma}$, a constant.

10.11.5 DOUBLING TIME - THE RULE OF 72

The "rule of 72" is an approximation for how long it takes the value of money to double. The computation is simple: divide the yearly interest rate into the number 72, the result is a number of years needed. The value 72 is used since it has many small divisors (e.g., 2,3,4,6,8); the value 69 would be more accurate.

EXAMPLE Investing $100 at an annual interest rate is 9%, it will take $\frac{72}{9} \approx 8$ years for the money to double to $200. (The actual return is $(1.09)^8 = \$199.26 \approx \$200$.)

10.11.6 CONTINUOUS COMPOUNDING

For a nominal annual rate r the effective annual interest rate under continuous compounding is $y = e^r - 1$.

EXAMPLE Consider the final amount F for an investment X using a nomimal interest rate of $r = 8\%$ and different compounding periods:

1. compounded quarterly: $F = X(1 + 0.08243) = X\left(1 + \frac{r}{4}\right)^4$
2. compounded monthly: $F = X(1 + 0.08300) = X\left(1 + \frac{r}{12}\right)^{12}$
3. compounded continuously: $F = X(1 + 0.08329) = X(1 + y) = Xe^r$

10.11.7 ECONOMIC ORDER QUANTITY FORMULA

- A order cost (dollars per order)
- D demand rate (pieces per unit time)
- H holding cost (dollars per item)
- Q order quantity (pieces per order)

The optimal order size is $Q^* = \sqrt{\frac{2AD}{H}}$

10.11.8 FINANCIAL TABLES

10.11.8.1 Compound interest: find final value

The following table uses Equation (10.11.2) to determine the final value in dollars (A) when one dollar ($P = 1$) is invested at an interest rate of i per time period, the length of investment time being n time periods.

For example, if $1 is invested at a return of 3% per time period, for $n = 60$ time periods, then the final value would be $5.89. Analogously, if $10 had been invested, then the final value would be $59.92.

	Interest rate (i)							
n	3%	4%	5%	6%	7%	8%	9%	10%
2	1.061	1.082	1.103	1.123	1.145	1.166	1.188	1.210
4	1.126	1.170	1.216	1.263	1.311	1.361	1.412	1.464
6	1.194	1.265	1.340	1.419	1.501	1.587	1.677	1.772
8	1.267	1.369	1.478	1.594	1.718	1.851	1.993	2.144
10	1.344	1.480	1.629	1.791	1.967	2.159	2.367	2.594
12	1.426	1.601	1.796	2.012	2.252	2.518	2.813	3.138
20	1.806	2.191	2.653	3.207	3.870	4.661	5.604	6.728
24	2.033	2.563	3.225	4.049	5.072	6.341	7.911	9.850
36	2.898	4.104	5.792	8.147	11.424	15.97	22.25	30.91
48	4.132	6.571	10.40	16.39	25.73	40.21	62.59	97.02
60	5.892	10.52	18.68	32.99	57.95	101.26	176.03	304.48

10.11.8.2 Compound interest: find interest rate

The following table uses Equation (10.11.2) to determine the compound interest rate i that must be obtained from an investment of one dollar ($P = 1$) to yield a final value of A (in dollars) when the initial amount is invested for n time periods.

For example, if \$1 is invested for $n = 60$ time periods, and the final amount obtained is \$4.00, then the interest rate was 2.34% per time period.

	Return after investing \$1 for n time periods (A)								
n	\$2.00	\$3.00	\$4.00	\$5.00	\$6.00	\$7.00	\$8.00	\$9.00	\$10.00
1	100	200	300	400	500	600	700	800	900
2	41.4	73.2	100	123	144	164	182	200	216
3	25.9	44.2	58.7	71.0	81.7	91.3	100	108	115
4	18.9	31.6	41.4	49.5	56.5	62.7	68.2	73.2	77.8
5	14.8	24.5	31.9	37.9	43.1	47.6	51.6	55.3	58.5
10	7.18	11.6	14.8	17.4	19.6	21.5	23.1	24.6	25.9
12	5.95	9.59	12.2	14.3	16.1	17.6	18.9	20.1	21.2
20	3.53	5.65	7.18	8.38	9.37	10.2	11.0	11.6	12.2
24	2.93	4.68	5.95	6.94	7.75	8.45	9.05	9.59	10.1
36	1.94	3.10	3.93	4.57	5.10	5.55	5.95	6.29	6.61
48	1.46	2.31	2.93	3.41	3.80	4.14	4.43	4.68	4.91
60	1.16	1.85	2.34	2.72	3.03	3.30	3.53	3.73	3.91

10.11.8.3 Compound interest: find annuity

The following table uses Equation (10.11.4) to determine the annuity (or mortgage) payment that must be paid each time period, for n time periods, at an interest rate of i% per time period, to pay off a loan of one dollar ($B = 1$).

For example, if \$100 is borrowed at 3% interest per time period, and the amount is to be paid back in equal amounts over $n = 10$ time periods, then the amount paid back per time period is \$11.72.

	Interest rate (i)								
n	2%	3%	4%	5%	6%	7%	8%	9%	10%
1	1.020	1.030	1.040	1.050	1.060	1.070	1.080	1.090	1.100
2	0.515	0.523	0.530	0.538	0.545	0.553	0.562	0.569	0.576
3	0.347	0.354	0.360	0.367	0.374	0.381	0.388	0.395	0.402
4	0.263	0.269	0.276	0.282	0.289	0.295	0.302	0.309	0.316
5	0.212	0.218	0.225	0.231	0.237	0.244	0.251	0.257	0.264
6	0.179	0.185	0.191	0.197	0.203	0.210	0.216	0.223	0.230
7	0.155	0.161	0.167	0.173	0.179	0.186	0.192	0.199	0.205
8	0.137	0.143	0.149	0.155	0.161	0.168	0.174	0.181	0.187
9	0.123	0.128	0.135	0.141	0.147	0.154	0.160	0.167	0.174
10	0.111	0.117	0.123	0.130	0.136	0.142	0.149	0.156	0.163
12	0.095	0.101	0.107	0.113	0.119	0.126	0.133	0.140	0.147
20	0.061	0.067	0.074	0.080	0.087	0.094	0.102	0.110	0.118
24	0.053	0.059	0.066	0.073	0.080	0.087	0.095	0.103	0.111
36	0.039	0.046	0.053	0.060	0.069	0.077	0.085	0.094	0.103

10.12 GAME THEORY

10.12.1 TWO PERSON NON-COOPERATIVE MATRIX GAMES

Given matrices $A = (a_{ij})$ and $B = (b_{ij})$ consider a game played as follows: Alice chooses action i (out of n possible actions) and Bob chooses action j (out of m possible actions). The *outcome* of the game is (i, j) and Alice and Bob receive payoffs of a_{ij} and b_{ij}, respectively. If using a *mixed* or *random strategy*, then Alice selects $\mathbf{x}^{\mathrm{T}} = (x_1, x_2, \ldots, x_n)$ where x_i corresponds to the probability that she chooses action i and Bob selects $\mathbf{y}^{\mathrm{T}} = (y_1, y_2, \ldots, y_m)$ where y_j corresponds to the probability that he chooses action j. Each player seeks to maximize their average payoff; Alice wants to maximize $\mathbf{x}^{\mathrm{T}}A\mathbf{y}$ while Bob wants to maximize $\mathbf{x}^{\mathrm{T}}B\mathbf{y}$.

If $a_{ij} \le a_{i^*j}$ for all i and j then i^* is a *dominant* strategy for Alice. Similarly, if $b_{ij} \le b_{ij^*}$ for all i and j then j^* is a *dominant* strategy for Bob. An outcome (i^*, j^*) is *Pareto optimal* if, for all i and j, the relation $a_{i^*j^*} < a_{ij}$ implies $b_{i^*j^*} > b_{ij}$.

If $A + B = 0$ then the game is a *zero sum game* and Bob equivalently is trying to minimize Alice's payoff. If $A + B \neq 0$ then the game is a *non-zero sum* game and there is potential for mutual gain or loss.

A *pure strategy* is one with no probabilistic component.

10.12.1.1 The pure zero sum game

Assume a zero sum game, $A + B = 0$. The outcome $(\widehat{i}, \widehat{j})$ is an *equilibrium* if $a_{i\widehat{j}} \le a_{\widehat{i}\widehat{j}} \le a_{\widehat{i}j}$ for all i and j. (Note that this implies $b_{\widehat{i}j} \le b_{\widehat{i}\widehat{j}} \le b_{i\widehat{j}}$.)

1. If $(\widehat{i}\widehat{j})$ is an equilibrium then $\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = a_{\widehat{i}\widehat{j}}$.
2. For all A, $\max_i \min_j a_{ij} \le \min_j \max_i a_{ij}$.
3. For all A, all outcomes are Pareto optimal as $a_{\widehat{i}\widehat{j}} < a_{ij}$ implies $b_{\widehat{i}\widehat{j}} = -a_{\widehat{i}\widehat{j}} > -a_{ij} = b_{ij}$.
4. For some A, there will be no equilibrium. For example, if $A = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]$ then $1 = \max_i \min_j a_{ij} < \min_j \max_i a_{ij} = 2$.
5. For some A, there will be an equilibrium which is not a dominant strategy. For example, if $A = \left[\begin{smallmatrix} 3 & 4 \\ 2 & 1 \end{smallmatrix}\right]$, then outcome (1,1) is an equilibrium with the property that Alice has the dominant strategy $i = 1$ as $a_{1,1} \ge a_{2,1}$ and $a_{1,2} \ge a_{2,2}$, but there is no dominant strategy for Bob as $a_{1,1} < a_{1,2}$ but $a_{2,1} > a_{2,2}$.
6. For some A, there will be multiple equilibria. If (i_1, j_1) and (i_2, j_2) are each an equilibrium then (i_1, j_2) and (i_2, j_1) are also equilibria. For example, if $A = \left[\begin{smallmatrix} 3 & 4 & 3 \\ 2 & 1 & 2 \\ 3 & 4 & 3 \end{smallmatrix}\right]$ then the outcomes (1,1), (1,3), (3,1) and (3,3) are each an equilibrium.
7. Order of actions: If Alice chooses her action before Bob and she chooses i then Bob will choose action $\beta(i) = \arg\min_j a_{ij}$. Hence Alice will choose $\widehat{i} = \arg\max_i a_{i\beta(i)} = \arg\max_i \min_j a_{ij}$ which is the *maximin* strategy. Alternatively, if Bob chooses his action before Alice he will choose the *minimax strategy* $\widehat{j} = \arg\min_j \max_i a_{ij}$. A player should never prefer to choose an action first but if there is an equilibrium the advantages of taking the second action can be eliminated.

10.12.1.2 The mixed zero sum game

The outcome $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is an *equilibrium* for the mixed zero sum game if

$$\mathbf{x}^{\mathrm{T}} A \widehat{\mathbf{y}} \leq \widehat{\mathbf{x}}^{\mathrm{T}} A \widehat{\mathbf{y}} \leq \widehat{\mathbf{x}}^{\mathrm{T}} A \mathbf{y} \qquad \text{for all } \mathbf{x} \in S_x \text{ and all } \mathbf{y} \in S_y$$

where $S_x = \{\mathbf{x} \mid 0 \leq x_i \text{ and } \sum_i x_i = 1\}$ and $S_y = \{\mathbf{y} \mid 0 \leq y_j \text{ and } \sum_j y_j = 1\}$.

1. If $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is an equilibrium then $\max_{\mathbf{x} \in S_x} \min_{\mathbf{y} \in S_y} \mathbf{x}^{\mathrm{T}} A \mathbf{y} = \min_{\mathbf{y} \in S_y} \max_{\mathbf{x} \in S_x} \mathbf{x}^{\mathrm{T}} A \mathbf{y} = \widehat{\mathbf{x}}^{\mathrm{T}} A \widehat{\mathbf{y}}$.
2. For all A, there exists at least one mixed strategy equilibrium. If there is more than one equilibrium, the average payoff to each player is independent of which equilibrium is used.
3. If $\min_j \sum_i x_i a_{ij} \leq \min_j \sum_i \widehat{x}_i a_{ij}$ for all $\mathbf{x} \in S_x$ and $\max_i \sum_j a_{ij} y_j \geq \max_i \sum_j a_{ij} \widehat{y}_j$ for all $\mathbf{y} \in S_y$ then $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is an equilibrium.
4. If $A \geq 0$ then an equilibrium strategy for Alice solves $\max_{\mathbf{x}} v$ subject to $v \leq \sum_i x_i a_{ij}$ for all j, with $\sum_i x_i = 1$ and $x_i \geq 0$. The optimal value of v corresponds to the average payoff to Alice in equilibrium. If $x_i' = x_i / v$ this is

$$\text{minimize} \sum_i x_i' \quad \text{subject to} \quad \begin{cases} \sum_i x_i' a_{ij} \geq 1, & \text{for all } j, \\ x_i' \geq 0, & \text{for all } i. \end{cases}$$

Similarly, if $\mathbf{y}' = (y_1', \dots, y_n')$ is a solution to

$$\text{maximize} \sum_j y_j' \quad \text{subject to} \quad \begin{cases} \sum_j a_{ij} y_j' \leq 1, & \text{for all } i, \\ y_j' \geq 0, & \text{for all } j, \end{cases}$$

then $\mathbf{y} = \mathbf{y}' / \sum_j y_j'$ is the equilibrium strategy for Bob. The payoff to Alice in equilibrium is $v = 1/\sum_i x_i' = 1/\sum_j y_j'$. For example, if $A = \left[\begin{smallmatrix} 1 & 2 \\ 2 & 1 \end{smallmatrix}\right]$ then

$$\text{minimize } x_1' + x_2' \quad \text{subject to} \quad \begin{cases} x_1' + 2x_2' \geq 1, & x_1' \geq 0, \\ 2x_1' + x_2' \geq 1, & x_2' \geq 0, \end{cases}$$

has the solution $x_1' = x_2' = \frac{1}{3}$. The payoff to Alice in equilibrium is $\frac{3}{2}$ and the equilibrium is $(x_1, x_2) = (\frac{1}{2}, \frac{1}{2})$.
5. If A is 2×2 and $a_{1,1} < a_{1,2}$, $a_{1,1} < a_{2,1}$, $a_{2,2} < a_{2,1}$, and $a_{2,2} < a_{1,2}$ (so neither player has a dominant strategy) then the mixed strategy equilibrium is

$$\begin{aligned} \widehat{\mathbf{x}} &= \left(\frac{a_{2,1} - a_{2,2}}{a_{1,2} + a_{2,1} - a_{1,1} - a_{2,2}}, \frac{a_{1,2} - a_{1,1}}{a_{1,2} + a_{2,1} - a_{1,1} - a_{2,2}} \right), \\ \widehat{\mathbf{y}} &= \left(\frac{a_{1,2} - a_{2,2}}{a_{1,2} + a_{2,1} - a_{1,1} - a_{2,2}}, \frac{a_{2,1} - a_{1,1}}{a_{1,2} + a_{2,1} - a_{1,1} - a_{2,2}} \right) \end{aligned} \tag{10.12.1}$$

and the payoff to Alice in equilibrium is $\frac{a_{1,2} a_{2,1} - a_{1,1} a_{2,2}}{a_{1,2} + a_{2,1} - a_{1,1} - a_{2,2}}$. which is

$$\max_{\widehat{\mathbf{x}} \in S_x} \min_{\widehat{\mathbf{y}} \in S_y} \sum_i \sum_j \widehat{x}_i a_{ij} \widehat{y}_j = \min_{\widehat{\mathbf{y}} \in S_y} \max_{\widehat{\mathbf{x}} \in S_x} \sum_i \sum_j \widehat{x}_i a_{ij} \widehat{y}_j \tag{10.12.2}$$

For example, if $A = \left[\begin{smallmatrix}1 & 2\\ 2 & 1\end{smallmatrix}\right]$ then the solution is $\widehat{\mathbf{x}} = \widehat{\mathbf{y}} = (\frac{1}{2}, \frac{1}{2})$ and the payoff to Alice in equilibrium is $\frac{3}{2}$.

6. Order of actions: It is never an advantage to take the first action. However, if the first player uses a mixed strategy equilibrium, then the advantages of taking the second action can always be eliminated.

10.12.1.3 The non-zero sum game

The outcome $(\widehat{ij})$ is a *Nash equilibrium* if $a_{i\widehat{j}} \le a_{\widehat{ij}}$ for all i and $b_{\widehat{i}j} \le b_{\widehat{ij}}$ for all j. For the mixed strategy game the outcome $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is a Nash equilibrium if $\mathbf{x}^{\mathrm{T}} A \widehat{\mathbf{y}} \le \widehat{\mathbf{x}}^{\mathrm{T}} A \widehat{\mathbf{y}}$ for all $\mathbf{x} \in S_x$ and $\widehat{\mathbf{x}}^{\mathrm{T}} A \mathbf{y} \le \widehat{\mathbf{x}}^{\mathrm{T}} A \widehat{\mathbf{y}}$ far all $\mathbf{y} \in S_y$.

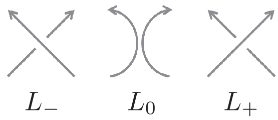
1. For all A and B, there exists at least one mixed strategy Nash equilibrium.
2. For all A and B, $\max_i \sum_j a_{ij} \widehat{y}_j \le \sum_i \sum_j \widehat{x}_i a_{ij} \widehat{y}_j$ and $\max_j \sum_i \widehat{x}_i b_{ij} \le \sum_i \sum_j \widehat{x}_i b_{ij} \widehat{y}_j$ is a necessary and sufficient condition for $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ to be a Nash equilibrium. For example, if $A = \left[\begin{smallmatrix}4 & 1\\ 1 & 3\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}3 & 1\\ 1 & 4\end{smallmatrix}\right]$ then there are three Nash equilibra: (i) $\{\widehat{\mathbf{x}} = (0, 1), \widehat{\mathbf{y}} = (0, 1)\}$, (ii) $\{\widehat{\mathbf{x}} = (\frac{3}{5}, \frac{2}{5}), \widehat{\mathbf{y}} = (\frac{2}{5}, \frac{3}{5})\}$, and (iii) $\{\widehat{\mathbf{x}} = (1, 0), \widehat{\mathbf{y}} = (1, 0)\}$.
3. For all A and B, if $\sum_j a_{ij} \widehat{y}_j$ is a constant for all i and $\sum_i \widehat{x}_i b_{ij}$ is a constant for all j (i.e., each player chooses an action to make the other indifferent to their action) then $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is a Nash equilibrium.
4. For some A and B if (i_1, j_1) and (i_2, j_2) are each a (pure strategy) Nash equilibrium then, unlike the case for a zero sum game, $a_{i_1 j_1}$ need not equal $a_{i_2 j_2}$, $b_{i_1 j_1}$ need not equal $b_{i_2 j_2}$, and neither (i_1, j_2) nor (i_2, j_1) need be a Nash equilibrium. For example, if $A = \left[\begin{smallmatrix}4 & 1\\ 1 & 3\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}3 & 1\\ 1 & 4\end{smallmatrix}\right]$ then both of the outcomes (1,1) and (2,2) are Nash equilibra yet neither (1,2) nor (2,1) are Nash equilibra.
5. *Prisoners' Dilemma*: A game in which there is a dominant strategy for both players but it is not Pareto optimal. For example, if $A = \left[\begin{smallmatrix}10 & 2\\ 15 & 5\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}10 & 15\\ 2 & 5\end{smallmatrix}\right]$ then the dominant (and equilibrium) outcome is (2,2) since $a_{2j} > a_{1j}$ for all j and and $b_{i2} > b_{i1}$ for all i. Here, Alice and Bob receive a payoff of 5 although the outcome (1,1) would be preferred by both because each would receive a payoff of 10.
6. *Braess Paradox*: A game in which the Nash equilibrium has a worse payoff for all players than the Nash equilibrium which would result if there were fewer possible actions. For example, if $A = \begin{bmatrix}10 & 2 & 1\\ 15 & 5 & 2\\ 20 & 6 & 4\end{bmatrix}$ and $B = \begin{bmatrix}10 & 15 & 20\\ 2 & 5 & 6\\ 1 & 2 & 4\end{bmatrix}$ then the Nash equilibrium is (3,3) which is worse for both players than the Nash equilibrium (2,2) which would occur if the third option for each player was unavailable.
7. Order of actions: If Alice chooses action i before Bob chooses an action, then Bob will choose $\beta(i) = \arg\max_j b_{ij}$ and hence Alice will choose $\widehat{i} = \arg\max_i a_{i\beta(i)}$. Alternatively, if Bob chooses an action before Alice, he will choose $\widehat{j} = \arg\max_j b_{\alpha(j)j}$ where $\alpha(j) = \arg\max_i a_{ij}$. Unlike the zero sum game there might be an advantage to choosing the action first, e.g., for $A = \left[\begin{smallmatrix}4 & 1\\ 1 & 3\end{smallmatrix}\right]$ and $B = \left[\begin{smallmatrix}3 & 1\\ 1 & 4\end{smallmatrix}\right]$ if Alice is first, the outcome will be (1,1) with a payoff of 4 to Alice and 3 to Bob. If Bob is first the outcome will be (2,2) with a payoff of 4 to Bob and 3 to Alice.

10.13 KNOT THEORY

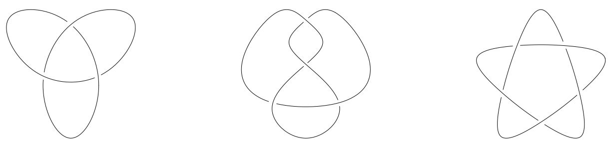
1. A *knot* is a closed, non-self-intersecting curve that is embedded in three dimensions and cannot be untangled to produce a simple loop (i.e., the unknot). Knots do not exist in higher-dimensional spaces.
2. Each place in a knot where 2 strands touch and one passes over (or under) the other is called a *crossing*. The number of crossings in a knot is called the *crossing number*. Let $N(n)$ be the number of distinct prime knots with n crossings, counting chiral versions of the same knot separately. Then $\frac{1}{3}\left(2^{n-2}-1\right) \leq N(n) \lesssim e^n$.

n	3	4	5	6	7	8	9	10	11
Number of knots with n crossings	1	1	2	3	7	21	49	165	552

3. The traditional *Alexander–Briggs notation* for a knot (e.g., 3_1, 4_1, 5_2) is the number of crossings with a subscript. The subscript starts at 1 and counts the number of knots with that many crossings; the order is arbitrary.
4. The *unknot*, or *trivial knot*, is a closed loop that is not knotted.
5. The *trefoil knot*, or 3_1 knot or *overhand knot*, has three crossings.
6. The *square knot* is the knot sum of two trefoils with opposite chiralities: $3_1\#3_1^*$. The *granny knot* is the knot sum of two trefoils with the same chirality: $3_1\#3_1$. The *knot sum* operation is denoted $\#$. Two knots are summed by placing them side-by-side and joining them by straight bars so the orientation is preserved in the sum; this is not a well-defined operation as it depends on the representation. The knot sum of two unknots is another unknot.
7. There are many knot representation schemes, including *Alexander polynomials* ($\Delta(x)$), *Conway polynomials* ($C(x)$), and *HOMFLY polynomials* ($P(\ell, m)$). The left and right trefoil knots have the same Alexander polynomials and different HOMFLY polynomials.
8. Some knot representations use *skeins* (see figure to right) to recursively simplify a knot. For example, the Conway polynomial of a knot may be determined by $C(L_+) = C(L_-) + xC(L_0)$ with $C(\text{unknot}) = 1$.



9. Information about the simplest knots:



Name:	trefoil knot	figure eight knot	Solomon seal knot
Notation:	3_1	4_1	5_1
$\Delta(x)$:	$x-1+x^{-1}$	$-x^{-1}+3-x$	$x^2-x+1-x^{-1}+x^{-2}$
$P(\ell, m)$:	$-\ell^4+m^2\ell^2-2\ell^2$	$m^2-\left(\ell^2+\ell^{-2}+1\right)$	$m^4\ell^4+m^2(-\ell^6-4\ell^4)+(3\ell^4+2\ell^6)$
$C(x)$:	x^2+1	$1-x^2$	x^4+3x^2+1

10.13.1 KNOTS UP TO EIGHT CROSSINGS

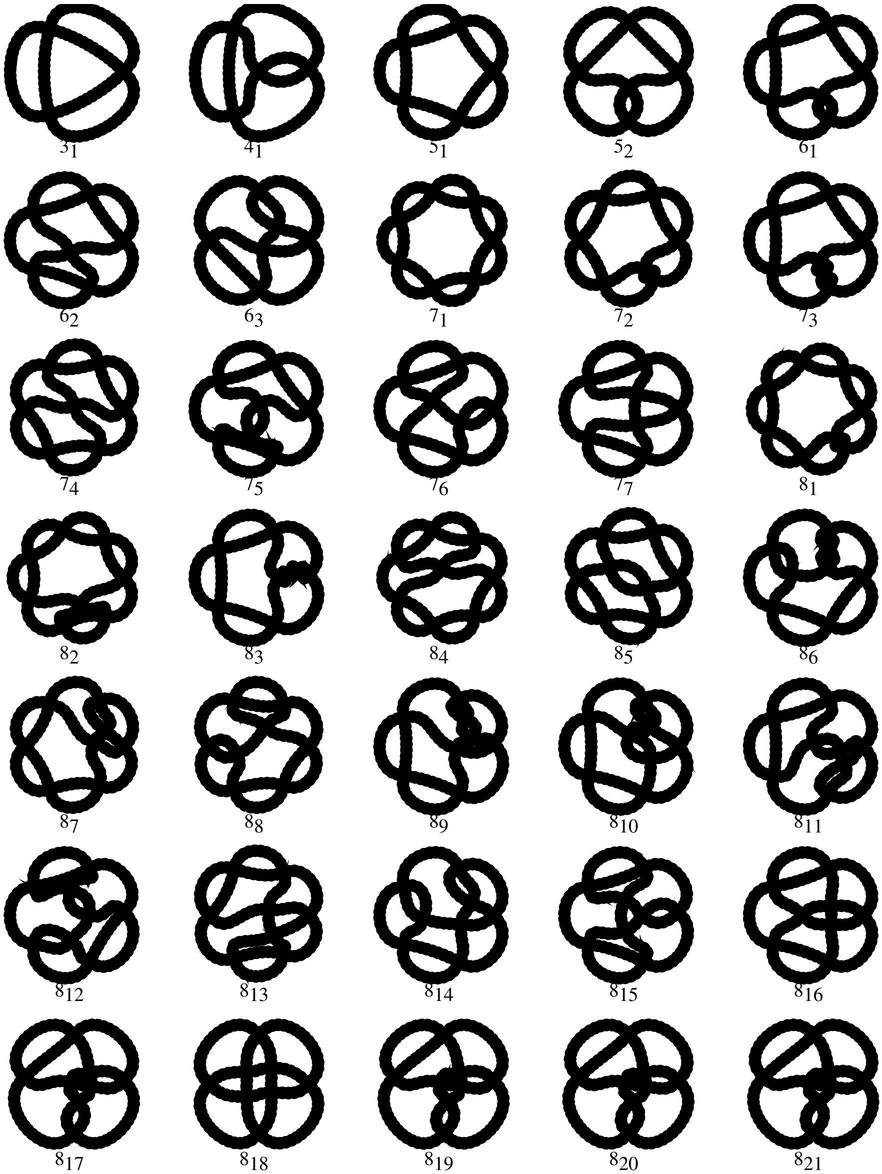


Image by Charlie Gunn and David Broman. Copyright The Geometry Center, University of Minnesota. With permission.

10.14 LATTICES

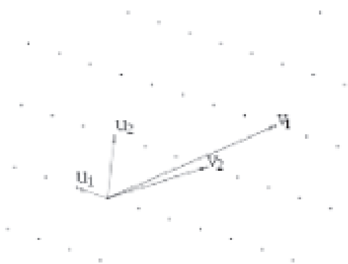
Let $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_m \end{bmatrix}$ be a set of vectors in $\mathbb{R}^n$. If the $\{\mathbf{b}_k\}$ are independent, then points in $\mathbb{R}^n$ can be written as linear combinations of the $\{\mathbf{b}_k\}$: $\mathbf{x} = \sum_{k=1}^{m} r_k \mathbf{b}_k$ where the $\{r_k\}$ are real numbers. Now, instead, only allow integer coefficients:

$$L = \left\{ \sum_{k=0}^{m} d_k \mathbf{b}_k \qquad d_k \text{ are integers} \right\}$$

Then L is a *lattice* of dimension n and rank m.

Given a lattice there are many bases that could be used to describe it. The goal of lattice basis reduction is to determine, from an integer lattice basis, a basis with "short," nearly orthogonal vectors.

The size of a vector is its Euclidean length: $\|\mathbf{b}\|_2 = \sqrt{\mathbf{b}^{\mathrm{T}}\mathbf{b}}$. The size of the basis, in the full rank case, is the volume $\sqrt{\det(B^{\mathrm{T}}B)}$. This is constant for a given lattice (up to sign) and is called the *lattice constant*. The figure shows a lattice with basis $V = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{bmatrix}$ and a "shorter" basis $U = \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{bmatrix}$.



EXAMPLE

Consider the basis $B = \left[\begin{smallmatrix} 1 & 4 \\ 9 & 37 \end{smallmatrix}\right] = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 \end{bmatrix}$ with $\mathbf{b}_1 = \begin{bmatrix} 1 & 9 \end{bmatrix}^{\mathrm{T}}$ and $\mathbf{b}_2 = \begin{bmatrix} 4 & 37 \end{bmatrix}^{\mathrm{T}}$. Defining $\mathbf{b}_1' = \mathbf{b}_2 - 4\mathbf{b}_1 = \begin{bmatrix} 0 & 1 \end{bmatrix}^{\mathrm{T}}$ and $\mathbf{b}_2' = \mathbf{b}_1 - 9\mathbf{b}_1' = \begin{bmatrix} 1 & 0 \end{bmatrix}^{\mathrm{T}}$ results in the basis $B' = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right] = \begin{bmatrix} \mathbf{b}_1' & \mathbf{b}_2' \end{bmatrix}$ with shorter vectors. Here the lattice constant is 1.

Given the $m \times n$ integer matrix A there is the lattice $L(A) = \{\mathbf{y} \in \mathbb{R}^m \mid \mathbf{y} = A\mathbf{x}, \mathbf{x} \in \mathbb{Z}^n\}$. This lattice is said to be generated by the columns of A. Computational problems involving integer lattices are the following:

1. Given an integral vector $\mathbf{b}$, determine if $\mathbf{b}$ is in $L(A)$. If $\mathbf{b} \in L(A)$, determine $\mathbf{b}$ as an integral linear combination of the columns of A. This problem is called the *linear equation integer feasibility problem*.
2. Given A and $\mathbf{b}$, the *closest vector problem* is to find $\min_{\mathbf{y}} \{|\mathbf{b} - \mathbf{y}| \mid \mathbf{b} \in L(A)\}$
3. The *shortest vector problem* is the *closest vector problem* with $\mathbf{b} = \mathbf{0}$ and the resulting vector being non-zero. That is, find $\min_{\mathbf{z}} \{\mathbf{z} \mid \mathbf{z} \in L(A), \mathbf{z} \neq \mathbf{0}\}$.

It is difficult to determine the shortest vector for all bases. An approximate solution is obtained by the LLL (for Lenstra, Lenstra, and Lovász) algorithm which runs in polynomial time. It outputs a short vector $\mathbf{b}_1'$ with $\|\mathbf{b}'\|_1 \leq 2^{(n-1)/4} \det(L)^{1/n}$.

There are many applications of finding the shortest vector:

- *Given the value* $\alpha = \sqrt{17} - \sqrt{13}$, *find a polynomial that has* α *as a root.*
  If α satisfies a polynomial of degree p, then there is an integer relation among the values $\{1, \alpha, \alpha^2, \dots, \alpha^p\}$. Choosing $p = 4$, consider the basis

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ C & C\alpha & C\alpha^2 & C\alpha^3 & C\alpha^4 \end{bmatrix} \qquad (10.14.1)$$

  where C is an arbitrary constant, the bigger the better. Using $C = 10^{15}$ evaluate B to 10 decimal places and then round the values to the nearest integer. Now B has integer values. This lattice has a shortest vector $\begin{bmatrix} 16 & 0 & -60 & 0 & 1 & -3 \end{bmatrix}^{\mathrm{T}}$ suggesting that α is a solution of the polynomial $16 - 60p^2 + p^4 = 0$, which is correct.
- *Suspecting that* $\beta = \cot\frac{\pi}{8} + \cot\frac{2\pi}{8} + \cot\frac{3\pi}{8}$ *is a linear combination of* $\{1, \sqrt{2}\}$, *find that combination.* Use the basis $B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ C & C\sqrt{2} & C\beta \end{bmatrix}$, evaluate it numerically for $C = 10^{20}$, round to integer values, and then use LLL to find the shortest vector $\begin{bmatrix} -1 & -2 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$. This suggests that $\cot\frac{\pi}{8} + \cot\frac{2\pi}{8} + \cot\frac{3\pi}{8} = 1 + 2\sqrt{2}$, which is correct.
- *Suspecting that* $\pi = \sum_{k=0}^{\infty} \frac{1}{16^k} \left[\frac{a_1}{8k+1} + \frac{a_2}{8k+2} + \dots + \frac{a_7}{8k+7} \right]$, *find* $\{a_i\}$.
  Use the basis

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ 0 & 0 & 0 & 0 & & 0 \\ 0 & 0 & 0 & 0 & & 1 \\ C\pi & C\gamma_1 & C\gamma_2 & C\gamma_3 & \dots & C\gamma_8 \end{bmatrix}, \qquad (10.14.2)$$

  where $\gamma_i = \sum_{k=0}^{\infty} \frac{1}{16^k} \frac{1}{8k+i}$. Use a large value of C and proceed as before. In this case there are two short vectors ($\begin{bmatrix} -1 & 4 & 0 & 0 & -2 & -1 & -1 & 0 & 0 \end{bmatrix}$ and $\begin{bmatrix} -2 & 0 & 8 & 4 & 4 & 0 & 0 & -1 & 0 \end{bmatrix}$) suggesting:

$$\begin{aligned} \pi &= \sum_{k=0}^{\infty} \frac{1}{16^k} \left[\frac{4}{8k+1} - \frac{2}{8k+4} - \frac{1}{8k+5} - \frac{1}{8k+6} \right] \\ 2\pi &= \sum_{k=0}^{\infty} \frac{1}{16^k} \left[\frac{8}{8k+2} + \frac{4}{8k+3} + \frac{4}{8k+4} - \frac{1}{8k+7} \right] \end{aligned} \qquad (10.14.3)$$

  both of which are correct.

10.15 LOGIC

10.15.1 SYMBOLIC LOGIC

10.15.1.1 Propositional calculus

Propositional calculus is the study of statements: how they are combined and how to determine their truth. Statements (or propositions) are combined by means of *connectives* such as *and* ($\wedge$), *or* ($\vee$), *not* ($\neg$, or sometimes $\sim$), *implies* ($\rightarrow$), and *if and only if* ($\longleftrightarrow$, sometimes written $\iff$ or "iff"). Propositions are denoted by letters $\{p, q, r, \ldots\}$. For example, if p is the statement "$x = 3$," and q the statement "$y = 4$," then $p \vee \neg q$ would be interpreted as "$x = 3$ or $y \neq 4$." To determine the truth of a statement, *truth tables* are used. Using T (for true) and F (for false), the truth tables for these connectives are as follows:

p	q	$p \wedge q$	$p \vee q$	$p \rightarrow q$	$p \longleftrightarrow q$
T	T	T	T	T	T
T	F	F	T	F	F
F	T	F	T	T	F
F	F	F	F	T	T

p	$\neg p$
T	F
F	T

The proposition $p \rightarrow q$ can be read "*If p then q*" or, less often, "*q if p*." The table shows that "$p \vee q$" is an *inclusive or* because it is true even when p and q are both true. Thus, the statement "I'm watching TV or I'm doing homework" is a true statement if the narrator happens to be both watching TV and doing homework. Note that $p \rightarrow q$ is false only when p is true and q is false. Thus, *a false statement implies any statement* and *a true statement is implied by any statement*.

10.15.1.2 Truth tables as functions

If we assign the value 1 to T, and 0 to F, then the truth table for $p \wedge q$ is simply the value pq. This can be done with all the connectives as follows:

Connective	Arithmetic function
$p \wedge q$	pq
$p \vee q$	$p + q - pq$
$p \rightarrow q$	$1 - p + pq$
$p \longleftrightarrow q$	$1 - p - q + 2pq$
$\neg p$	$1 - p$

These formulas may be used to verify tautologies, because, from this point of view, a tautology is a function whose value is identically 1. In using them, it is useful to remember that $pp = p^2 = p$, since $p = 0$ or $p = 1$.

10.15.1.3 Tautologies

A statement such as $(p \to (q \wedge r)) \vee \neg p$ is a *compound statement* composed of the *atomic propositions* p, q, and r. The letters P, Q, and R are used to designate compound statements. A *tautology* is a compound statement which always is true, regardless of the truth values of the atomic statements used to define it. For example, a simple tautology is $(\neg\neg p) \longleftrightarrow p$. Tautologies are logical truths. More examples:

Law of the excluded middle	$p \vee \neg p$
De Morgan's laws	$\neg(p \vee q) \longleftrightarrow (\neg p \wedge \neg q)$
	$\neg(p \wedge q) \longleftrightarrow (\neg p \vee \neg q)$
Modus ponens	$(p \wedge (p \to q)) \to q$
Contrapositive law	$(p \to q) \longleftrightarrow (\neg q \to \neg p)$
Reductio ad absurdum	$(\neg p \to p) \to p$
Elimination of cases	$((p \vee q) \wedge \neg p) \to q$
Transitivity of implication	$((p \to q) \wedge (q \to r)) \to (p \to r)$
Proof by cases	$((p \to q) \wedge (\neg p \to q)) \to q$

Idempotent laws	$p \wedge p \longleftrightarrow p$;	$p \vee p \longleftrightarrow p$
Commutative laws	$(p \wedge q) \longleftrightarrow (q \wedge p)$;	$(p \vee q) \longleftrightarrow (q \vee p)$
Associative laws	$(p \wedge (q \wedge r)) \longleftrightarrow ((p \wedge q) \wedge r)$	
	$(p \vee (q \vee r)) \longleftrightarrow ((p \vee q) \vee r)$	

10.15.1.4 Rules of inference

A *rule of inference* in propositional calculus is a method of arriving at a valid (true) *conclusion*, given certain statements, assumed to be true, which are called the *hypotheses*. For example, suppose that P and Q are compound statements. Then if P and $P \to Q$ are true, then Q must necessarily be true. This follows from the *modus ponens* tautology in the above list of tautologies. We write this rule of inference $P,\ P \to Q \Rightarrow Q$. It is also classically written

$$\begin{array}{c} P \\ P \to Q \\ \hline Q \end{array}$$

Some examples of rules of inferences follow, all derived from the above list of tautologies:

Modus ponens	$P,\ P \to Q \Rightarrow Q$
Contrapositive	$P \to Q \Rightarrow \neg Q \to \neg P$
Modus tollens	$P \to Q, \neg Q \Rightarrow \neg P$
Transitivity	$P \to Q, Q \to R \Rightarrow P \to R$
Elimination of cases	$P \vee Q, \neg P \Rightarrow Q$
"And" usage	$P \wedge Q \Rightarrow P, Q$

10.15.1.5 Deductions

A *deduction* from hypotheses is a list of statements, each one of which is either one of the hypotheses, a tautology, or follows from previous statements in the list by a valid rule of inference. It follows that if the hypotheses are true, then the conclusion must be true. Suppose for example, that we are given *hypotheses* $\neg q \to p,\ q \to \neg r,\ r$; it is required to deduce the *conclusion* p. A deduction showing this, with reasons for each step is as follows:

	Statement	Reason
1.	$q \to \neg r$	Hypothesis
2.	r	Hypothesis
3.	$\neg q$	Modus tollens (1,2)
4.	$\neg q \to p$	Hypothesis
5.	p	Modus ponens (3,4)

10.15.1.6 Predicate calculus

Unlike propositional calculus, which may be considered the skeleton of logical discourse, *predicate calculus* is the language in which most mathematical reasoning takes place. It uses the symbols of propositional calculus, with the exception of the propositional variables p, q, ... Predicate calculus uses the *universal quantifier* $\forall$, the *existential quantifier* $\exists$, *predicates* $P(x), Q(x,y), \ldots$, *variables* $x, y, \ldots$, and assumes a *universe* U from which the variables are taken. The quantifiers are illustrated in the following table.

Symbol	Read as	Usage	Interpretation
$\exists$	There exists an	$\exists x(x > 10)$	There is an x such that $x > 10$
$\forall$	For all	$\forall x(x^2 + 1 \neq 0)$	For all x, $x^2 + 1 \neq 0$

Predicates are variable statements which may be true or false, depending on the values of its variable. In the above table, "$x > 10$" is a predicate in the one variable x as is "$x^2 + 1 \neq 0$." Without a given universe, we cannot decide if a statement is true or false. Thus $\forall x(x^2 + 1 \neq 0)$ is true if the universe U is the real numbers, but false if U is the complex numbers. A useful rule for manipulating quantifiers is

$$\neg\exists x P(x) \longleftrightarrow \exists x \neg P(x),$$
$$\neg\forall x P(x) \longleftrightarrow \exists x \neg P(x).$$

For example, it is not true that all people are mortal if and only if there is a person who is immortal. Here the universe U is the set of people, and $P(x)$ is the predicate "x is mortal." This works with more than one quantifier. Thus,

$$\neg\forall x \exists y P(x,y) \longleftrightarrow \exists x \forall y \neg P(x,y).$$

For example, if it is not true that every person loves someone, then it follows that there is a person who loves no one (and vice versa).

Fermat's last theorem, stated in terms of the predicate calculus (U = the positive integers), is $\quad \forall n \forall a \forall b \forall c[(n > 2) \to (a^n + b^n \neq c^n)]$

10.15.2 FUZZY LOGIC

Notation

- A, B, C fuzzy sets
- x element of a set
- μ_A membership function for set A; in the range $0 \leq \mu_A(x) \leq 1$

Concepts

1. An ordinary set (not a fuzzy set) is a *crisp set*.
2. **Definitions**
   (a) The point x belongs to the set A if and only if $\mu_A(x) = 1$.
   (b) *Set equality* $A = B$ if and only if for all x, $\mu_A(x) = \mu_B(x)$
   (c) *Set inclusion* $A \subset B$ if and only if for all x, $\mu_A(x) \leq \mu_B(x)$
3. **Fuzzy set Operations**
   (a) Complement $\mu_{\text{not } A}(x) = 1 - \mu_A(x)$
   (b) Intersection $\mu_{A\cap B}(x) = \min\left(\mu_A(x), \mu_B(x)\right)$
   (c) Union $\mu_{A\cup B}(x) = \max\left(\mu_A(x), \mu_B(x)\right)$
4. **Classical set theory operations that apply to fuzzy sets**
   (a) De Morgans law $\overline{(A \cap B)} = \overline{A} \cup \overline{B}$
   $\overline{(A \cup B)} = \overline{A} \cap \overline{B}$
   (b) Associativity $(A \cap B) \cap C = A \cap (B \cap C)$
   $(A \cup B) \cup C = A \cup (B \cup C)$
   (c) Commutativity $A \cap B = B \cap A$ $A \cup B = B \cup A$
   (d) Distributivity $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
5. **Hedges**
   Hedges are word concepts that can be interpreted as operators modifying fuzzy values. For example:
   (a) *Very* $\mu_A(x) = [\mu_B(x)]^2$
   (b) *Somewhat* $\mu_A(x) = [\mu_B(x)]^{1/2}$
6. **Fuzzy controllers**
   A common approach to creating a fuzzy controller is the sequence of steps:
   (a) *Fuzzification* Define fuzzy sets and their membership functions. Example: $P = \{\text{Alice}, \text{Bob}, \text{Charlie}\}$; $\mu_{\text{like cool}}(p) = (0.1, 0.3, 0.3)$, $\mu_{\text{like warm}}(p) = (0.4, 0.8, 0.9)$, $\mu_{\text{come to party}}(p) = (0.7, 0.3, 0.6), \ldots$
   (b) *Rule evaluation* Implement rules such as "If *variable* is *property* then *action*." Example: "If very cool inside then turn on heat," "If very warm inside then turn on A/C," "If cold outside then it takes several hours to heat house," ...
   (c) *Defuzzification* Obtain a crisp result from fuzzy analysis. Example: "Turn on heat at 5 PM for party tonight."

10.15.3 LINEAR TEMPORAL LOGIC

Linear Temporal Logic has propositions and operators. It represents infinite sequences of states where each point in time has a unique successor, based on a linear-time perspective.

1. Propositions: “True" ($\top$), “False" ($\bot$), ϕ, ψ, θ, ...
2. Basic (“ordinary") logical operators:

   (a) $\neg$ negation
   (b) $\vee$ disjunction or “or"
   (c) $\vdash$ provable: “$x \vdash y$ means y is provable from x"

3. Derived logical operators:

   (a) $\wedge$ conjunction or “and": $\phi \wedge \psi \equiv \neg(\neg\phi \vee \neg\psi)$
   (b) $\Rightarrow$ implication: $\phi \Rightarrow \psi \equiv \neg\phi \vee \psi$
   (c) $\leftrightarrow$ equivalence: $\phi \leftrightarrow \psi \equiv (\phi \Rightarrow \psi) \wedge (\psi \Rightarrow \phi)$

4. Basic Temporal operators:

   (a) $\bigcirc$ next (or an “X" is used for “NeXt time")
   (b) $\mathcal{U}$ until (or a “U" is used)

5. Derived Temporal operators:

   (a) $\Diamond$ eventually (or “sometime in the Future"): $\Diamond\psi \equiv \top\,\mathcal{U}\,\psi$
   (b) $\Box$ always (or “Globally in the future"): $\Box\psi \equiv \neg\Diamond\neg\psi$
   (c) $\mathcal{R}$ release (or an “R" is used) $\phi\,\mathcal{R}\,\psi \equiv \neg(\neg\phi\,\mathcal{U}\,\neg\psi)$
   (d) $\Box\Diamond$ infinitely often
   (e) $\Diamond\Box$ eventually forever

Properties

1. Distributive Laws
   (a) $\bigcirc(\phi \vee \psi) \equiv (\bigcirc\,\phi) \wedge (\bigcirc\,\psi)$
   (b) $\bigcirc(\phi \wedge \psi) \equiv (\bigcirc\,\phi) \vee (\bigcirc\,\psi)$
   (c) $\bigcirc(\phi\,\mathcal{U}\,\psi) \equiv (\bigcirc\,\phi)\,\mathcal{U}(\bigcirc\,\psi)$
   (d) $\Diamond(\phi \vee \psi) \equiv (\Diamond\phi) \vee (\Diamond\psi)$
   (e) $\Box(\phi \wedge \psi) \equiv (\Box\phi) \wedge (\Box\psi)$
   (f) $\theta\,\mathcal{U}(\phi \vee \psi) \equiv (\theta\,\mathcal{U}\,\phi) \vee (\theta\,\mathcal{U}\,\psi)$
   (g) $(\phi \wedge \psi)\,\mathcal{U}\,\theta \equiv (\phi\,\mathcal{U}\,\theta) \wedge (\psi\,\mathcal{U}\,\theta)$
2. Idempotency Laws
   (a) $\Diamond\Diamond\phi \equiv \Diamond\phi$
   (b) $\Box\Box\phi \equiv \Box\phi$
   (c) $\phi\,\mathcal{U}(\phi\,\mathcal{U}\,\psi) \equiv \phi\,\mathcal{U}\,\psi$
   (d) $(\phi\,\mathcal{U}\,\psi)\,\mathcal{U}\,\psi \equiv \phi\,\mathcal{U}\,\psi$
3. Absorption Laws
   (a) $\Diamond\Box\Diamond\phi \equiv \Box\Diamond\phi$
   (b) $\Box\Diamond\Box\phi \equiv \Diamond\Box\phi$
4. Expansion Laws
   (a) $\phi\,\mathcal{U}\,\psi \equiv \phi \vee (\phi \wedge \bigcirc(\phi\,\mathcal{U}\,\psi))$
   (b) $\Diamond\phi \equiv \phi \vee (\bigcirc\,\Diamond\phi)$
   (c) $\Box\phi \equiv \phi \wedge (\bigcirc\,\Box\phi)$
5. Negation propagation
   (a) $\neg\bigcirc\,\phi \equiv \bigcirc\,\neg\phi$
   (b) $\neg\Box\phi \equiv \Diamond\neg\phi$
   (c) $\neg\Diamond\phi \equiv \Box\neg\phi$
   (d) $\neg(\phi\,\mathcal{U}\,\psi) \equiv (\neg\phi\,\mathcal{R}\,\neg\psi)$
   (e) $\neg(\phi\,\mathcal{R}\,\psi) \equiv (\neg\phi\,\mathcal{U}\,\neg\psi)$

10.16 MOMENTS OF INERTIA

The moment of inertia is $\int_V \rho(r) r_\perp^2 \, dV$ where $\rho(r)$ is the density and $r_\perp$ is the perpendicular distance from the axis of rotation.

	Body	Axis	Moment of inertia
(1)	Uniform thin rod	Normal to the length, at one end	$m\frac{l^2}{3}$
(2)	Uniform thin rod	Normal to the length, at the center	$m\frac{l^2}{12}$
(3)	Thin rectangular sheet, sides a and b	Through the center parallel to b	$m\frac{a^2}{12}$
(4)	Thin rectangular sheet, sides a and b	Through the center perpendicular to the sheet	$m\frac{a^2+b^2}{12}$
(5)	Thin circular sheet of radius r	Normal to the plate through the center	$m\frac{r^2}{2}$
(6)	Thin circular sheet of radius r	Along any diameter	$m\frac{r^2}{4}$
(7)	Thin circular ring, radii r_1 and r_2	Through center normal to plane of ring	$m\frac{r_1^2+r_2^2}{2}$
(8)	Thin circular ring, radii r_1 and r_2	Along any diameter	$m\frac{r_1^2+r_2^2}{4}$
(9)	Rectangular parallelopiped, edges a, b, and c	Through center perpendicular to face ab (parallel to edge c)	$m\frac{a^2+b^2}{12}$
(10)	Sphere, radius r	Any diameter	$m\frac{2}{5}r^2$
(11)	Spherical shell, external radius r_1, internal radius r_2	Any diameter	$m\frac{2}{5}\frac{r_1^5-r_2^5}{r_1^3-r_2^3}$
(12)	Spherical shell, very thin, mean radius r	Any diameter	$m\frac{2}{3}r^2$
(13)	Right circular cylinder of radius r, length l	Longitudinal axis of the slide	$m\frac{r^2}{2}$
(14)	Right circular cylinder of radius r, length l	Transverse diameter	$m\left(\frac{r^2}{4}+\frac{l^2}{12}\right)$
(15)	Hollow circular cylinder, radii r_1 and r_2, length l	Longitudinal axis of the figure	$m\frac{r_1^2+r_2^2}{2}$
(16)	Thin cylindrical shell, length l, mean radius r	Longitudinal axis of the figure	mr^2
(17)	Hollow circular cylinder, radii r_1 and r_2, length l	Transverse diameter	$m\left(\frac{r_1^2+r_2^2}{4}+\frac{l^2}{12}\right)$
(18)	Hollow circular cylinder, very thin, length l, mean radius r	Transverse diameter	$m\left(\frac{r^2}{2}+\frac{l^2}{12}\right)$
(19)	Elliptic cylinder, length l, transverse semiaxes a and b	Longitudinal axis	$m\left(\frac{a^2+b^2}{4}\right)$
(20)	Right cone, altitude h, radius of base r	Axis of the figure	$m\frac{3}{10}r^2$

10.17 MUSIC

Music principles do not depend on a specific key; the key of C is used in examples.

Definitions

1. Map the pitches to numbers from 0 to 11 as follows:
   - $0 = C$
   - $1 = C\sharp = D\flat$
   - $2 = D$
   - $3 = D\sharp = E\flat$
   - $4 = E$
   - $5 = F$
   - $6 = F\sharp = G\flat$
   - $7 = G$
   - $8 = G\sharp = A\flat$
   - $9 = A$
   - $10 = A\sharp = B\flat$
   - $11 = B$
2. A *pitch class set* (or pcset), shown as $\{\ldots\}$, is an unordered set of numbers representing notes. For example, the C major chord $\{C, E, G\}$ is $\{0, 4, 7\}$.
3. A *pitch class segment* (or pcseg), shown as $\langle \ldots \rangle$, is an ordered set of numbers representing notes. For example, the first part of the main theme in Haydin's Surprise Symphony is $\langle C, C, E, E, G, G, E, F, F, D, D, B, B, G\rangle = \langle 0, 0, 4, 4, 7, 7, 4, 5, 5, 2, 2, 11, 11, 7\rangle$.
4. Define the *transposition by* n operator T_n by $T_n(x) = x + n \pmod{12}$. For example, the transposition of a C major pcset by 7 steps is
$$\begin{aligned} T_7(\text{C major pcset}) &= T_7(\{C, E, G\}) = T_7(\{0, 4, 7\}) \\ &= \{T_7(0), T_7(4), T_7(7)\} = \{7, 11, 2\} = \{G, B, D\} \\ &= \text{G major chord} \end{aligned}$$
5. Define the *inversion about* n operator I_n by $I_n(x) = -x + n \pmod{12}$.
6. Given a specific pcseg, define the collection of transpositions of it to be *prime forms*, define the collection of inversions of it to be *inverted forms*. Let S be the collection of all prime and inverted forms.
   For example, given the pcseg $\langle G, C, D\rangle = \langle 7, 0, 2\rangle$, S has 24 elements consisting of the following:

Prime forms	Inverted forms
$T_0\langle 7,0,2\rangle = \langle 7,0,2\rangle$	$I_0\langle 7,0,2\rangle = \langle 5,0,10\rangle$
$T_1\langle 7,0,2\rangle = \langle 8,1,3\rangle$	$I_1\langle 7,0,2\rangle = \langle 6,1,11\rangle$
$T_2\langle 7,0,2\rangle = \langle 9,2,4\rangle$	$I_2\langle 7,0,2\rangle = \langle 7,2,0\rangle$
$\vdots$	$\vdots$

7. Given a 3-element pcseg x, define the operators P, L, and R as follows:
   (a) P (parallel) is the form opposite in type to x with the first and third notes switched: $P(\langle a, b, c\rangle) = I_{a+c}(\langle a, b, c\rangle)$.
   (b) L (leading tone exchange) is the form opposite in type to x with the second and third notes switched: $L(\langle a, b, c\rangle) = I_{b+c}(\langle a, b, c\rangle)$.
   (c) R (relative) is the form opposite in type to x with the first and second notes switched: $R(\langle a, b, c\rangle) = I_{a+b}(\langle a, b, c\rangle)$.

   For example:
   - $P\langle 0,4,7\rangle = \langle 7,3,0\rangle$ and $P\langle 3,11,8\rangle = \langle 8,0,3\rangle$.
   - $L\langle 0,4,7\rangle = \langle 11,7,4\rangle$ and $L\langle 3,11,8\rangle = \langle 4,8,11\rangle$.
   - $R\langle 0,4,7\rangle = \langle 4,0,9\rangle$ and $R\langle 3,11,8\rangle = \langle 1,3,6\rangle$.

EXAMPLE Starting with the C major chord $\langle 0,4,7\rangle$, the set S consists of the following:

Prime forms	Inverted forms
$C = \langle 0,4,7\rangle$	$\langle 0,8,5\rangle = f$
$C\sharp = D\flat = \langle 1,5,8\rangle$	$\langle 1,9,6\rangle = f\sharp = g\flat$
$D = \langle 2,6,9\rangle$	$\langle 2,10,7\rangle = g$
$D\sharp = E\flat = \langle 3,7,10\rangle$	$\langle 3,11,8\rangle = g\flat = a\flat$
$E = \langle 4,8,11\rangle$	$\langle 4,0,9\rangle = a$
$F = \langle 5,9,0\rangle$	$\langle 5,1,10\rangle = a\flat = b\flat$
$F\sharp = G\flat = \langle 6,10,1\rangle$	$\langle 6,2,11\rangle = b$
$G = \langle 7,11,2\rangle$	$\langle 7,3,0\rangle = c$
$G\sharp = A\flat = \langle 8,0,3\rangle$	$\langle 8,4,1\rangle = c\flat = d\flat$
$A = \langle 9,1,4\rangle$	$\langle 9,5,2\rangle = d$
$A\sharp = B\flat = \langle 10,2,5\rangle$	$\langle 10,6,3\rangle = d\flat = e\flat$
$B = \langle 11,3,6\rangle$	$\langle 11,7,4\rangle = e$

where major chords have been given a capital letter and minor chords have been given a lowercase letter.

Results

1. The transposition and inversions operators can be interpreted to apply to the set S of forms. In this context, they satisfy the following rules (where "$\circ$" represents composition):
$$\begin{aligned} T_m \circ T_n &= T_{m+n} \\ T_m \circ I_n &= I_{m+n} \\ I_m \circ T_n &= I_{m-n} \\ I_m \circ I_n &= T_{m-n} \end{aligned}$$
where the indices are interpreted mod 12. The collection of 12 transposition operators and 12 inversion operators form a group called the T/I group.
2. The P, L, and R operators under composition form a group called the PLR group.
3. The T/I group and the PLR groups are each isomorphic to the dihedral group of order 24.
4. Consider the group of all permutations of S. In this larger group:
   (a) the group T/I is the centralizer[2] of the PLR group; and
   (b) the group PLR is the centralizer of the T/I group.
5. The P, L, and R function have musical significance:
   (a) The function P takes a chord and maps it to its parallel minor or major. For example $P(C \text{ major}) = c$ minor and $P(c \text{ minor}) = C$ major.
   (b) The function L is a leading tone exchange; $L(C \text{ major}) = e$ minor.
   (c) The function R takes a chord and maps it to its relative minor or major. For example $R(C \text{ major}) = a$ minor and $R(a \text{ minor}) = C$ major.

[2]The centralizer of a subgroup H of a group G is the set of elements of G which commute with all elements of H.

10.18 OPERATIONS RESEARCH

Operations research integrates mathematical modeling and analysis with engineering in an effort to design and control systems.

10.18.1 LINEAR PROGRAMMING

Linear programming (LP) is a technique for modeling problems with linear objective functions and linear constraints. The *standard form* for an LP model with n decision variables and m resource constraints is

$$\text{Minimize} \quad \sum_{j=1}^{n} c_j x_j \qquad \qquad (\textit{objective function}),$$

$$\text{Subject to} \quad \begin{cases} x_j \geq 0 \quad \text{for } j = 1, \cdots, n, & \text{(non-negativity requirement)}, \\ \displaystyle\sum_{j=1}^{n} a_{ij} x_j = b_i \quad \text{for } i = 1, \cdots, m, & (\textit{constraint functions}), \end{cases}$$

where x_j is the amount of decision variable j used, c_j is decision $j's$ per unit contribution to the objective, a_{ij} is decision $j's$ per unit usage of resource i, and b_i is the total amount of resource i to be used.

Let $\mathbf{x}$ represent the $(n \times 1)$ vector $(x_1, \cdots, x_n)^{\mathrm{T}}$, $\mathbf{c}$ the $(n \times 1)$ vector $(c_1, \cdots, c_n)^{\mathrm{T}}$, $\mathbf{b}$ the $(m \times 1)$ vector $(b_1, \cdots, b_m)^{\mathrm{T}}$, A the $(m \times n)$ matrix (a_{ij}), and $\mathbf{A}_{\cdot j}$ the $(n \times 1)$ column of A associated with x_j. Then the standard model, in matrix notation, is "minimize $\mathbf{c}^{\mathrm{T}}\mathbf{x}$ subject to $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$." A vector $\mathbf{x}$ is called *feasible* if and only if $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} \geq \mathbf{0}$.

10.18.1.1 Modeling in LP

LP is an appropriate modeling technique if the following four assumptions are satisfied by the situation:

1. All data coefficients are known with certainty.
2. There is a single objective.
3. The problem relationships are linear functions of the decisions.
4. The decisions can take on continuous values.

Here are two examples of LP modeling:

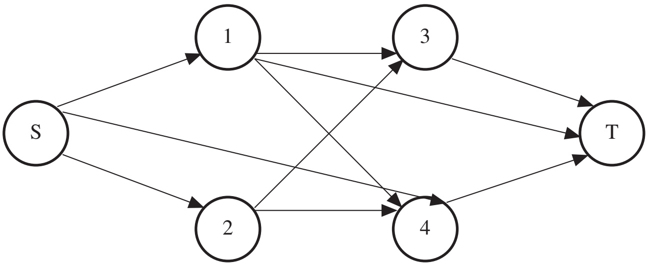
1. *Product mix problem* — Consider a company with three products to sell. Each product requires four operations and the per unit data are given as follows:

Product	Drilling	Assembly	Finishing	Packing	Profit
A	2	3	1	2	45
B	3	6	2	4	90
C	2	1	4	1	55
Hours available	480	960	540	320	

Let x_A, x_B, and x_C represent the number of units of A, B, and C manufactured daily. A model to maximize profit subject to the labor restrictions is

$$\begin{array}{ll} \text{Maximize} & 45x_A + 90x_B + 55x_C \qquad \text{(total profit)}, \\ \text{Subject to:} & \begin{cases} 2x_A + 3x_B + 2x_C \leq 480 & \text{(drilling hours)}, \\ 3x_A + 6x_B + 1x_C \leq 960 & \text{(assembly hours)}, \\ 1x_A + 2x_B + 4x_C \leq 540 & \text{(finishing hours)}, \\ 2x_A + 4x_B + 1x_C \leq 320 & \text{(packing hours)}, \\ x_A \geq 0, \quad x_B \geq 0, \quad x_C \geq 0. \end{cases} \end{array}$$

2. *Maximum flow through a network* — Consider the directed network below. Nodes S (and T) are the source (and terminal) nodes. On each arc material up to the arc capacity C_{ij} can be shipped. Material is neither created nor destroyed at nodes other than S and T. The goal is to maximize the amount of material that can be shipped through the network from S to T. If x_{ij} represents the amount of material shipped from node i to node j, a model that determines the maximum flow is shown below.



$$\begin{array}{ll} \text{Maximize} & x_{1T} + x_{3T} + x_{4T}, \\ \text{Subject to} & \begin{cases} x_{S1} = x_{13} + x_{14} + x_{1T} & \text{(node 1 conservation)}, \\ x_{S2} = x_{23} + x_{24} & \text{(node 2 conservation)}, \\ x_{13} + x_{23} = x_{3T} & \text{(node 3 conservation)}, \\ x_{S4} + x_{24} + x_{14} = x_{4T} & \text{(node 4 conservation)}, \\ 0 \leq x_{ij} \leq C_{ij} \quad \text{for all pairs } (i,j) & \text{(arc capacity)}. \end{cases} \end{array}$$

10.18.1.2 Transformation to standard form

Any LP model can be transformed to *standard form* as follows:

1. If the problem has "maximize objective," change it to "minimize objective" and multiply every c_j by -1
2. If the problem has "$\leq$" constraints, change them to "$=$" constraints by introducing non-negative *slack variables* $\{S_i\}$ to $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$.

   That is, the new formulation is: $\sum_{j=1}^{n} a_{ij}x_j + S_i = b_i$ with $S_i \geq 0$.

3. If the problem has "$\geq$" constraints, change them to "$=$" constraints by introducing non-negative *surplus variables* $\{U_i\}$ to $\sum_{j=1}^{n} a_{ij}x_j \geq b_i$.

   That is, the new formulation is: $\sum_{j=1}^{n} a_{ij}x_j - U_i = b_i$ with $U_i \geq 0$.

All slack and surplus variables have $c_j = 0$.

10.18.1.3 Solving LP models: simplex method

Assume that there is at least one feasible **x** vector, and that A has rank m. Geometrically, because all constraints are linear, the set of feasible **x** forms a convex polyhedral set (bounded or unbounded) that must have at least one extreme point. The motivation for the simplex method for solving LP models is the following:

> For any LP model with a bounded optimal solution, an optimal solution exists at an extreme point of the feasible set.

Given a feasible solution **x**, let $\mathbf{x}^B$ be the components of **x** with $x_j > 0$ and $\mathbf{x}^N$ be the components with $x_j = 0$. Associated with $\mathbf{x}^B$, define B as the columns of A associated with each x_j in $\mathbf{x}^B$. For example, if x_2, x_4, and S_1 are positive in **x**, then $x^B = (x_2, x_4, S_1)^{\mathrm{T}}$, and B is the matrix with columns $\mathbf{A}_{\cdot 2}, \mathbf{A}_{\cdot 4}, \mathbf{A}_{\cdot S_1}$. Define N as the remaining columns of A, i.e., those associated with $\mathbf{x}^N$. A *basic feasible solution* (BFS) is a feasible solution where the columns of B are linearly independent. The following theorem relates a BFS with extreme points:

> A feasible solution **x** is at an extreme point of the feasible region if and only if **x** is a BFS.

The following simplex method finds an optimal solution to the LP by finding the optimal partition of **x** into $\mathbf{x}^B$ and $\mathbf{x}^N$:

Step (1) Find an initial basic feasible solution. Define $\mathbf{x}^B$, $\mathbf{x}^N$, B, N, $\mathbf{c}^B$, and $\mathbf{c}^N$ as above.

Step (2) Compute the vector $\mathbf{c}' = (\mathbf{c}^N - \mathbf{c}^B B^{-1} N)$. If $\mathbf{c}'_j \geq \mathbf{0}$ for all j, then stop; the solution $\mathbf{x}^B = B^{-1}\mathbf{b}$ is optimal with objective value $\mathbf{c}^B B^{-1}\mathbf{b}$. Otherwise, select the variable x_j in $\mathbf{x}^N$ with the most negative c'_j value, and go to Step (3).

Step (3) Compute $\mathbf{A}'_{\cdot j} = B^{-1}\mathbf{A}_{\cdot j}$. If $A'_{\cdot j} \leq \mathbf{0}$ for all j, then stop; the problem is unbounded and the objective can decrease to $-\infty$. Otherwise, compute $\mathbf{b}' = B^{-1}\mathbf{b}$ and find $\min_{i | a'_{ij} > 0} \frac{b'_i}{a'_{ij}}$. Assume the minimum ratio occurs in row r. Insert x_j into the r^{th} position of $\mathbf{x}^B$, take the variable that was in this position, and move it to the $\mathbf{x}^N$ partition. Update B, N, $\mathbf{c}^B$, and $\mathbf{c}^N$ accordingly. Go to Step (2).

Ties in the selections in Steps (2) and (3) can be broken arbitrarily. The unboundedness signal suggests that the model is missing constraints or that there has been an incorrect data entry or computational error, because, in real problems, a profit or cost cannot be unbounded. For maximization problems, only Step (2) changes. The solution is optimal when $c'_j \leq 0$ for all j, then choose the variable with the maximum c'_j value to move into $\mathbf{x}^B$. Effective methods for updating B^{-1} in each iteration of Step (3) exist to ease the computational burden.

To find an initial basic feasible solution define a variable A_i and add this variable to the left-hand side of constraint i transforming to $\sum_{j=1}^{n} a_{ij} + A_i = b_i$. The new constraint is equivalent to the original constraint if and only if $A_i = 0$. Also, because A_i appears only in constraint i and there are m A_i variables, the columns corresponding to the A_i variables are of rank m. We now solve a "new" LP model with the adjusted constraints and the new objective "minimize $\sum_{i=1}^{m} A_i$." If the optimal solution to this new model is 0, the solution is a basic feasible solution to the original problem and we can use it in step (1). Otherwise, no basic feasible solution exists for the original problem.

10.18.1.4 Solving LP models: interior point method

An alternative way to determine extreme points is to cut through the middle of the polyhedron and go directly towards the optimal solution. Karmarkar's method assumes that the LP model has the following form:

$$\text{Minimize} \quad \sum_{j=1}^{n} c_j x_j,$$
$$\text{Subject to} \quad \begin{cases} \displaystyle\sum_{j=1}^{n} a_{ij} x_j = 0 & \text{for } i = 1, \cdots, m, \\ \displaystyle\sum_{j=1}^{n} x_j = 1, & \\ x_j \geq 0, & \text{for } j = 1, \cdots, n. \end{cases} \tag{10.18.1}$$

Also, assume that the optimal objective value is 0 and that $x_j = 1/n$ for $j = 1, \cdots, n$ is feasible. Any model can be transformed so that these assumptions hold.

The following *centering transformation*, relative to the k^{th} estimate of solution vector $\mathbf{x}^k$, takes a feasible solution vector $\mathbf{x}$ and transforms it to $\mathbf{y}$ such that $\mathbf{x}^k$ is transformed to the center of the feasible simplex: $\mathbf{y}_j = \dfrac{x_j/x_j^k}{\sum_{r=1}^{n}(x_r/x_r^k)}$. Let $\text{Diag}(x^k)$ represent an $n \times n$ matrix with off-diagonal entries equal to 0 and the diagonal entry in row j equal to x_j^k. The formal algorithm is as follows:

Step (1) Initialize $x_j^0 = 1/n$ and set the iteration count $k = 0$.

Step (2) If $\displaystyle\sum_{j=1}^{n} x_j^k c_j$ is sufficiently close to 0, then stop; $\mathbf{x}^k$ is optimal. Otherwise go to Step (3).

Step (3) Move from the center of the transformed space in an improving direction using

$$\mathbf{y}^{k+1} = \left[\frac{1}{n}, \frac{1}{n}, \cdots, \frac{1}{n}\right]^{\mathrm{T}} - \frac{\theta(I - P^T(PP^T)^{-1}P)[\mathrm{Diag}(\mathbf{x}^k)]\mathbf{c}^{\mathrm{T}}}{||C_p||\sqrt{n(n-1)}},$$

where $||C_p||$ is the length of the vector $(I - P^T(PP^T)^{-1}P)[\mathrm{Diag}(\mathbf{x}^k)]\mathbf{c}^T$, P is an $(m+1) \times n$ matrix whose first m rows are $A[\mathrm{Diag}(\mathbf{x}^k)]$ and whose last row is a vector of $1's$, and θ is a parameter between 0 and 1.

Step (4) Find the new point $\mathbf{x}^{k+1}$ in the original space by applying the inverse transformation of the centering transformation to $\mathbf{y}^{k+1}$. Set $k = k + 1$ and go to Step (2).

The method is guaranteed to converge to the optimal solution when $\theta = \frac{1}{4}$ is used.

10.18.2 DUALITY AND COMPLEMENTARY SLACKNESS

Define y_i as the *dual variable (shadow price)* representing the purchase price for a unit of resource i. The *dual problem* to the primal model (maximize objective, all constraints of the form "$\leq$") is

$$\begin{aligned}
&\text{Minimize} \quad \sum_{i=1}^{m} b_i y_i, \\
&\text{Subject to} \quad \begin{cases} \displaystyle\sum_{i=1}^{m} a_{ij} y_i \geq c_j & \text{for } j = 1, \dots, n, \\ \qquad y_i \geq 0, & \text{for } i = 1, \dots, m. \end{cases}
\end{aligned} \tag{10.18.2}$$

The following results link the dual model (minimization) with its primal model (maximization).

1. *Weak duality theorem*: If that $\mathbf{x}$ and $\mathbf{y}$ are feasible solutions to the respective primal and dual problems, then $\displaystyle\sum_{j=1}^{n} c_j x_j \leq \sum_{i=1}^{m} b_i y_i$.
2. *Strong duality theorem*: If the primal model has a finite optimal solution $\mathbf{x}^*$, then the dual has a finite optimal solution $\mathbf{y}^*$, and $\displaystyle\sum_{j=1}^{n} c_j x_j^* = \sum_{i=1}^{m} b_i y_i^*$.
3. *Complementary slackness theorem*: If $\mathbf{x}$ and $\mathbf{y}$ are feasible solutions to the respective primal and dual problems. Then, $\mathbf{x}$ is optimal for the primal and $\mathbf{y}$ is optimal for the dual if and only if:

$$y_i \cdot \left(b_i - \sum_{j=1}^{n} a_{ij} x_j\right) = 0, \qquad \text{for } i = 1, \cdots, m, \quad \text{and}$$

$$x_j \cdot \left(\sum_{i=1}^{m} a_{ij} y_i - c_j\right) = 0, \qquad \text{for } j = 1, \cdots, n.$$

10.18.3 LINEAR INTEGER PROGRAMMING

Linear integer programming models result from restricting the decisions in linear programming models to be integer valued. The standard form is

$$\begin{aligned}
&\text{Minimize} \quad \sum_{j=1}^{n} c_j x_j \qquad \text{(objective function)},\\
&\text{Subject to} \quad \begin{cases} x_j \geq 0, & \text{and integer for } j = 1, \ldots, n, \\ \displaystyle\sum_{j=1}^{n} a_{ij} x_j = b_i & \text{for } i = 1, \ldots, m \quad \text{(constraint functions).} \end{cases}
\end{aligned} \tag{10.18.3}$$

As long as the variable values are bounded, then the general model can be transformed into a model where all variable values are restricted to either 0 or 1. Hence, algorithms that solve 0–1 integer programming models are usually sufficient.

10.18.3.1 Branch and bound

Branch and bound implicitly enumerates all feasible integer solutions to find the optimal solution. The main idea is to break the feasible set into subsets (branching) and then evaluate the best solution in each subset or determine that the subset cannot contain the optimal solution (bounding). When a subset is evaluated, it is said to be *fathomed*. The following algorithm performs the branching by partitioning on variables with fractional values and uses a linear programming relaxation to generate a bound on the best solution in a subset:

Step (1) Assume that a feasible integer solution, called the *incumbent*, is known whose objective function value is z (initially, z may be set to infinity if no feasible solution is known). Set p, the subset counter equal to 1. Make the original model be the first problem in the subset list.

Step (2) If $p = 0$, then stop. The incumbent solution is the optimal solution. Otherwise go to Step (3).

Step (3) Solve the LP relaxation of the p^{th} problem in the subset list (allow all integer valued variables to take on continuous values). Denote the LP objective value by v. If $v \geq z$ or the LP is infeasible, then set $p = p - 1$ (fathom by bound or infeasibility), and go to Step (2). If the LP solution is integer valued, then update the incumbent to the LP solution, set $z = \min(z, v)$ and $p = p - 1$, and go to Step (2). Otherwise, go to Step (4).

Step (4) Take any variable x_j with fractional value in the LP solution. Replace problem p with two problems created by individually adding the constraints $x_j \leq \lfloor x_j \rfloor$ and $x_j \geq \lceil x_j \rceil$ to problem p. Add these two problems to the bottom of the subset list replacing the p^{th} problem, set $p = p + 1$, and go to Step (2).

10.18.4 NETWORK FLOW METHODS

A *network* consists of N (a set of nodes) and A (a set of arcs). Each arc (i, j) defines a connection from node i to node j. Depending on the application, arc (i, j) may have an associated cost and upper and lower capacity on flow.

Decision problems on networks can often be modeled using linear programming models, and these models usually have the property that solutions from the simplex method are integer valued (the *total unimodularity property*). Because of the underlying graphical structure, more efficient algorithms are also available. We present the augmenting path algorithm for the maximum flow problem and the Hungarian method for the assignment problem.

10.18.4.1 Maximum flow

Let x_{ij} represent the flow on arc (i, j), c_{ij} the flow capacity of (i, j), S be the source node, and T be the terminal node. The maximum flow problem is to ship as much flow from S to T without violating the capacity on any arc, and all flow sent into a node must leave that node (except for S and T). The following algorithm solves the problem by continually adding flow-carrying paths until no path can be found:

Step (1) Initialize $x_{ij} = 0$ for all (i, j).

Step (2) Find a flow-augmenting path from S to T using the following labeling method. Start by labeling S with a^*. From any labeled node i, label node j with the label i if j is unlabeled and $x_{ij} < c_{ij}$ (*forward labeling arc*). From any labeled node i, label node j with the label i if j is unlabeled and $x_{ji} > 0$ (*backward labeling arc*). Perform labeling until no additional nodes can be labeled. If T cannot be labeled, then stop. The current x_{ij} values are optimal. Otherwise, go to Step (3).

Step (3) There is a path from S to T where flow is increased on the forward labeling arcs, decreased on the backward labeling arcs, and gets more flow from S to T. Let F be the minimum of $c_{ij} - x_{ij}$ over all forward labeling arcs and of x_{ij} over all backward labeling arcs. Set $x_{ij} = x_{ij} + F$ for the forward arcs and $x_{ij} = x_{ij} - F$ for the backward arcs. Go to Step (2).

The algorithm terminates with a set of arcs with $x_{ij} = c_{ij}$ and if these are deleted, then S and T are in two disconnected pieces of the network. The algorithm finds the maximum flow by finding the minimum capacity set of arcs that disconnects S and T (minimum capacity cutset).

10.18.5 ASSIGNMENT PROBLEM

Consider a set J of m jobs and a set I of m employees. Each employee can do 1 job, and each job must be done by 1 employee. If job j is assigned to employee i, then the cost to the company is c_{ij}. The problem is to assign employees to jobs to minimize the overall cost.

This problem can be formulated as an optimization problem on a bipartite graph where the jobs are one part and the employees are the other. Let C be the $m \times m$ matrix of costs c_{ij}. The following algorithm solves for the optimal assignment.

Step (1) Find $l_i = \min_j c_{ij}$ for each row i. Let $c_{ij} = c_{ij} - l_i$. Find $n_j = \min_i c_{ij}$ for each column j. Let $c_{ij} = c_{ij} - n_j$.

Step (2) Construct a graph with nodes for S, T, and each element of the sets J and I. Construct an arc from S to each node in I, and set its capacity to 1. Construct an arc from each node in J to T, and set its capacity to 1. If $c_{ij} = 0$, then construct an arc from $i \in I$ to $j \in J$, and set its capacity to 2. Solve a maximum flow problem on the constructed graph. If m units of flow can go through the network, then stop. The maximum flow solution on the arcs between I and J represents the optimal assignment. Otherwise, go to Step (3).

Step (3) Update C using the following rules based on the labels in the solution to the maximum flow problem: Let L_I and L_J be the set of elements of I and J, respectively, with labels when the maximum flow algorithm terminates. Let $\delta = \min_{i \in L_I, j \in (J - L_J)} c_{ij}$; note that $\delta > 0$. For $i \in L_I$ and $j \in (J - L_J)$, set $c_{ij} = c_{ij} - \delta$. For $i \in (I - L_I)$ and $j \in L_J$, set $c_{ij} = c_{ij} + \delta$. Leave all other c_{ij} values unchanged. Go to Step (2).

In Step (3), the algorithm creates new arcs, eliminates some unused arcs, and leaves unchanged arcs with $x_{ij} = 1$. When returning to Step (2), you can solve the next maximum flow problem by adding and deleting the appropriate arcs and starting with the flows and labels of the preceding execution of the maximum flow algorithm.

10.18.6 SHORTEST PATH PROBLEM

Consider a network (N, A) where N is the set of nodes, A the set of arcs, and d_{ij} represents the "distance" of traveling on arc (i, j) (if no arc exists between i and j, $d_{ij} = \infty$). For any two nodes R and S, the shortest path problem is to find the shortest distance route through the network from R to S. Let the state space be N and a stage representing travel along one arc. $f^*(i)$ is the optimal distance from node i to S. The resultant recursive equations are $f^*(i) = \min_{j \in N} \left[d_{ij} + f^*(j)\right], \qquad \text{for all } i.$.

Dijkstra's algorithm can be used successively to approximate the solution to the equations when $d_{ij} > 0$ for all (i, j).

Step (1) Set $f^*(S) = 0$ and $f^*(i) = d_{iS}$ for all $i \in N$. Let P be the set of permanently labeled nodes; $P = \{S\}$. Let $T = N - P$ be the set of temporarily labeled nodes.

Step (2) Find $i \in T$ with $f^*(i) = \min_{j \in T} f^*(j)$. Set $T = T - \{i\}$ and $P = P \cup \{i\}$. If $T = \emptyset$ (the empty set), then stop; $f^*(R)$ is the optimal path length. Otherwise, go to Step (3).

Step (3) Set $f^*(j) = \min[f^*(j), f^*(i) + d_{ij}]$ for all $j \in T$. Go to Step (2).

10.18.7 HEURISTIC SEARCH TECHNIQUES

Heuristic search techniques are commonly used for combinatorial optimization problems. A method starts with an initial vector $\mathbf{x}_0$ and attempts to find improved solutions. Define $\mathbf{x}$ to be the current solution vector, $f(\mathbf{x})$ to be its objective value and $N(\mathbf{x})$ to be its neighborhood. For the remainder of this section, assume that we seek the minimum value of $f(\mathbf{x})$. In each iteration of a neighborhood search algorithm, we generate a number of solutions $\mathbf{x}' \in N(\mathbf{x})$, and compare $f(\mathbf{x})$ with each $f(\mathbf{x}')$. If $f(\mathbf{x}')$ is a better value than $f(\mathbf{x})$, then we update the current solution to $\mathbf{x}'$ (termed *accepting* $\mathbf{x}'$) and we iterate again searching from $\mathbf{x}'$. The process continues until none of the generated neighbors of the current solution yield lower solutions. Common improvement procedures such as *pairwise interchange* and "k-opt," are specific instances of neighborhood search methods. Note that at each iteration, we move to a better solution or we terminate with the best solution seen so far.

The key issues involved in designing a neighborhood search method are the definition of the neighborhood of $\mathbf{x}$, and the number of neighbors to generate in each iteration. When the neighborhood of $\mathbf{x}$ is easily computed and evaluated, then one can generate the entire neighborhood to ensure finding a better solution if one exists in the neighborhood. It is also possible to generate only a portion of the neighborhood (however, this could lead to premature termination with a poorer solution). Generally, deterministic neighborhood search can only guarantee finding a local minimum solution to the optimization problem. One can make multiple runs, each with different $\mathbf{x}_0$, to increase the chances of finding the global minimum solution.

10.18.7.1 Simulated annealing (SA)

Simulated annealing is a neighborhood search method that uses randomization to avoid terminating at a locally optimal point. In each iteration of SA, a single neighbor $\mathbf{x}'$ of $\mathbf{x}$ is generated. If $f(\mathbf{x}') \leq f(\mathbf{x})$, then $\mathbf{x}'$ is accepted. Otherwise, $\mathbf{x}'$ is accepted with a probability that depends upon $f(\mathbf{x}') - f(\mathbf{x})$, and a non-stationary control parameter. Define the following:

1. C_k is the k^{th} control parameter
2. L_k is the maximum number of neighbors evaluated while the k^{th} control parameter is in use
3. I_k is the counter for the number of solutions currently evaluated at the k^{th} control parameter

To initialize the algorithm, assume that we have an initial value $\mathbf{x}_0$ and sequences $\{C_k\}$ (termed the *cooling schedule*) and $\{L_k\}$ such that $C_k \to 0$ as $k \to \infty$. The SA algorithm to minimize $f(\mathbf{x})$ is:

Step (1) Set $\mathbf{x} = \mathbf{x}_0$, $k = 1$ and $I_k = 0$.

Step (2) Generate a neighbor $\mathbf{x}'$ of $\mathbf{x}$, compute $f(\mathbf{x}')$ and increment I_k by 1.

Step (3) If $f(\mathbf{x}') \leq f(\mathbf{x})$, then replace $\mathbf{x}$ with $\mathbf{x}'$ and go to Step (5).

Step (4) If $f(\mathbf{x}') > f(\mathbf{x})$, then with probability $e^{(f(\mathbf{x})-f(\mathbf{x}'))/C_k}$, replace $\mathbf{x}$ with $\mathbf{x}'$.

Step (5) If $I_k = L_k$, then increment k by 1, reset $I_k = 0$ and check for the termination criterion. If the termination criterion is met, then stop, $\mathbf{x}$ is the solution, if not then go to Step (2).

The algorithm terminates when C_k approaches 0, or there has been no improvement in the solution over a number of C_k values. In the above description only the current solution is stored; not the best solution encountered. In practice however, in Step (2), we also compare $f(\mathbf{x}')$ with an incumbent solution. If the incumbent solution is worse, then we replace the incumbent with $\mathbf{x}'$, otherwise we retain the current incumbent. Note that SA only generates one neighbor of $\mathbf{x}$ in each iteration. However, we may not accept $\mathbf{x}'$ and therefore may generate another neighbor of $\mathbf{x}$.

Several issues must be resolved when implementing SA. If C_k goes to zero too quickly, then the algorithm can easily get stuck in a local minimum solution. If C_1 is large and C_k tends to zero too slowly, then the algorithm requires more computation to achieve convergence. Under various technical conditions, there are convergence proofs for this method.

10.18.7.2 Tabu search (TS)

Tabu search is similar to simulated annealing. A problem in SA is that one can start to climb out of a local minimum solution, only to return via a sequence of "better" solutions that leads directly back to where the algorithm has already searched. Also, the neighborhood structure may permit moving to areas where one knows that no optimal solutions exist. Finally, a modeler may have insight on where to look for good solutions and SA cannot easily enable searching of specific areas. Tabu search tries to remedy each of these deficiencies.

The key terminology in TS is $S(\mathbf{x})$, the set of moves from $\mathbf{x}$. This is similar to the neighborhood of $\mathbf{x}$, and is all the solutions that you can get to from $\mathbf{x}$ in one *move*. A "move" is similar to an SA "step," but is more general because it can be applied to both continuous and discrete variable problems. Let s be a move in $S(\mathbf{x})$. For example, consider a 5-city traveling salesman problem where $\mathbf{x}$ is a vector containing the sequence of cities visited, including the return to the initial city. Let $\mathbf{x} = (2, 1, 5, 3, 4, 2)$. If we define a move to be an "adjacent pairwise interchange," then the set of moves $S(\mathbf{x})$ is:

$$\{\{(1, 2, 5, 3, 4, 1), (2, 5, 1, 3, 4, 2), (2, 1, 3, 5, 4, 2), (2, 1, 5, 4, 3, 2), (4, 1, 5, 3, 2, 4)\}$$

As another example, if we use a standard non-linear programming direction–step size search algorithm, then one can construct a family of moves of the form $S(\mathbf{x}) = \mathbf{x} + u\mathbf{d}$. Here, u is a step size scalar and $\mathbf{d}$ is the direction of movement and the family depends on the values of u and $\mathbf{d}$ selected.

To run a Tabu search, define T as the *tabu set*; these are a set of moves that the method should not use. Define OPT to be the function that selects a particular $s \in S(\mathbf{x})$ that creates an eventual improvement in the objective. Then

Step (1) Start with initial incumbent $\mathbf{x}_0$. Set $\mathbf{x} = \mathbf{x}_0$, $k = 0$, $T = \phi$.

Step (2) If $S(\mathbf{x}) - T = \phi$, then go to Step (4). Otherwise set $k = k + 1$ and select $s_k \in S(\mathbf{x}) - T$ such that $s_k(\mathbf{x}) = OPT(s \mid s \in S(\mathbf{x}) - T)$.

Step (3) Let $\mathbf{x} = s_k(\mathbf{x})$. If $f(\mathbf{x}) < f(\mathbf{x}_0)$, then $\mathbf{x}_0 = \mathbf{x}$.

Step (4) If a chosen number of iterations has elapsed either in total or since $\mathbf{x}_0$ was last improved, or if $S(\mathbf{x}) - T = \phi$ from Step (2), then stop. Otherwise, update T (if necessary) and go to Step (2).

The method is more effective if the user understands the solution space and can guide the search somewhat. Often the tabu list contains solutions that were previously visited or solutions that would reverse properties of good solutions. Early in the method, it is important that the search space is evaluated in a coarse manner so that one does not skip an area where the optimal solution is located. Tabu search can move to inferior solutions temporarily when OPT returns a solution that has a worse objective value than $f(\mathbf{x})$ and, in fact, this happens every time when $\mathbf{x}$ is a local minimum solution.

10.18.7.3 Genetic algorithms

A different approach to heuristically solving difficult combinatorial optimization problems mimics evolutionary theory within an algorithmic process. A population of individuals is represented by K various feasible solutions $\mathbf{x}_k$ for $k = 1, \dots, K$. The collection of such solutions at any iteration of a genetic algorithm is referred to as a *generation*, and the individual elements of each solution $\mathbf{x}_k$ are called *chromosomes*. For instance, in the context of the traveling salesman problem, a generation would consist of a set of traveling salesman tours, and the chromosomes of an individual solution would represent cities.

To continue drawing parallels with the evolutionary process, each iteration of a genetic algorithm creates a new generation by computing new solutions based on the previous population. More specifically, an individual of the new generation is created from parent solutions by a *crossover operator*. The crossover operator describes how a solution is created by combining characteristics of the parent solutions. The selection of the crossover operator is a key aspect of designing an algorithm.

The rules for composing a new generation differ among implementations, but often consist of selecting some of the best solutions from the previous generation along with some new solutions created by crossovers from the previous generation. Additionally, these solutions may mutate from generation to generation in order to introduce new elements and chromosomal patterns into the population. The objective value of each new solution in the new generation is computed, and the best solution found thus far in the algorithm is updated if applicable. The creation of the new generation of solutions concludes a genetic algorithm iteration. The algorithm stops once some termination criteria is reached (e.g., after a specified number of generations are evolved, or perhaps if no new best solution was recorded in the last q generations). A typical genetic algorithm for minimization is:

Step (1) Choose a population size K, a maximum number of generations Q, and a number of survivors $S < K$, where K, Q, and S are integers. Also, choose a mutation probability p (typically, p is small, perhaps close to 0.05). Create an initial set of solutions $\mathbf{x}_k$, for $k = 1, \dots, K$, and define Generation 0 to be these solutions. Calculate the objective function of each solution in Generation 0, and let $\mathbf{x}^*$ with objective function $f(\mathbf{x}^*)$ denote the best such solution. Initialize the generation counter $i = 0$.

Step (2) Copy the best (according to objective function value) S solutions from Generation i into Generation $i + 1$.

Step (3) Create the remaining $K - S$ solutions for Generation $i + 1$ by executing a crossover operation on randomly selected parents from Generation i. For each new solution $\mathbf{x}_k$ created, calculate its objective function value $f(\mathbf{x}_k)$. If $f(\mathbf{x}_k) < f(\mathbf{x}^*)$, then set $\mathbf{x}^* = \mathbf{x}_k$ and $f(\mathbf{x}^*) = f(\mathbf{x}_k)$.

Step (4) For $k = 1, \ldots, K$, mutate solution $\mathbf{x}_k$ in Generation i with probability p. Calculate the new objective function value $f(\mathbf{x}_k)$, and if $f(\mathbf{x}_k) < f(\mathbf{x}^*)$, then set $\mathbf{x}^* = \mathbf{x}_k$ and $f(\mathbf{x}^*) = f(\mathbf{x}_k)$.

Step (5) Set $i = i + 1$. If $i = Q$, then terminate with solution $\mathbf{x}^*$. Otherwise, go to Step 2.

Three specific process concerns are addressed below.

1. *The initial population* — Careful consideration should be given to the creation of the initial set of solutions for Generation 0. For instance, a rudimentary constructive heuristic may be used to create a set of good initial solutions rather than using some blindly random approach. However, it is important that the set of heuristic solutions is sufficiently diverse. That is, if all initial solutions are nearly identical, then the solutions created in the next generation may closely resemble those of the previous generation – limiting the scope of the genetic algorithm search space. Hence, one may penalize solutions having too close a resemblance to previously generated solutions in the initial step.
2. *The crossover process* — The crossover operator is the most important consideration in a genetic algorithm. While the selection of the parents for the crossover operation is done randomly, preference should be given to parent solutions having better quality objective function values (imitating mating of the most fit individuals, as in evolution theory). However, feasibility restrictions on the structure of a solution must be recognized and addressed.

   EXAMPLE For example, in the traveling salesman problem, a feasibility restriction is that each solution be a permutation of integers. Consider the following two parent solutions, where a return to the first city is implied:

   $$(1, 3, 5, 2, 4, 6) \quad \text{and} \quad (1, 2, 3, 6, 5, 4).$$

   A crossover operator that takes the first (last) three chromosomes from the first (second) parent would result in the solution (1, 3, 5, 6, 5, 4), which is not a permutation and cannot be used. A better operator would start with the preceding operator and post-process the result by replacing repeated chromosomes with omitted chromosomes. In the above example, city 5 is repeated while city 2 is omitted, and thus one of the following two solutions would be generated:

   $$(1, 3, 2, 6, 5, 4) \quad \text{or} \quad (1, 3, 5, 6, 2, 4).$$

3. *The mutation operator* — A genetic algorithm without a mutation operator may have the solutions within the same generation converge to a small set of distinct solutions, from which radically different (and perhaps optimal) solutions cannot be created via the crossover operator. The purpose of the mutation operator is to inject diverse elements into future iterations. The mutation operator must generate significant enough change in the solution to ensure that there exists a chance of having this modification propagate into solutions in future generations.

10.19 PROOF METHODS

1. Combinatorial Proof — *Prove that P and Q represent the same thing*
   - Show that P and Q are counting some item, but in different ways.
   - Example: Show that $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.
2. Direct Proof — *Prove $P \implies Q$*
   - Assume that P is true.
   - Use the fact that P is true to show that Q must be true.
   - Example: The sum of 2 adjacent integers is odd.
3. Geometric Proof — *Prove a geometric relationship*
   - Assemble needed definitions, axioms, postulates
   - Combine the assembly, using logical deductions, to reach the goal
   - Example: An angle and the lengths of the two adjacent sides characterize a triangle.
4. Nonconstructive Proof — *Prove P is true.*
   - Create an example, for which all the specifics are not known, which forces P to be true.
   - Example: Show there are irrational numbers a and b with a^b is rational.
5. Proof by Cases/Proof by Exhaustion — *Prove $P(s)$ is true for all $s \in S$*
   - Methodically show, for each $s \in S$, that $P(s)$ is true.
   - Example: Every positive integer less than 10 has a square less than 100.
6. Proof by Construction — *There exists an entity having a property P*
   - Explicitly construct such an entity.
   - Example: There is a prime large than 100.
7. Proof by Contradiction — *Prove P is false*
   - Assume that P is true.
   - Use the fact that P is true to show via logic that something is true, when we know that thing to be false.
   - Example: $\sqrt{2}$ is a rational number.
8. Proof by Contrapositive — *Prove $P \implies Q$*
   - Logically, the proposition $P \implies Q$ is equivalent $\neg Q \implies \neg P$.
   - Instead of proving $P \implies Q$, prove the equivalent statement.
   - Example: If x^2 is even, then x is even.
9. Proof by Induction

   Prove an indexed proposition, $P(n)$, is true for $n = 1, 2, \ldots$
   - Show that $P(1)$ is true.
   - Show that $P(m) \implies P(m+1)$ for $m \geq 1$
   - Example: If a and b are consecutive integers, then $a + b$ is odd.
10. Visual Proof – see page 3

10.20 RECREATIONAL MATHEMATICS

10.20.1 MAGIC SQUARES

A *magic square* is a square array of integers with the property that the sum of the integers in each row or column is the same. If $(c, n) = (d, n) = (e, n) = (f, n) = (cf - en, n) = 1$, then the array $A = (a_{ij})$ will be magic (and use the n^2 numbers $k = 0, 1, \ldots, n^2 - 1$) if $a_{ij} = k$ with

$$i \equiv ck + e \left\lfloor \frac{k}{n} \right\rfloor \pmod{n} \qquad \text{and} \qquad j \equiv dk + f \left\lfloor \frac{k}{n} \right\rfloor \pmod{n}$$

For example, with $c = 1$, $d = e = f = 2$, and $n = 3$, a magic square is

6	1	5
2	3	7
4	8	0

.

An antimagic square is one for which the sums of each row, column, and main diagonal are different and are also sequential integers. Examples include:

15	2	12	4
1	14	10	5
8	9	3	16
11	13	6	7

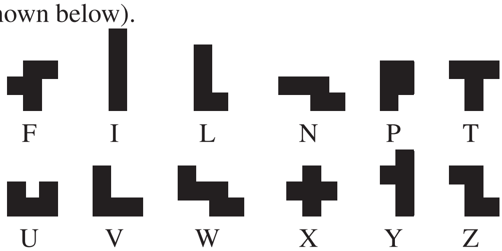
21	18	6	17	4
7	3	13	16	24
5	20	23	11	1
15	8	19	2	25
14	12	9	22	10

10.20.2 POLYOMINOES

Consider the two-dimensional connected shapes formed from N unit squares. Allow them to move in space (e.g., they can be flipped over).

1. For $N = 1$ there is only the single monomino.
2. For $N = 2$ there is only the single domino.
3. For $N = 3$ there are two distinct trominoes.
4. For $N = 4$ there are five distinct tetrominoes. They are used in the game tetris.

5. For $N = 5$ there are 12 distinct pentominoes (where the customary name for each shape is shown below).



6. For $N = 6, 7, 8, 9, \ldots$ there are $35, 108, 369, 1285, \ldots$ unique polyominoes.

The number of three-dimensional polyominoes with N cubes (for $N = 1, 2, \ldots$) is $1, 1, 2, 8, 29, 166, 1023, 6922, 48311, 346543, 2522522, 18598427, 138462649, \ldots$.

10.21 RISK ANALYSIS AND DECISION RULES

Decision rules are actions that are taken based on the state $\{\theta_1, \theta_2, \dots\}$ of a system. For example, in making a decision about a trip, the states may be *rain* and *no rain* and the decision rules are *stay home*, *go with an umbrella*, and *go without an umbrella*.

A *loss function* is a function that depends on a specific state and a decision rule. For example, consider the following loss function $\ell(\theta, a)$:

Loss function data

Possible actions		System state	
		θ_1 (rain)	θ_2 (no rain)
Stay home	a_1	4	4
Go without an umbrella	a_2	5	0
Go with an umbrella	a_3	2	5

It is possible to determine the "best" decision even without obtaining any data.

1. **Minimax principle**
   With this principle one prepares for the worst. For each action it is possible to determine the minimum possible loss that may be incurred. This loss is assigned to each action; the action with the smallest (or minimum) maximum loss is the action chosen.
   For the given loss function data the maximum loss is 4 for action a_1 and 5 for action a_2 or a_3. Under a minimax principle, the chosen action would be a_1 and the minimax loss would be 4.
2. **Minimax principle for mixed actions**
   Assume that action a_i is taken with probability p_i (with $p_1 + p_2 + p_3 = 1$). Then the expected loss $L(\theta_i)$ is given by $L(\theta_i) = p_1\ell(\theta_i, a_1) + p_2\ell(\theta_i, a_2) + p_3\ell(\theta_i, a_3)$. The above data results in the following expected losses:

$$\begin{bmatrix} L(\theta_1) \\ L(\theta_2) \end{bmatrix} = p_1 \begin{bmatrix} 4 \\ 4 \end{bmatrix} + p_2 \begin{bmatrix} 5 \\ 0 \end{bmatrix} + p_3 \begin{bmatrix} 2 \\ 5 \end{bmatrix} \tag{10.21.1}$$

   The minimax point of this mixed action case satisfies $L(\theta_1) = L(\theta_2)$. Using this and $p_1 + p_2 + p_3 = 1$ results in $L(\theta_1) = L(\theta_2) = 4 - 7p_3/5$. Hence, p_3 should be as large as possible; or $\mathbf{p} = (\frac{0}{8}, \frac{3}{8}, \frac{5}{8})$.
   Hence, if action a_2 is chosen $3/8$'s of the time, and action a_3 is chosen $5/8$'s of the time, then the minimax loss is equal to $L = 25/8$. This is a smaller loss than using a pure strategy of only choosing a single action.
3. **Bayes actions**
   If the probability distribution of the states $\{\theta_1, \theta_2, \dots\}$ has the density function $g(\theta_i)$ then the loss has an expectation of $B(a) = \sum_i g(\theta_i)\ell(\theta_i, a)$; this is the *Bayes loss* for action a. A *Bayes action* is an action with minimal Bayes loss.
   For example, assume that $g(\theta_1) = 0.4$ and $g(\theta_2) = 0.6$. Then $B(a_1) = 4$, $B(a_2) = 2$, and $B(a_3) = 3.8$ which leads to the choice of action a_2.

A course of action can also be based on data about the states of interest. For example, a weather report Z will give data for the predictions of *rain* and *no rain*. Continuing the example, assume that the correctness of these predictions is given as follows:

		θ_1 (rain)	θ_2 (no rain)
Predict rain	z_1	0.8	0.1
Predict no rain	z_2	0.2	0.9

That is, when it will rain, then the prediction is correct 80% of the time.

A *decision function* is an assignment of data to actions. For this example there are $3^2 = 9$ possible decision functions, $\{d_1, d_2, \ldots, d_9\}$; they are:

	Decision functions								
	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9
Predict z_1, take action	a_1	a_2	a_3	a_1	a_2	a_1	a_3	a_2	a_3
Predict z_2, take action	a_1	a_2	a_3	a_2	a_1	a_3	a_1	a_3	a_2

The *risk function* $R(\theta, d_i)$ is the expected value of the loss when a specific decision function is being used. This results in the following values:

Risk function evaluation		
Decision Function	θ_1 (rain)	θ_2 (no rain)
d_1	4	4
d_2	5	0
d_3	2	5
d_4	4.2	0.4
d_5	4.8	3.6
d_6	3.6	4.9
d_7	2.4	4.1
d_8	4.4	4.5
d_9	2.6	0.5

This array can now be treated as though it gave the loss function in a no–data problem. The minimax principle for mixed action results in the "best" solution being rule d_3 for $\frac{7}{17}$'s of the time and rule d_9 for $\frac{10}{17}$'s of the time. This leads to a minimax loss of $\frac{40}{17}$. Before the data Z is received, the minimax loss was $\frac{25}{8}$. Hence, the data Z is "worth" $\frac{25}{8} - \frac{40}{17} = \frac{105}{136}$ in using the minimax approach.

The *regret function* $r(\theta, a)$ is the loss, $\ell(\theta, a)$, minus the minimum loss for that θ: $r(\theta, a) = \ell(\theta, a) - \min_b \ell(\theta, b)$. This is the contribution to loss that even a good decision cannot avoid. Hence $r(\theta, a)$ is a loss that could have been avoided had the state been known—hence the term "regret." For the given loss function data, the minimum loss for $\theta = \theta_1$ is 2, for $\theta = \theta_2$ it is 0. In this case the regret function is

	θ_1 (rain)	θ_2 (no rain)
a_1	2	4
a_2	3	0
a_3	0	5

If the minimax principle is used to determine the "best" action, then, in this example, the "best" action is a_2.

10.22 SIGNAL PROCESSING

10.22.1 ESTIMATION

Let $\{e_t\}$ be a white noise process (white noise satisfies $\mathrm{E}[e_t] = \mu$, $\mathrm{Var}[e_t] = \sigma^2$, and $\mathrm{Cov}[e_t, e_s] = 0$ for $s \neq t$). Suppose that $\{X_t\}$ is a time series (sequence of values in time). A non-anticipating linear model presumes that $\sum_{u=0}^{\infty} h_u X_{t-u} = e_t$, where the $\{h_u\}$ are constants. This can be written $H(z)X_t = e_t$ where $H(z) = \sum_{u=0}^{\infty} h_u z^u$ and $z^n X_t = X_{t-n}$. Alternately, $X_t = H^{-1}(z)e_t$. In practice, several types of models are used:

1. AR(k), autoregressive model of order k: This assumes that $H(z) = 1 + a_1 z + \cdots + a_k z_k$ and so
$$X_t + a_1 X_{t-1} + \ldots a_k X_{t-k} = e_t. \tag{10.22.1}$$
2. MA(l), moving average of order l: This assumes that $H^{-1}(z) = 1 + b_1 z + \cdots + b_k z_k$ and so
$$X_t = e_t + b_1 e_{t-1} + \ldots b_l e_{t-l}. \tag{10.22.2}$$
3. ARMA(k, l), mixed autoregressive/moving average of order (k, l): This assumes that $H^{-1}(z) = \frac{1+b_1 z+\cdots+b_l z_l}{1+a_1 z+\cdots+a_k z_k}$ and so
$$X_t + a_1 X_{t-1} + \ldots a_k X_{t-k} = e_t + b_1 e_{t-1} + \cdots + b_l e_{t-l}. \tag{10.22.3}$$

10.22.2 WINDOWING

A signal $\{x(n)\}$ can be truncated by multiplying by a windowing sequence: $x_w(n) = x(n)\, w(n)$ where

$$w(n) = \begin{cases} f(n) & n = 0, 1, \ldots, N-1 \\ 0 & \text{otherwise} \end{cases}$$

and $f(n) \leq 1$. Windowing functions are often symmetric to avoid changing the linear phase of the signal.

Window	Expression	First sidelobe	Main lobe width	peak reduction
Rectangular	$w(n) = 1$	-13.46 dB	1	1
Hamming	$w(n) = 0.54 - 0.46\cos\left(\frac{2\pi n}{N-1}\right)$	-41 dB	2	0.73
Hanning	$w(n) = \frac{1}{2}\left[1 - \cos\left(\frac{2\pi n}{N-1}\right)\right]$	-32 dB	2	0.66

10.22.3 MATCHED FILTERING

Let $X(t)$ represent a signal to be recovered, let $N(t)$ represent noise, and let $Y(t) = X(t) + N(t)$ represent the observable signal. A prediction of the signal is

$$X_p(t) = \int_0^\infty K(z)Y(t-z)\,dz, \tag{10.22.4}$$

where $K(z)$ is a filter. The mean square error is $\mathrm{E}\left[(X(t) - X_p(t))^2\right]$; this is minimized by the optimal (*Wiener*) filter $K_{\text{opt}}(z)$.

When X and Y are stationary, define their autocorrelation functions as $R_{XX}(t-s) = \mathrm{E}\left[X(t)X(s)\right]$ and $R_{YY}(t-s) = \mathrm{E}\left[Y(t)Y(s)\right]$. If $\mathcal{F}$ represents the Fourier transform, then the optimal filter is given by

$$\mathcal{F}\left[K_{\text{opt}}(t)\right] = \frac{1}{2\pi}\frac{\mathcal{F}\left[R_{XX}(t)\right]}{\mathcal{F}\left[R_{YY}(t)\right]}. \tag{10.22.5}$$

For example, if X and N are uncorrelated, then

$$\mathcal{F}\left[K_{\text{opt}}(t)\right] = \frac{1}{2\pi}\frac{\mathcal{F}\left[R_{XX}(t)\right]}{\mathcal{F}\left[R_{XX}(t)\right] + \mathcal{F}\left[R_{NN}(t)\right]}. \tag{10.22.6}$$

In the case of no noise, $\mathcal{F}\left[K_{\text{opt}}(t)\right] = \frac{1}{2\pi}$, $K_{\text{opt}}(t) = \delta(t)$, and $S_p(t) = Y(t)$.

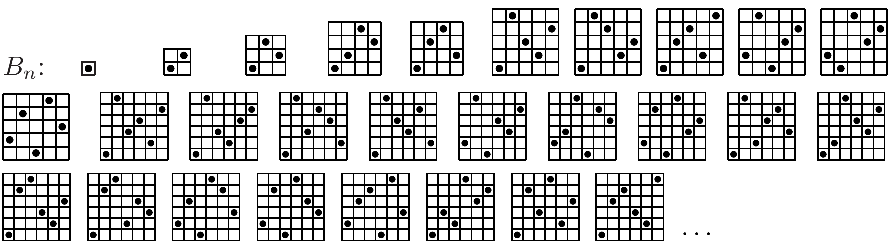
10.22.4 COSTAS ARRAYS

An $n \times n$ *Costas array* is an array of zeros and ones whose two-dimensional autocorrelation function is n at the origin and no more than 1 anywhere else. There are B_n basic Costas arrays; there are C_n arrays when rotations and flips are allowed. Each array can be interpreted as a permutation.

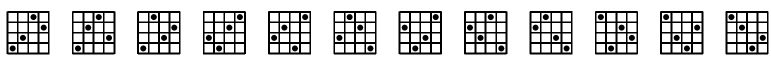
n	1	2	3	4	5	6	7	8	9	10	11	12
B_n	1	1	1	2	6	17	13	17	30	60	555	990
C_n	1	2	4	12	40	116	200	444	760	2160	4368	7852

n	13	14	15	16	17	18	19	20
B_n	1616	2168	2467	2648	2294	1892	1283	810
C_n	12828	17252	19612	21104	18276	15096	10240	6464

B_n:



C_4:



10.22.5 KALMAN FILTERS

Kalman filtering is a linear least squares recursive estimator. It is used when the state space has a higher dimension than the observation space. For example, in some airport radars the distance to aircraft is measured and the velocity of each aircraft is inferred.

1. $\mathbf{x}$ is the unknown state to be estimated
2. $\widehat{\mathbf{x}}$ is the estimate of $\mathbf{x}$
3. $\mathbf{z}$ is an observation
4. $\{\mathbf{w}, \mathbf{v}\}$ are noise terms
5. $\{Q, R\}$ are spectral density matrices
6. "$\mathbf{a} \sim N(\mathbf{b}, C)$" means that the random variable $\mathbf{a}$ has a normal distribution with a mean of $\mathbf{b}$ and a covariance matrix of C
7. "Extended Kalman filter": State propagation is achieved through sequential linearizations of the system model and the measurement model
8. "$(-)$" is the value before a new discrete observation and "$(+)$" is the value after a new discrete observation

10.22.5.1 Discrete Kalman filter

1. System model $\quad \mathbf{x}_k = \Phi_{k-1}\mathbf{x}_{k-1} + \mathbf{w}_{k-1}; \qquad \mathbf{w}_k \sim N(\mathbf{0}, Q_k)$
2. Measurement model $\quad \mathbf{z}_k = H_k\mathbf{x}_k + \mathbf{v}_k; \qquad \mathbf{v}_k \sim N(\mathbf{0}, R_k)$
3. Initial conditions $\quad \mathrm{E}\left[\mathbf{x}(0)\right] = \widehat{\mathbf{x}}_0,$
   $\mathrm{E}\left[(\mathbf{x}(0) - \widehat{\mathbf{x}}_0)(\mathbf{x}(0) - \widehat{\mathbf{x}}_0)^{\mathrm{T}}\right] = P_0$
4. Other assumptions $\quad \mathrm{E}\left[\mathbf{w}_k\mathbf{v}_j^{\mathrm{T}}\right] = 0$ for all j and k
5. State estimate extrapolation $\quad \widehat{\mathbf{x}}_k(-) = \Phi_{k-1}\widehat{\mathbf{x}}_{k-1}(+)$
6. Error covariance extrapolation $\quad P_k(-) = \Phi_{k-1}P_{k-1}(+)\Phi_{k-1}^{\mathrm{T}} + Q_{k-1}$
7. State estimate update $\quad \widehat{\mathbf{x}}_k(+) = \widehat{\mathbf{x}}_k(-) + K_k[\mathbf{z}_k - H_k\widehat{\mathbf{x}}_k(-)]$
8. Error covariance update $\quad P_k(+) = [I - K_kH_k]\,P_k(-)$
9. Kalman gain matrix $\quad K_k = P_k(-)H_k^{\mathrm{T}}\left[H_kP_k(-)H_k^{\mathrm{T}} + R_k\right]^{-1}$

10.22.5.2 Continuous Kalman filter

1. System model $\quad \dot{\mathbf{x}}(t) = F(t)\mathbf{x}(t) + G(t)\mathbf{w}(t); \qquad \mathbf{w}(t) \sim N(\mathbf{0}, Q(t))$
2. Measurement model $\quad \mathbf{z}(t) = H(t)\mathbf{x}(t) + \mathbf{v}(t); \qquad \mathbf{v}(t) \sim N(\mathbf{0}, R(t))$
3. Initial conditions $\quad \widehat{\mathbf{x}}(0) = \widehat{\mathbf{x}}_0, \quad P(0) = P_0,$
   $\mathrm{E}\left[\mathbf{x}(0)\right] = \widehat{\mathbf{x}}_0, \quad \mathrm{E}\left[(\mathbf{x}(0) - \widehat{\mathbf{x}}_0)(\mathbf{x}(0) - \widehat{\mathbf{x}}_0)^{\mathrm{T}}\right] = P_0$
4. Other assumptions $\quad R^{-1}(t)$ exists, $\quad \mathrm{E}\left[\mathbf{w}(t)\mathbf{v}^{\mathrm{T}}(\tau)\right] = C(t)\delta(t-\tau)$
5. State estimate propagation $\quad \dot{\widehat{\mathbf{x}}}(t) = F(t)\widehat{\mathbf{x}}(t) + K(t)\left[\mathbf{z}(t) - H(t)\widehat{\mathbf{x}}(t)\right]$
6. Error covariance propagation
   $$\dot{P}(t) = F(t)P(t) + P(t)F^{\mathrm{T}}(t) + G(t)Q(t)G^{\mathrm{T}}(t) - K(t)R(t)K^{\mathrm{T}}(t)$$
7. Kalman gain matrix $\quad K(t) = \left[P(t)H^{\mathrm{T}}(t) + G(t)C(t)\right]R^{-1}(t)$

10.22.5.3 Continuous extended Kalman filter

1. System model $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{w}(t);\quad \mathbf{w}(t) \sim N(\mathbf{0}, Q(t))$
2. Measurement model $\mathbf{z}(t) = \mathbf{h}(\mathbf{x}(t)) + \mathbf{v}(t);\quad \mathbf{v}(t) \sim N(\mathbf{0}, R(t))$
3. Initial conditions $\mathbf{x}(0) \sim N(\widehat{\mathbf{x}}_0, P_0)$
4. Other assumptions $\mathrm{E}\left[\mathbf{w}(t)\mathbf{v}^{\mathrm{T}}(\tau)\right] = 0$ for all t and all τ
5. State estimate propagation $\dot{\widehat{\mathbf{x}}}(t) = \mathbf{f}(\widehat{\mathbf{x}}(t), t) + K(t)\left[\mathbf{z}(t) - \mathbf{h}(\widehat{\mathbf{x}}(t), t)\right]$
6. Error covariance propagation
$$\begin{aligned}\dot{P}(t) = F(\widehat{\mathbf{x}}(t), t)P(t) + P(t)F^{\mathrm{T}}(\widehat{\mathbf{x}}(t), t) + Q(t)\\ -P(t)H^{\mathrm{T}}(\widehat{\mathbf{x}}(t), t)R^{-1}(t)H(\widehat{\mathbf{x}}(t), t)P(t)\end{aligned}$$
7. Gain equation $K(t) = P(t)H^{\mathrm{T}}(\widehat{\mathbf{x}}(t), t)R^{-1}(t)$
8. Definitions
$$F(\widehat{\mathbf{x}}(t), t) = \left.\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right|_{\mathbf{x}(t)=\widehat{\mathbf{x}}(t)}$$
$$H(\widehat{\mathbf{x}}(t), t) = \left.\frac{\partial \mathbf{h}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right|_{\mathbf{x}(t)=\widehat{\mathbf{x}}(t)}$$

10.22.5.4 Continuous-discrete extended Kalman filter

1. System model $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), t) + \mathbf{w}(t);\quad \mathbf{w}(t) \sim N(\mathbf{0}, Q(t))$
2. Measurement model $\mathbf{z}_k = \mathbf{h}_k(\mathbf{x}(t_k)) + \mathbf{v}_k;\quad \mathbf{v}_k \sim N(\mathbf{0}, R_k)$
$k = 1, 2, \dots$
3. Initial conditions $\mathbf{x}(0) \sim N(\widehat{\mathbf{x}}_0, P_0)$
4. Other assumptions $\mathrm{E}\left[\mathbf{w}(t)\mathbf{v}_k^{\mathrm{T}}\right] = 0$ for all k and all t
5. State estimate propagation $\dot{\widehat{\mathbf{x}}}(t) = \mathbf{f}(\widehat{\mathbf{x}}(t), t)$
6. Error covariance propagation
$$\dot{P}(t) = F(\widehat{\mathbf{x}}(t), t)P(t) + P(t)F^{\mathrm{T}}(\widehat{\mathbf{x}}(t), t) + Q(t)$$
7. State estimate update $\widehat{\mathbf{x}}_k(+) = \widehat{\mathbf{x}}_k(-) + K_k\left[\mathbf{z}_k - \mathbf{h}_k(\widehat{\mathbf{x}}_k(-))\right]$
8. Error covariance update $P_k(+) = \left[I - K_k H_k(\widehat{\mathbf{x}}_k(-))\right]P_k(-)$
9. Gain matrix
$$K_k = P_k(-)H_k^{\mathrm{T}}(\widehat{\mathbf{x}}_k(-))\left[H_k(\widehat{\mathbf{x}}_k(-))P_k(-)H_k^{\mathrm{T}}(\widehat{\mathbf{x}}_k(-)) + R_k\right]^{-1}$$
10. Definitions
$$F(\widehat{\mathbf{x}}(t), t) = \left.\frac{\partial \mathbf{f}(\mathbf{x}(t), t)}{\partial \mathbf{x}(t)}\right|_{\mathbf{x}(t)=\widehat{\mathbf{x}}(t)}$$
$$H_k(\widehat{\mathbf{x}}_k(-)) = \left.\frac{\partial \mathbf{h}_k(\mathbf{x}(t_k))}{\partial \mathbf{x}(t_k)}\right|_{\mathbf{x}(t_k)=\widehat{\mathbf{x}}_k(-)}$$

10.22.6 WAVELETS

Constructing wavelet orthonormal bases $\{\psi_{j,k}\}_{j,k=-\infty}^{+\infty}$ begins by choosing real coefficients $h_0, \dots, h_n$ satisfying the following (set $h_k = 0$ if $k < 0$ or $k > n$):

1. *Normalization*: $\sum_k h_k = \sqrt{2}$.
2. *Orthogonality*: $\sum_k h_k\, h_{k-2j} = 1$ if $j = 0$ and 0 if $j \neq 0$.
3. *Accuracy* p: $\sum_k (-1)^k k^j h_k = 0$ for $j = 0, \dots, p-1$ with $p > 0$.
4. *Cohen–Lawton criterion*: A technical condition rarely violated by coefficients which satisfy the normalization, orthogonality, and accuracy p conditions.

These four conditions imply the existence of a solution $\varphi \in L^2(\mathbb{R})$, called the *scaling function*, to the following *refinement equation*:

$$\varphi(x) = \sqrt{2} \sum_{k=0}^{n} h_k \, \varphi(2x - k). \tag{10.22.7}$$

The scaling function is normalized so that $\int \varphi(x)\, dx = 1$. Then $\varphi(x)$ is unique, and it vanishes outside of the interval $[0, n]$. The maximum possible accuracy is $p = (n+1)/2$.

For each fixed integer j, let V_j be the closed subspace of $L^2(\mathbb{R})$ spanned by the functions $\{\varphi_{j,k}\}_{k=-\infty}^{+\infty}$ where $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$. The projection of $f(x)$ onto the subspace V_j is an approximation at resolution level 2^{-j}. It is given by $f_j(x) = \sum_k c_{j,k}\, \varphi_{j,k}(x)$ with $c_{j,k} = \langle f, \varphi_{j,k} \rangle = \int f(x)\, \varphi_{j,k}(x)\, dx$.

The *wavelet* ψ is derived from the scaling function φ by the formula,

$$\psi(x) = \sqrt{2} \sum_{k=0}^{n} g_k \, \varphi(2x - k), \qquad \text{where } g_k = (-1)^k \, h_{n-k}. \tag{10.22.8}$$

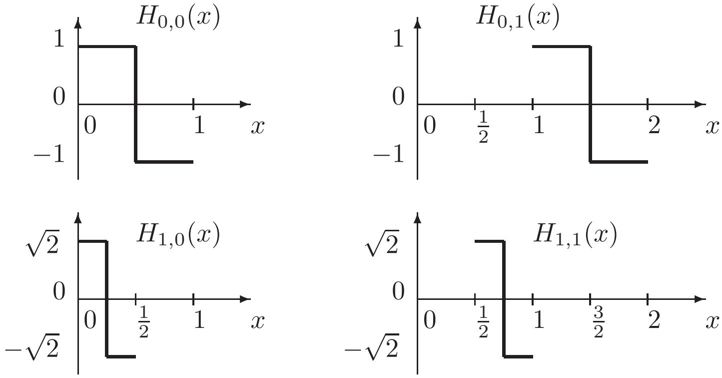
The wavelet ψ has the same smoothness as φ, and the accuracy p condition implies vanishing moments for ψ: $\int x^j \psi(x)\, dx = 0$ for $j = 0, \dots, p-1$. The functions $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ are orthonormal and the entire collection $\{\psi_{j,k}\}$ (where $j \geq 0$ and $0 \leq k < 2^j$) forms an orthonormal basis for the Hilbert space $L^2(\mathbb{R})$. That is, functions f with finite energy, i.e., $\int_{-\infty}^{+\infty} |f(x)|^2\, dx < \infty$.

10.22.6.1 Haar wavelets

The *Haar wavelet* $H(x)$ and its corresponding scaling function are

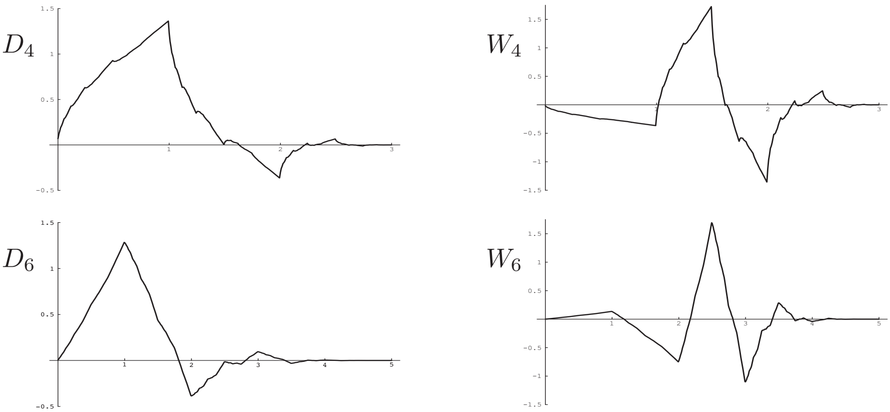
$$H(x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2} \\ -1 & \text{if } \frac{1}{2} \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \qquad \phi(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{10.22.9}$$

Define the *Haar system* $\{H_{j,k}\}$ by $H_{j,k}(x) = 2^{j/2} H(2^j x - k)$. While the Haar wavelet is the simplest possible, it is not continuous and therefore not differentiable.



10.22.6.2 Daubechies wavelets

For each *even* integer $N > 0$, there is a unique set of coefficients $h_0, \dots, h_{N-1}$ which satisfy the normalization and orthogonality conditions with maximal accuracy $p = N/2$. The corresponding φ and ψ are the *Daubechies scaling function* D_N and *Daubechies wavelet* W_N. The Haar wavelet H is the Daubechies wavelet W_2. For the Haar wavelet, the coefficients are $h_0 = h_1 = 1/\sqrt{2}$ and the subspace V_j consists of all functions which are piecewise constant on each interval $[k2^{-j}, (k+1)2^{-j})$.



10.22.7 THE CONTINUOUS WAVELET TRANSFORM

The *continuous wavelet transform* of a function $f(t)$ with respect to a *mother wavelet* ψ is defined as

$$W_\psi[f](a,b) = \int_{-\infty}^{\infty} f(x)\overline{\psi}_{a,b}(x)\, dx, \tag{10.22.10}$$

where $a \neq 0$ and b are real numbers and

$$\psi_{a,b}(x) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{x-b}{a}\right). \tag{10.22.11}$$

The parameters a and b are called the scale (frequency) and translation parameters, respectively.

The mother wavelet is assumed to satisfy the condition

$$C_\psi = 2\pi \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(w)|^2}{|w|} dw < \infty, \tag{10.22.12}$$

where $\widehat{\psi} = \mathcal{F}(\psi)$ is the Fourier transform of ψ. Condition (10.22.12) is the *admissibility condition*.

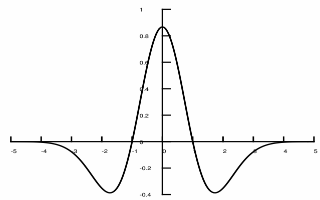
The constant $1/\sqrt{|a|}$ in (10.22.11) is a normalizing constant introduced so that

$$\|\psi_{a,b}\|^2 = \int_{-\infty}^{\infty} |\psi_{a,b}(x)|^2\, dx = \int_{-\infty}^{\infty} |\psi(x)|^2\, dx = \|\psi\|^2.$$

If ψ is *integrable*, i.e., $\int_{-\infty}^{\infty} |\psi(x)|\, dx < \infty$, then the admissibility condition is satisfied only if

$$\widehat{\psi}(0) = 0 \quad \text{or} \quad \int_{-\infty}^{\infty} \psi(x) dx = 0. \tag{10.22.13}$$

Example: A commonly used mother wavelet is the *Mexican Hat function*; it is the second derivative of a Gaussian. When normalized so that $||\psi||^2 = 1$, it is $\psi(x) = \dfrac{2}{\pi^{1/4}\sqrt{3}}(1-x^2)e^{-x^2/2}$.



Properties of The Continuous Wavelet Transform

1. **Another Version of the Wavelet Transform:** Since

$$\mathcal{F}(\psi_{a,b}(x))(w) = \sqrt{|a|}\, e^{ibw}\widehat{\psi}(aw), \tag{10.22.14}$$

another version of the continuous wavelet transform in terms of the Fourier transform of the wavelet is

$$W_\psi[f](a,b) = \sqrt{|a|}\int_{-\infty}^{\infty} \widehat{f}(w)e^{-ibw}\overline{\widehat{\psi}}(aw)\, dw. \tag{10.22.15}$$

2. **Linearity:** The continuous wavelet transform is a continuous linear transformation, i.e.,

$$W_\psi[\alpha f(x) + \beta g(x)] = \alpha W_\psi[f(x)] + \beta W_\psi[g(x)],$$

where α and β are complex numbers. This maps $L^2(\mathbb{R})$ into the Hilbert space

$$L^2\left(\mathbb{R}^2, a^{-2}\, da\, db\right) = \left(F(a,b) : \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F(a,b)|^2 \frac{da\, db}{a^2} < \infty\right),$$

with inner product

$$\langle F, G\rangle_{L^2(\mathbb{R}^2,\mu)} = \langle F, G\rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(a,b)\overline{G}(a,b)\, d\mu(a,b)$$

where $d\mu(a,b) = da\, db/a^2$.

3. **Translation:**

$$W_\psi[f(x+c)](a,b) = W_\psi[f(x)](a, b+c) \tag{10.22.16}$$

4. **Dilation:**

$$W_\psi\left[f\left(\frac{x}{c}\right)\right](a,b) = \sqrt{c}\, W_\psi[f(x)]\left(\frac{a}{c}, \frac{b}{c}\right), \quad c > 0. \tag{10.22.17}$$

5. **Interchanging of Mother Wavelets:** If ϕ and ψ are admissible mother wavelets, then

$$W_\psi[\phi](a,b) = \overline{W}_\phi[\psi]\left(\frac{1}{a}, -\frac{b}{a}\right). \tag{10.22.18}$$

6. **Inversion formula:** The inversion formula of the wavelet transform given by (10.22.10) is

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_\psi[f](a,b)\psi_{a,b}(x) \frac{da\, db}{a^2}, \tag{10.22.19}$$

where C_ψ is given by (10.22.12).
The inversion formula for the wavelet transform given by (10.22.15) is

$$\widehat{f}(w) = \frac{1}{2\pi \sqrt{|a|}\, \overline{\widehat{\psi}}(aw)} \int_{-\infty}^{\infty} W_\psi[f](a,b) e^{iwb} db. \tag{10.22.20}$$

7. **Parseval's relation:** Let $f, g \in L^2(\mathbb{R})$ and denote their continuous wavelet transforms with respect to the same mother wavelet by F and G, respectively. In view of (10.22.15) and (10.22.20), we obtain

$$\langle W_\psi[f], W_\psi[g] \rangle_{L^2(\mathbb{R}^2,\mu)} = \langle F, G \rangle_{L^2(\mathbb{R}^2,\mu)} = C_\psi \langle \widehat{f}, \widehat{g} \rangle_{L^2(\mathbb{R})} = C_\psi \langle f, g \rangle_{L^2(\mathbb{R})}, \tag{10.22.21}$$

where C_ψ is given by (10.22.12). In particular,

$$\|F\|^2_{L^2(\mathbb{R}^2,\mu)} = C_\psi \|f\|^2_{L^2(\mathbb{R})}. \tag{10.22.22}$$

If ψ is normalized so that $C_\psi = 1$, it follows that the continuous wavelet transform is an isometry from $L^2(\mathbb{R})$ into $L^2(\mathbb{R}^2, \mu)$.

8. **A Generalized Parseval's Relation:** For the two mother wavelets ψ and ϕ define

$$C_{\psi,\phi} = 2\pi \int_{-\infty}^{\infty} \frac{\widehat{\phi}(w)\overline{\widehat{\psi}}(w)}{|w|} dw < \infty, \tag{10.22.23}$$

and denote the continuous wavelet transform of $f, g \in L^2(\mathbb{R})$ with respect to ψ and ϕ by $W_\psi[f]$ and $W_\phi[g]$, respectively. Then

$$\langle W_\psi[f], W_\phi[g] \rangle_{L^2(\mathbb{R}^2,\mu)} = C_{\psi,\phi} \langle \widehat{f}, \widehat{g} \rangle = C_{\psi,\phi} \langle f, g \rangle. \tag{10.22.24}$$

If $\psi = \phi$, Formula (10.22.24) reduces to (10.22.21).

9. **A Generalized Inversion Formula:** The following inversion formula is a consequence of (10.22.24). It shows that if the continuous wavelet transform is taken with respect to a mother wavelet ψ, the inversion formula may be taken with respect to a different mother wavelet ϕ. In fact, the following relation holds

$$f(x) = \frac{1}{C_{\psi,\phi}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_\psi[f](a,b)\phi_{a,b}(x) \frac{da\, db}{a^2}. \tag{10.22.25}$$

In particular if $\phi = \psi$, we obtain

$$f(x) = \frac{1}{C_\psi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W_\psi[f](a,b)\psi_{a,b}(x) \frac{da\, db}{a^2}. \tag{10.22.26}$$

10.22.8 THE DISCRETE WAVELET TRANSFORM AND WAVELET SERIES

The discrete wavelet transform of f may be viewed as the values of the continuous wavelet transform of f taken at a lattice of points obtained by discretizing the parameters a and b. The standard choices are: $a = a_0^m$ with $a_0 > 0$ and $a_0 \neq 1$, $b = nb_0a_0^m$ with $b_0 \neq 0$, and m and n are integers $(0, \pm 1, \pm 2, \ldots)$. We choose, without loss of generality, $a_0 > 1$ and $b_0 > 0$.

The discrete wavelet transform of f, also called the *wavelet coefficients* of f, is defined as the doubly indexed sequence

$$\tilde{f}(m,n) = W_\psi[f](m,n) = \langle f, \psi_{m,n}\rangle = \int_{-\infty}^{\infty} f(x)\overline{\psi}_{m,n}(x)\,dx,$$

where

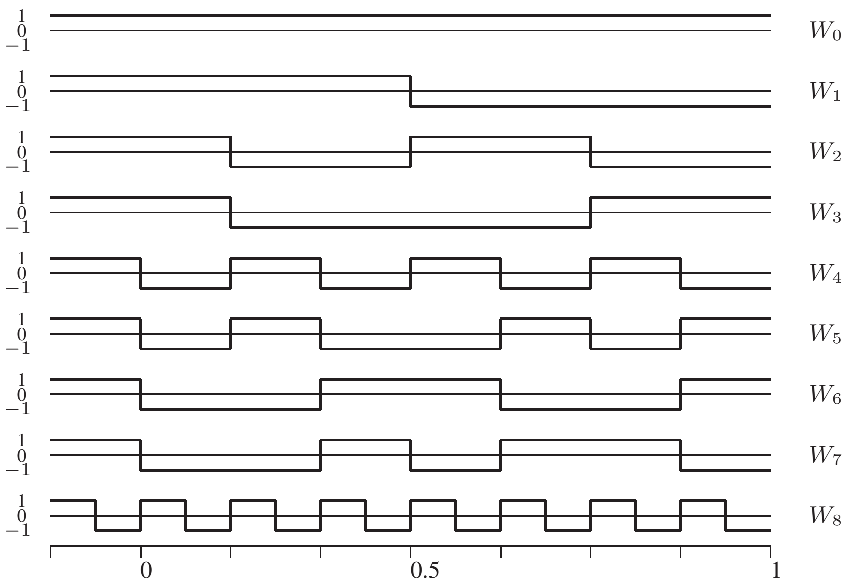
$$\psi_{m,n}(x) = a_0^{m/2}\psi(a_0^m x - nb_0). \tag{10.22.27}$$

We call the series $\sum_{m,n=-\infty}^{\infty} \tilde{f}(m,n)\psi_{m,n}(x)$ *the wavelet series* of f, the function $\psi(x)$ the mother wavelet, and the functions $\{\psi_{m,n}(x)\}$ *the wavelets*.

Thus, wavelets are functions generated from one single function (the mother wavelet) by dilation (scaling) and translation. The standard choice for a and b are $a = 2$ and $b = 1$, and in this case the wavelets are called *dyadic wavelets*.

10.22.9 WALSH FUNCTIONS

The *Rademacher functions* are defined by $r_k(x) = \operatorname{sgn}\sin\left(2^{k+1}\pi x\right)$. If the binary expansion of n has the form $n = 2^{i_1} + 2^{i_2} + \cdots + 2^{i_m}$, then the *Walsh function* of order n is $W_n(x) = r_{i_1}(x)r_{i_2}(x)\ldots r_{i_m}(x)$. The Rademacher functions are an orthogonal system but they are not complete.



10.23 UNITS

10.23.1 SI SYSTEM OF MEASUREMENT

SI, the abbreviation of the French words "Système Internationale d'Unités," is the accepted abbreviation for the International Metric System.

1. There are 7 base units

Quantity measured	**Unit**	**Symbol**
Length	meter	m
Mass	kilogram	kg
Time	second	s
Amount of substance	mole	mol
Electric current	ampere	A
Luminous intensity	candela	cd
Thermodynamic temperature	kelvin	K

2. There are 22 derived units with special names and symbols.
   When a unit is named after a person, the unit's name is in lower case letters and the unit's symbol starts with a capital letter.

Quantity	**SI Name**	**Symbol**	**Combination of other SI units (or base units)**
Absorbed dose	gray	Gy	J/kg
Activity (radiation source)	becquerel	Bq	1/s
Capacitance	farad	F	C/V
Catalytic activity	katal	kat	s^{-1} mol
Celsius temperature	Celsius	°C	K
Conductance	siemen	S	A/V
Dose equivalent	sievert	Sv	J/kg
Electric charge	coulomb	C	A s
Electric potential	volt	V	W/A
Electric resistance	ohm	Ω	V/A
Energy	joule	J	N m
Force	newton	N	kg m/s^2
Frequency	hertz	Hz	1/s
Illuminance	lux	lx	lm/m^2
Inductance	henry	H	Wb/A
Luminous flux	lumen	lm	cd sr
Magnetic flux density	tesla	T	Wb/m^2
Magnetic flux	weber	Wb	V s
Plane angle (unitless)	radian	rad	$m \cdot m^{-1}$
Power	watt	W	J/s
Pressure or stress	pascal	Pa	N/m^2
Solid angle (unitless)	steradian	sr	$m^2 \cdot m^{-2}$

3. The following units are accepted for use with SI units.

Name	Symbol	Value in SI units
(angle) degree	$\circ$	$1^\circ = (\pi/180)$ rad
(angle) minute	$'$	$1' = (1/60)^\circ = (\pi/10800)$ rad
(angle) second	$''$	$1'' = (1/60)' = (\pi/648000)$ rad
(time) day	d	1 d = 24 h = 86400 s
(time) hour	h	1 h = 60 min = 3600 s
(time) minute	min	1 min = 60 s
astronomical unit	au	1 au $\approx 1.49598 \times 10^{11}$ m
bel	B	1 B = (1/2) ln 10 Np
		(Note that 1 dB = 0.1 B)
electronvolt	eV	1 eV $\approx 1.6021764 \times 10^{-19}$ C
liter	L	1 L = 1 dm^3 = 10^{-3} m^3
metric ton	t	1 t = 10^3 kg
neper	Np	1 Np = 1 (unitless)
unified atomic mass unit	u	1 u $\approx 1.66054 \times 10^{-27}$ kg

4. The following units are currently accepted for use with SI units

Name	Symbol	Value in SI units
angstrom	Å	1Å$= 0.1$ nm $= 10^{-10}$ m
are	a	1 a = 1 dam$^2 = 10^2$ m^2
barn	b	1 b = 100 fm$^2 = 10^{-28}$ m^2
bar	bar	1 bar = 0.1 MPa = 100 kPa = 1000 hPa $= 10^5$ Pa
curie	Ci	1 Ci $= 3.7 \times 10^{10}$ Bq
hectare	ha	1 ha = 1 hm$^2 = 10^4$ m^2
knot		1 nautical mile per hour = (1852/3600) m/s
nautical mile		1 nautical mile = 1852 m
rad	rad	1 rad = 1 cGy $= 10^{-2}$ Gy
rem	rem	1 rem = 1 cSv $= 10^{-2}$ Sv
roentgen	R	1 R $= 2.58 \times 10^{-4}$ C/kg

10.23.2 TEMPERATURE CONVERSION

If t_F is the temperature in degrees Fahrenheit and t_C is the temperature in degrees Celsius, then

$$t_C = \frac{5}{9}(t_F - 32) \quad \text{and} \quad t_F = \frac{9}{5}t_C + 32. \tag{10.23.1}$$

−40°C	0°C	10°C	20°C	37°C	100°C
−40°F	32°F	50°F	68°F	98.6°F	212°F

If T_K is the temperature in kelvin and T_R is the temperature in degrees Rankine, then

$$T_R = t_F + 459.69 \qquad \text{and} \qquad T_K = t_C + 273.15 = \frac{5}{9}T_R. \tag{10.23.2}$$

10.23.3 UNITED STATES CUSTOMARY SYSTEM OF WEIGHTS AND MEASURES

1. Linear measure

   1 mile = 5280 feet or 320 rods
   1 rod = 16.5 feet or 5.5 yards
   1 yard = 3 feet
   1 foot = 12 inches

2. Linear measure: Nautical

   1 fathom = 6 feet
   1° of latitude = 69 miles
   1° of longitude at 40° latitude = 46 nautical miles $\approx$ 53 miles
   1 nautical mile = 6076.1 feet $\approx$ 1.1508 statute miles

3. Square measure

   1 square mile = 640 acres
   1 acre = 43,560 square feet

4. Volume measure

   1 cubic yard = 27 cubic feet
   1 cubic foot = 1728 cubic inches

5. Dry measure

   1 bushel = 4 pecks
   1 peck = 8 quarts
   1 quart = 2 pints

6. Liquid measure

   1 cubic foot = 7.4805 gallons
   1 gallon = 4 quarts
   1 quart = 2 pints
   1 pint = 4 gills

7. Liquid measure: Apothecaries'

   1 pint = 16 fluid ounces
   1 fluid ounce = 8 drams
   1 fluid dram = 60 minims

8. Weight: Apothecaries'

   1 pound = 12 ounces
   1 ounce = 8 drams
   1 dram = 3 scruples
   1 scruple = 20 grains

9. Weight: Avoirdupois

   1 ton = 2000 pounds
   1 pound = 16 ounces or 7000 grains
   1 ounce = 16 drams or 437.5 grains

10. Weight: Troy

    1 pound = 12 ounces
    1 ounce = 20 pennyweights
    1 pennyweight = 24 grains

10.23.4 UNITS OF PHYSICAL QUANTITIES

The units of a system's parameters constrain all the derivable quantities, regardless of the equations describing the system. In particular, all derived quantities are functions of dimensionless combinations of parameters. The number of dimensionless parameters and their forms are given by the *Buckingham pi theorem*.

Suppose $u = f(W_1, \dots, W_n)$ is to be determined in terms of n measurable variables and parameters $\{W_i\}$ where f is an unknown function. Let $\{u, W_i\}$ involve m fundamental dimensions labeled by $L_1, \dots, L_m$ (e.g., length or mass). The dimensions of the $\{u, W_i\}$ are given by a product of powers of the fundamental dimensions. For example, the dimensions of W_i are $L_1^{b_{i1}} L_2^{b_{i2}} \cdots L_m^{b_{im}}$ where the $\{b_{ij}\}$ are real and called the *dimensional exponents*. A quantity is called dimensionless if all of its dimensional exponents are zero. Let $\mathbf{b}_i = \begin{bmatrix} b_{i1} & \dots & b_{im} \end{bmatrix}^{\mathrm{T}}$ be the dimension vector of W_i and let $B = \begin{bmatrix} \mathbf{b}_1 & \dots & \mathbf{b}_n \end{bmatrix}$ be the $m \times n$ dimension matrix of the system. Let $\mathbf{a} = \begin{bmatrix} a_1 & \dots & a_m \end{bmatrix}^{\mathrm{T}}$ be the dimension vector of u and let $\mathbf{y} = \begin{bmatrix} y_1 & \dots & y_n \end{bmatrix}^{\mathrm{T}}$ represent a solution of $B\mathbf{y} = -\mathbf{a}$. Then,

1. The number of dimensionless quantities is $k + 1 = n + 1 - \operatorname{rank}(B)$.
2. The quantity u can be expressed in terms of dimensionless parameters as

$$u = W_1^{-y_1} W_2^{-y_2} \cdots W_n^{-y_n} g(\pi_1, \pi_2, \dots, \pi_k) \tag{10.23.3}$$

where $\{\pi_i\}$ are dimensionless quantities. Specifically, let $\mathbf{x}^{(i)} = \begin{bmatrix} x_{1i} & \dots & x_{ni} \end{bmatrix}^{\mathrm{T}}$ be one of $k = n - r(B)$ linearly independent solutions of the system $B\mathbf{x} = \mathbf{0}$ and define $\pi_i = W_1^{x_{1i}} W_2^{x_{2i}} \cdots W_n^{x_{ni}}$.

In the following, it may be easier to think "kilograms" for the mass M, "meters" for the length L, "seconds" for the time T, and "degrees" for the temperature θ. For example, acceleration is measured in units of L/T^2, or meters per second squared.

Quantity	Dimensions
Acceleration	L/T^2
Angular acceleration	$1/T^2$
Angular frequency	$1/T$
Angular momentum	ML^2/T
Angular velocity	$1/T$
Area	L^2
Displacement	L
Energy	ML^2/T^2
Force	ML/T^2
Frequency	$1/T$
Gravitational field strength	ML/T^2
Gravitational potential	ML^2/T^2
Length	L
Mass	M
Mass density	M/L^3

Quantity	Dimensions
Momentum	ML/T
Period	T
Power	ML^2/T^3
Pressure	M/LT^2
Moment of inertia	ML^2
Time	T
Torque	ML^2/T^2
Velocity	L/T
Volume	L^3
Wavelength	L
Work	ML^2/T^2
Entropy	$ML^2/T^2\theta$
Internal energy	ML^2/T^2
Heat	ML^2/T^2

EXAMPLES

1. *Shock wave propagation from a point explosion*

   Suppose a nuclear explosion, represented as a point explosion of energy E, creates a shock wave propagating into air of density ρ. After time t and we want to know the radius r. Using the ordering $\begin{bmatrix} M & L & T \end{bmatrix}$ the physical quantities have dimensions:

   - $u = r$ has dimensions L or $\mathbf{a} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_1 = t$ has dimensions T or $\mathbf{b}_1 = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$
   - $W_2 = \rho$ has dimensions ML^{-3} or $\mathbf{b}_2 = \begin{bmatrix} 1 & -3 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_3 = E$ has dimensions ML^2T^{-2} or $\mathbf{b}_3 = \begin{bmatrix} 1 & 2 & -2 \end{bmatrix}^{\mathrm{T}}$

   so that $n = 3$ and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 1 \\ 0 & -3 & 2 \\ 1 & 0 & -2 \end{bmatrix}$. The rank of B is 3, so there is $n + 1 - \operatorname{rank}(B) = 1$ dimensionless quantity. The unique solution of $B\mathbf{y} = -\mathbf{a}$ is $\mathbf{y} = \begin{bmatrix} -\frac{2}{5} & \frac{1}{5} & -\frac{1}{5} \end{bmatrix}^{\mathrm{T}}$. Putting this together, the solution can be represented as $u = r = W_1^{-y_1} W_2^{-y_2} W_3^{-y_3} g(\pi_1) = t^{2/5} \rho^{-1/5} E^{1/5} g(\pi_1)$ or $r = C \left(\frac{Et^2}{\rho} \right)^{1/5}$ where C is a constant. Note, in particular, the scaling law $r_{\text{later}} = r_{\text{earlier}} \left(\frac{t_{\text{later}}}{t_{\text{earlier}}} \right)^{2/5}$.

2. *A simple pendulum*

   A pendulum swinging under the force of gravity g consists of a (massless) rod of length ℓ and a mass m at the tip. We want to find the oscillation period T_p. Using the ordering $\begin{bmatrix} M & L & T \end{bmatrix}$ the physical quantities have dimensions:

   - $u = T_p$ has dimensions T or $\mathbf{a} = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^{\mathrm{T}}$
   - $W_1 = \ell$ has dimensions L or $\mathbf{b}_1 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_2 = m$ has dimensions M or $\mathbf{b}_2 = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_3 = g$ has dimensions LT^{-2} or $\mathbf{b}_3 = \begin{bmatrix} 0 & 1 & -2 \end{bmatrix}^{\mathrm{T}}$

   so that $n = 3$ and $B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & -2 \end{bmatrix}$. The rank of B is 3, so there is $n + 1 - \operatorname{rank}(B) = 1$ dimensionless quantity. The unique solution of $B\mathbf{y} = -\mathbf{a}$ is $\mathbf{y} = \begin{bmatrix} -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}^{\mathrm{T}}$. Putting this together, the solution can be represented as $u = T_p = W_1^{-y_1} W_2^{-y_2} W_3^{-y_3} h(\pi_1) = \ell^{1/2} m^0 g^{-1/2} h(\pi_1)$ or $T_p = C\sqrt{\frac{\ell}{g}}$ where C is a constant. Note, in particular, the period does not depend on the mass m.

   For small oscillations a physics analysis shows that $C = 2\pi$.

3. *Drag in a fluid*

   A sphere of size ℓ, which is subject to gravity g, is in a fluid with viscosity μ, density ρ, and velocity v. We will determine the drag D. Using the ordering $\begin{bmatrix} M & L & T \end{bmatrix}$:

   - $u = D$ has dimensions MLT^{-2} or $\mathbf{a} = \begin{bmatrix} 1 & 1 & -2 \end{bmatrix}^{\mathrm{T}}$
   - $W_1 = \ell$ has dimensions L or $\mathbf{b}_1 = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_2 = \mu$ has dimensions $ML^{-1}T^{-1}$ or $\mathbf{b}_2 = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix}^{\mathrm{T}}$
   - $W_3 = \rho$ has dimensions ML^{-3} or $\mathbf{b}_3 = \begin{bmatrix} 1 & -3 & 0 \end{bmatrix}^{\mathrm{T}}$
   - $W_4 = v$ has dimensions LT^{-1} or $\mathbf{b}_4 = \begin{bmatrix} 0 & 1 & -1 \end{bmatrix}^{\mathrm{T}}$
   - $W_5 = g$ has dimensions LT^{-2} or $\mathbf{b}_5 = \begin{bmatrix} 0 & 1 & -2 \end{bmatrix}^{\mathrm{T}}$

   A solution of $B\mathbf{y} = -\mathbf{a}$ is $\mathbf{y} = \begin{bmatrix} -2 & 0 & -1 & -2 & 0 \end{bmatrix}^{\mathrm{T}}$. The null solutions are: $B\begin{bmatrix} -1 & 0 & 0 & 2 & -1 \end{bmatrix}^{\mathrm{T}} = \mathbf{0}$ and $B\begin{bmatrix} 1 & -1 & 1 & 1 & 0 \end{bmatrix}^{\mathrm{T}} = \mathbf{0}$.

   Hence, $u = D = \rho \ell^2 v^2 f(\mathrm{Re}, \mathrm{Fr})$ where Re$= \ell\rho v/\mu$ is the *Reynold's number*, Fr$= v/\sqrt{\ell g}$ is the *Froude number*, and $f(\cdot, \cdot)$ is a dimensionless function.

10.23.5 CONVERSION: METRIC TO ENGLISH

Multiply	By	To obtain
centimeters	0.3937008	inches
cubic meters	1.307951	cubic yards
cubic meters	35.31467	cubic feet
grams	0.03527396	ounces
kilograms	2.204623	pounds
kilometers	0.6213712	miles
liters	0.2641721	gallons (US)
meters	1.093613	yards
meters	3.280840	feet
milliliters	0.03381402	fluid ounces
milliliters	0.06102374	cubic inches
square centimeters	0.1550003	square inches
square meters	1.195990	square yards
square meters	10.76391	square feet

10.23.6 CONVERSION: ENGLISH TO METRIC

Multiply	By	To obtain
cubic feet	0.02831685	cubic meters
cubic inches	16.38706	milliliters
cubic yards	0.7645549	cubic meters
feet	0.3048000	meters
fluid ounces	29.57353	milliliters
gallons (US)	3.785412	liters
inches	2.540000	centimeters
miles	1.609344	kilometers
mils	25.4	micrometers
ounces	28.34952	grams
pounds	0.4535924	kilograms
square feet	0.09290304	square meters
square inches	6.451600	square centimeters
square yards	0.8361274	square meters
yards	0.9144000	meters

10.23.7 MISCELLANEOUS CONVERSIONS

Multiply	By	To obtain
feet of water at 4°C	2.950×10^{-2}	atmospheres
inches of mercury at 4°C	3.342×10^{-2}	atmospheres
pounds per square inch	6.804×10^{-2}	atmospheres
foot-pounds	1.285×10^{-3}	BTU
joules	9.480×10^{-4}	BTU
cords	128	cubic feet
radian	57.29578	degree (angle)
foot-pounds	1.356×10^{7}	ergs
atmospheres	33.90	feet of water at 4°C
miles	5280	feet
horsepower	3.3×10^{4}	foot-pounds per minute
horsepower-hours	1.98×10^{6}	foot-pounds
kilowatt-hours	2.655×10^{6}	foot-pounds
foot-pounds per second	1.818×10^{-3}	horsepower
atmospheres	29.92	inches of mercury at 0°C
BTU	1.055060×10^{3}	joules
foot-pounds	1.35582	joules
BTU per minute	1.758×10^{-2}	kilowatts
foot-pounds per minute	2.26×10^{-5}	kilowatts
horsepower	0.7457	kilowatts
miles per hour	0.8689762	knots
feet	1.893939×10^{-4}	miles
miles	0.8689762	nautical miles
degrees	1.745329×10^{-2}	radians
acres	43560	square feet
BTU per minute	17.5796	watts

10.23.8 PHYSICAL CONSTANTS

1. c (speed of light) = 299,792,458 m/s (exact value)
2. e (charge of electron) $\approx 1.6021764 \times 10^{-19}$ C
3. $\hbar$ (Plank constant over 2π) $\approx 1.0545716 \times 10^{-34}$ J s
4. Acceleration, sea level, latitude 45° ≈ 9.806194 m/s^2 ≈ 32.1726 ft/s^2
5. Avogadro's number $\approx 6.022142 \times 10^{23}$
6. Density of mercury, at 0°C ≈ 13.5951 g/mL
7. Density of water (maximum), at 3.98°C ≈ 0.99997496 g/mL
8. Density of water, at 0°C ≈ 0.9998426 g/mL
9. Density of dry air, at 0°C, 760 mm of Hg ≈ 1.2927 g/L
10. Heat of fusion of water, at 0°C ≈ 333.6 J/g
11. Heat of vaporization of water, at 100°C ≈ 2256.8 J/g
12. Mass of hydrogen atom $\approx 1.67353 \times 10^{-24}$ g
13. Velocity of sound, dry air, at 0°C ≈ 331.36 m/s ≈ 1087.1 ft/s

10.24 VOTING POWER

A *weighted voting* game is represented by the vector $[q; w_1, w_2, \dots, w_n]$:

1. There are n players.
2. Player i has w_i votes (with $w_i > 0$).
3. A *coalition* is a subset of players.
4. A coalition S is *winning* if $\sum_{i \in S} w_i \geq q$, where q is the *quota*.
5. A game is *proper* if $\frac{1}{2} \sum w_i < q$.

A player

1. Can have *veto power*: no coalition can win without this player
2. Can be a *dictator*: has more votes than the quota
3. Can be a *dummy*: cannot affect any coalitions

10.24.1 SHAPLEY–SHUBIK POWER INDEX

Consider all permutations of players. Scan each permutation from beginning to end adding together the votes that each player contributes. Eventually a total of at least q will be arrived at, this occurs at the *pivotal* player. The *Shapley–Shubik power index* (ϕ) of player i is the number of permutations for which player i is pivotal; divided by the total number of permutations.

EXAMPLE

1. Consider the $[5; 4, 2, 1, 1]$ game, the players are $\{$A,B,C,D$\}$.
2. For the $4! = 24$ permutations of four players the pivotal player is underlined:

A$\underline{\text{B}}$CD	B$\underline{\text{A}}$CD	C$\underline{\text{A}}$BD	D$\underline{\text{A}}$BC
A$\underline{\text{B}}$DC	B$\underline{\text{A}}$DC	C$\underline{\text{A}}$DB	D$\underline{\text{A}}$CB
A$\underline{\text{C}}$BD	BC$\underline{\text{A}}$D	CB$\underline{\text{A}}$D	DB$\underline{\text{A}}$C
A$\underline{\text{C}}$DB	BCD$\underline{\text{A}}$	CBD$\underline{\text{A}}$	DBC$\underline{\text{A}}$
A$\underline{\text{D}}$BC	BD$\underline{\text{A}}$C	CD$\underline{\text{A}}$B	DC$\underline{\text{A}}$B
A$\underline{\text{D}}$CB	BDC$\underline{\text{A}}$	CDB$\underline{\text{A}}$	DCB$\underline{\text{A}}$

3. Hence player A has power $\phi(\text{A}) = \frac{18}{24} = 0.75$.
4. The other three players have equal power of $\frac{2}{24} \approx 0.083$.

10.24.2 BANZHAF POWER INDEX

Consider all 2^N possible coalitions of players. For each coalition, if player i can change the winning-ness of the coalition, by either entering it or leaving it, then i is *marginal* or *swing*. The *Banzhaf power index* (β) of player i is proportional to the number of times he is marginal; the total power of all players is 1.

EXAMPLE

1. Consider the $[5; 4, 2, 1, 1]$ game, the players are $\{$A,B,C,D$\}$.
2. There are 16 subsets of four players; each player is "in" (`I`) or is "out" (`O`) of a coalition. For each coalition the marginal players are listed

`O O O O`	$\Rightarrow$	$\{\}$	`I O O O`	$\Rightarrow$	$\{$B,C,D$\}$
`O O O I`	$\Rightarrow$	$\{$A$\}$	`I O O I`	$\Rightarrow$	$\{$A,D$\}$
`O O I O`	$\Rightarrow$	$\{$A$\}$	`I O I O`	$\Rightarrow$	$\{$A,C$\}$
`O O I I`	$\Rightarrow$	$\{$A$\}$	`I O I I`	$\Rightarrow$	$\{$A$\}$
`O I O O`	$\Rightarrow$	$\{$A$\}$	`I I O O`	$\Rightarrow$	$\{$A,B$\}$
`O I O I`	$\Rightarrow$	$\{$A$\}$	`I I O I`	$\Rightarrow$	$\{$A$\}$
`O I I O`	$\Rightarrow$	$\{$A$\}$	`I I I O`	$\Rightarrow$	$\{$A$\}$
`O I I I`	$\Rightarrow$	$\{$A$\}$	`I I I I`	$\Rightarrow$	$\{$A$\}$

3. Player A is marginal 14 times, and has power $\beta(\text{A}) = \frac{14}{20} = 0.7$
4. The players B, C, and D are each marginal 2 times and have equal power of $\frac{2}{20} = 0.1$.

10.24.2.1 Voting power examples

1. For the game $[51; 49, 48, 3]$ the winning coalitions are: $\{1, 2, 3\}$, $\{1, 2\}$, $\{1, 3\}$ and $\{2, 3\}$. These are the same winning coalitions as the game $[3; 1, 1, 1]$. Hence, all players have equal power even though the number of votes each player has is different.
2. The original EEC (1958) had France, Germany, Italy, Belgium, The Netherlands, and Luxembourg. They voted as $[12; 4, 4, 4, 2, 2, 1]$. Therefore:
$$\phi = \frac{1}{60}(14, 14, 14, 9, 9, 0) \quad \text{and} \quad \beta = \frac{1}{21}(5, 5, 5, 3, 3, 0). \tag{10.24.1}$$
3. The *UN security council* has 15 members. The five permanent members have veto power. For a motion to pass, it must be supported by at least 9 members of the council and it must not be vetoed. A game representation is: $[39; \underbrace{7, 7, 7, 7, 7}_{\text{5 members}}, \underbrace{1, 1, 1, 1, 1, 1, 1, 1, 1, 1}_{\text{10 members}}]$.
   (a) Shapley–Shubik powers: of each permanent member $\phi_{\text{major}} = \frac{421}{2145} \approx 0.196$, of each minor member $\phi_{\text{minor}} = \frac{4}{2145} \approx 0.002$.
   (b) Banzhaf powers: $\beta_{\text{major}} = \frac{106}{635} \approx 0.167$ and $\beta_{\text{minor}} = \frac{21}{1270} \approx 0.017$.
4. Changing the quota in a game may change the powers of the players:
   (a) For the game $[6; 4, 3, 2]$ have $\phi = \frac{1}{6}(4, 1, 1)$.
   (b) For the game $[7; 4, 3, 2]$ have $\phi = \frac{1}{2}(1, 1, 0)$.
   (c) For the game $[8; 4, 3, 2]$ have $\phi = \frac{1}{3}(1, 1, 1)$.
5. For the n-player game $[q; a, 1, 1, 1, 1, \dots, 1]$ with $1 < a < q \le n$ and $n < 2q + 1 - a$ the powers are $\phi_{\text{major}} = a/n$ and $\phi_{\text{minor}} = (n - a)/n(n - 1)$.
6. Four-person committee, one member is chair. Use majority rule until deadlock, then chair decides. This is a $[3; 2, 1, 1, 1]$ game so that $\phi = \frac{1}{6}(3, 1, 1, 1)$.
7. Five-person committee, with two co-chairs. Need a majority, and at least one co-chair. This is a $[7; 3, 3, 2, 2, 2]$ game so that $\phi = \frac{1}{12}(3, 3, 2, 2, 2)$ and $\beta = \frac{1}{29}(7, 7, 5, 5, 5)$

10.25 GREEK ALPHABET

Greek letter			Greek name	English equivalent
A	α		Alpha	a
B	β		Beta	b
Γ	γ		Gamma	g
Δ	δ		Delta	d
E	ϵ	ε	Epsilon	e
Z	ζ		Zeta	z
H	η		Eta	e
Θ	θ	ϑ	Theta	th
I	ι		Iota	i
K	κ		Kappa	k
Λ	λ		Lambda	l
M	μ		Mu	m

Greek letter			Greek name	English equivalent
N	ν		Nu	n
Ξ	ξ		Xi	x
O	o		Omicron	o
Π	π	ϖ	Pi	p
P	ρ	ϱ	Rho	r
Σ	σ	ς	Sigma	s
T	τ		Tau	t
Υ	υ		Upsilon	u
Φ	ϕ	φ	Phi	ph
X	χ		Chi	ch
Ψ	ψ		Psi	ps
Ω	ω		Omega	o

10.26 BRAILLE CODE

0	1	2	3	4
· ● · ● · ● ● ● ● ● · ·	· ● ● · · ● · · ● ● · ·	· ● ● ● · ● ● · ● ● · ·	· ● ● ● · ● · · ● ● · ·	· ● ● ● · ● · ● ● ● · ·

5	6	7	8	9
· ● ● · · ● · ● ● ● · ·	· ● ● ● · ● ● · ● ● · ·	· ● ● ● · ● ● ● ● ● · ·	· ● ● · · ● ● ● ● ● · ·	· ● · ● · ● ● · ● ● · ·

10.27 MORSE CODE

A	• —
B	— • • •
C	— • — •
D	— • •
E	•
F	• • — •
G	— — •
H	• • • •
I	• •
J	• — — —
K	— • —
L	• — • •
M	— —

N	— •
O	— — —
P	• — — •
Q	— — • —
R	• — •
S	• • •
T	—
U	• • —
V	• • • —
W	• — —
X	— • • —
Y	— • — —
Z	— — • •

Period	• — • — • —
Comma	— — • • — —
Question	• • — — • •
1	• — — — —
2	• • — — —
3	• • • — —
4	• • • • —
5	• • • • •
6	— • • • •
7	— — • • •
8	— — — • •
9	— — — — •
0	— — — — —

10.28 BAR CODES

All the codes shown here used “123456789” as the input.

1. Aztec
2. Code One
3. Code 11
4. Code 39
5. Code 128
6. Databar-14 Stack
7. Grid Matrix
8. ISBN
9. KIX Code
10. Maxicode
11. Micro QR code
12. PDF 417
13. Postnet
14. QR
15. UPC-A

List of References

Chapter 1 *Numbers and Elementary Mathematics*

1. G. H. Hardy, E. M. Wright, A. Wiles, R. Heath-Brown, J. Silverman, *An Introduction to the Theory of Numbers*, Oxford University Press, USA, 2008.
2. L. B. W. Jolley, *Summation of Series*, Dover Publications, New York, 1961.
3. S. G. Krantz, *The Elements of Advanced Mathematics*, CRC Press, Boca Raton, FL, 1995.
4. R. J. Mathar, *The series limit of* $\sum_k 1/[k \log k (\log \log k)^2]$, *ArXiv e-prints*, http://arxiv.org/abs/0902.0789, accessed 11 December 2016.
5. R. B. Nelsen, *Proofs Without Words: Exercises in Visual Thinking*, The Mathematical Association of America, Washington D.C., 1997.
6. R. B. Nelsen, *Proofs Without Words II: More Exercises in Visual Thinking*, The Mathematical Association of America, Washington D.C., 2001.
7. P. Ribenboim, *The book of prime number records*, Springer-Verlag, New York, 1988.
8. L. D. Servi, "Nested Square Roots of 2," *American Mathematical Monthly*, volume 110, number 4, April 2003.
9. N. J. A. Sloane and S. Plouffe, *Encyclopedia of Integer Sequences*, Academic Press, New York, 1995.

Chapter 2 *Algebra*

1. D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton University Press, 2005.
2. I. N. Herstein, *Topics in Algebra*, 2nd edition, John Wiley & Sons, New York, 1975.
3. G. Strang, *Introduction to Linear Algebra*, Third Edition, Wellesley Cambridge Press, Wellesley, MA, 2003.

Chapter 3 *Discrete Mathematics*

1. B. Bollobás, *Graph Theory*, Springer-Verlag, Berlin, 1979.
2. C. J. Colbourn and J. H. Dinitz, *Handbook of Combinatorial Designs*, CRC Press, Boca Raton, FL, 1996.
3. J. Gross, *Handbook of Graph Theory & Applications*, CRC Press, Boca Raton, FL, 1999.
4. K. H. Rosen, *Handbook of Discrete and Combinatorial Mathematics*, CRC Press, Boca Raton, FL, 2000.
5. R. P. Stanley, *Enumerative Combinatorics*, Volumes 1 & 2, Cambridge University Press, New York, 1997 & 2001.

Chapter 4 *Geometry*

1. A. Gray, *Modern Differential Geometry of Curves and Surfaces*, CRC Press, Boca Raton, FL, 1993.
2. J. O'Rourke and J. E. Goodman, *Handbook of Discrete and Computational Geometry*, CRC Press, Boca Raton, FL, 1997.
3. D. J. Struik, *Lectures in Classical Differential Geometry*, 2nd edition, Dover, New York, 1988.

Chapter 5 *Analysis*

1. J. W. Brown and R. V. Churchill, *Complex Variables and Applications*, 6th edition, McGraw-Hill, New York, 1996.
2. A. G. Butkovskiy, *Green's Functions and Transfer Functions Handbook*, Halstead Press, John Wiley & Sons, New York, 1982.
3. G. F. Carrier, M. Krook, and C. E. Pearson, *Functions of a Complex Variable: Theory and Technique*, McGraw–Hill, 1966.
4. I. S. Gradshteyn and M. Ryzhik, *Tables of Integrals, Series, and Products*, D. Zwillinger (ed), 8th edition, Academic Press, Orlando, FL, 2014.
5. A. J. Jerri, *Introduction to Integral Equations with Applications*, Marcel Dekker, New York, 1985.
6. S. G. Krantz, *Real Analysis and Foundations*, CRC Press, Boca Raton, FL, 1991.
7. P. Moon and D. E. Spencer, *Field Theory Handbook*, Springer-Verlag, Berlin, 1961.
8. A. D. Polyanin and V. F. Zaitsev, *Handbook of Exact Solution for Ordinary Differential Equations*, CRC Press, Boca Raton, FL, 1995.
9. J. L. Synge and A. Schild, *Tensor Calculus*, University of Toronto Press, Toronto, 1949.
10. D. Zwillinger, *Handbook of Differential Equations*, 3rd edition, Academic Press, Orlando, FL, 1997.
11. D. Zwillinger, *Handbook of Integration*, Jones & Bartlett, Boston, 1992.

Chapter 6 *Special Functions*

1. Staff of the Bateman Manuscript Project, A. Erdélyi, Editor, *Tables of Integral Transforms*, in 3 volumes, McGraw-Hill, New York, 1954.
2. W. Magnus, F. Oberhettinger, and R. P. Soni, *Formulas and Theorems for the Special Functions of Mathematical Physics*, Springer-Verlag, New York, 1966.
3. N. M. Temme, *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*, Wiley-Interscience, Indianapolis, IN, 1996.
4. N. I. A. Vilenkin, *Special Functions and the Theory of Group Representations*, American Mathematical Society, Providence, RI, 1968.

Chapter 7 *Probability and Statistics*

1. W. Feller, *An Introduction to Probability Theory and Its Applications*, Volume 1, John Wiley & Sons, New York, 1968.
2. S. K. Park and K. W. Miller, "Random number generators: good ones are hard to find," *Communications of the ACM*, October 1988, 31, 10, pages 1192–1201.
3. D. Zwillinger and S. Kokoska, *Standard Probability and Statistics Tables and Formulae*, Chapman & Hall/CRC, Boca Raton, FL, 2000.
4. MIL-STD-1916, Department of Defense Test Method Standard – "DoD Preferred Methods For Acceptance of Product", 1 April 1996, http://www.sqconline.com/sites/sqconline.com/files/MIL-STD-1916.pdf, accessed 1 Jan 2017.

Chapter 8 *Scientific Computing*

1. R. L. Burden and J. D. Faires, *Numerical Analysis*, 9th edition, Brooks Cole, Florence, KY, 2010.
2. G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed., The Johns Hopkins University Press, Baltimore, 1996.
3. J. Marin, "Computing Columns, Footings and Gates through Moments of Area," *Computers & Structures*, Vol. 18, No. 2, Feb 1984, pages 343–349.
4. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd edition, Cambridge University Press, New York, 2007.

Chapter 9 *Mathematical Formulas from the Sciences*

1. A. D. Polyanin and A. I. Chernoutsan, *A Concise Handbook of Mathematics, Physics, and Engineering Sciences*, Chapman & Hall/CRC, Boca Raton, FL, 2011.
2. "Greeks (finance)", *Wikipedia, the Free Encyclopedia*, https://en.wikipedia.org/wiki/Greeks_(finance), accessed 1 March 2017.

Chapter 10 *Miscellaneous*

1. I. Daubechies, *Ten Lectures on Wavelets*, SIAM Press, Philadelphia, 1992.
2. D. E. Knuth, "Two Notes on Notation", *ArXiv e-prints*, http://arxiv.org/abs/math/9205211, accessed 11 December 2016.
3. F. Glover, "Tabu Search: A Tutorial," *Interfaces*, 20(4), pages 74–94, 1990.
4. D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison–Wesley, Reading, MA, 1989.
5. J. P. Lambert, "Voting Games, Power Indices, and Presidential Elections," *The UMAP Journal*, Module 690, **9**, No. 3, pages 214–267, 1988.
6. C. Livingston, *Knot Theory*, The Mathematical Association of America, Washington, D.C., 1993.
7. D. Luce and H. Raiffa. *Games and Decision Theory*, Wiley, New York, 1957.
8. F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, Amsterdam, 1977.
9. A. O'Maley, "Elliptic Curves and Elliptic Curve Cryptography," *B. S. Undergraduate Mathematics Exchange*, Vol. 3, No. 1, pages 16–24, Fall 2005.
10. G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, MA, 1995.
11. E. W. Weisstein, *CRC Concise Encyclopedia of Mathematics*, 2nd edition, Chapman & Hall/CRC, Boca Raton, FL, 2002.
12. S. Wolfram, *A New Kind of Science*, Wolfram Media, IL, 2002.

List of Figures

List of Notations

Symbols

Greek Letters

Numbers

A

B

C

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S

T

U

V

W

X

Y

Z

Index

A

B

C

D

E

F

H

I

J

K

L

M

N

O

P

Q

T

U

V

W

X

Y

Z