The Open University

M303

Further pure mathematics

Handbook

This publication forms part of the Open University module M303 *Further pure mathematics*. Details of this and other Open University modules can be obtained from Student Recruitment, The Open University, PO Box 197, Milton Keynes MK7 6BJ, United Kingdom (tel. +44 (0)300 303 5303; email general-enquiries@open.ac.uk).

Alternatively, you may visit the Open University website at www.open.ac.uk where you can learn more about the wide range of modules and packs offered at all levels by The Open University.

The Open University, Walton Hall, Milton Keynes, MK7 6AA.

First published 2015. Second edition 2016.

Edited, designed and typeset by The Open University, using the Open University TEX System.

Printed in the United Kingdom by Halstan & Co. Ltd, Amersham, Bucks.

SUP 05188 5

2.1

Contents

Quick reference

1 Greek alphabet

α	A	alpha	ι	I	iota	ρ	P	rho
β	B	beta	κ	K	kappa	σ	Σ	sigma
γ	Γ	gamma	λ	Λ	lambda	τ	T	tau
δ	Δ	delta	μ	M	mu	υ	Υ	upsilon
ε	E	epsilon	ν	N	nu	ϕ	Φ	phi
ζ	Z	zeta	ξ	Ξ	xi	χ	X	chi
η	H	eta	o	O	omicron	ψ	Ψ	psi
θ	Θ	theta	π	Π	pi	ω	Ω	omega

2 Mathematical language

In mathematics we commonly use **implications** such as 'if P then Q', where P and Q are statements that can be either true or false. The statement P is the **hypothesis** and the statement Q is the **conclusion**. An implication is **true** if the conclusion is true whenever the hypothesis is true, and **false** otherwise. For example, the following implication is true:

if x is positive, then $x+1$ is positive.

There are various equivalent ways of stating such an implication:

(a) if $x>0$, then $x+1>0$

(b) $x>0 \implies x+1>0$

(c) for all $x>0$, we have $x+1>0$

(d) $x+1>0$, for all $x>0$

(e) $x+1>0$, whenever $x>0$

(f) for $x+1$ to be positive, it is sufficient that x be positive.

The **converse** of an implication is obtained by exchanging the hypothesis and the conclusion. For example, the converse of the (true) implication

if $x>0$, then $x+1>0$ is if $x+1>0$, then $x>0$,

which is false (try $x=0$).

An **equivalence** consists of an implication 'if P then Q' *and* its converse 'if Q then P'. The equivalence is true if both these implications are true. For example, the following equivalence is true:

$x>0$ is equivalent to $2x>0$.

It could alternatively be stated as follows:

(a) $x>0 \iff 2x>0$

(b) $x>0$ if, and only if, $2x>0$

(c) $x>0$ is necessary and sufficient for $2x>0$.

There are three ways of proving implications.

1. **Direct proof**: we begin by assuming that the hypothesis is true and then argue directly to show that the conclusion is true.
2. **Proof by contraposition**: we begin by assuming that the conclusion is false and then argue directly to show that the hypothesis is false.
3. **Proof by contradiction**: we begin by assuming that the hypothesis is true *and* that the conclusion is false, and then argue from both to obtain a contradiction.

Example To prove the implication 'if n is even then n^2 is even'.

1. **Direct proof**: assume that n is even and argue to deduce that n^2 is necessarily even.
2. **Proof by contraposition**: assume that n^2 is odd and argue to deduce that n is necessarily odd.
3. **Proof by contradiction**: assume that n is even *and* n^2 is odd, and argue to obtain a contradiction.

A **counterexample** to an implication satisfies the hypothesis of the implication but not the conclusion. Therefore *any* one counterexample is sufficient to disprove an implication, that is, to show that the implication is false.

Example The counterexample $m = n = 1$ disproves the implication 'if both m and n are odd then the sum $m + n$ is odd'.

3 Set notation

Notation	Meaning
$\{x, y, \ldots, z\}$	The set of elements listed in $\{\ldots\}$
$\{x : \ldots\}$	The set of all x such that $\ldots$ holds
$x \in A$	x belongs to A
$x \notin A$	x does not belong to A
$A \subseteq B$	A is a subset of B: each element of A belongs to B
$A = B$	A is equal to B: $A \subseteq B$ and $B \subseteq A$
$A \subset B$	A is a proper subset of B: $A \subseteq B$ but $A \neq B$
$A \cup B$	A union B: the set of all elements that belong to A or B (or both)
$A \cap B$	A intersection B: the set of all elements that belong to both A and B
$A - B$ or $A \backslash B$	A minus B: the set of all elements of A that do not belong to B
$\emptyset$	The empty set

Some texts use $A \subset B$ to mean $A \subseteq B$.

4 Real numbers

Definitions

A **real number** is a number that can be represented by a decimal of the form

$$\pm a_0.a_1a_2a_3\ldots,$$

where a_0 is a non-negative integer and a_1, a_2, a_3, ... are digits. **Rational numbers** (ratios of integers) are represented by recurring decimals and **irrational numbers** are represented by non-recurring decimals. Real numbers are often represented by points on a line, called the **real line**.

Some important subsets of the real numbers

Symbol	Subset
$\mathbb{N}$	The set of all natural numbers: $1, 2, 3, \ldots$
$\mathbb{Z}$	The set of all integers: $0, \pm1, \pm2, \pm3, \ldots$
$\mathbb{Q}$	The set of all rational numbers (numbers of the form p/q, where $p \in \mathbb{Z}, q \in \mathbb{N}$)
$\mathbb{R}$	The set of all real numbers
$\mathbb{R} - \mathbb{Q}$	The set of all irrational numbers (real numbers that are not rational, e.g. $\sqrt{2}$, π, e)
(a, b)	$\{x : a < x < b\}$, the open interval from a to b
$[a, b]$	$\{x : a \leq x \leq b\}$, the closed interval from a to b
$(a, b]$	$\{x : a < x \leq b\}$
$[a, b)$	$\{x : a \leq x < b\}$
$[a, \infty)$	$\{x : x \geq a\}$
(a, ∞)	$\{x : x > a\}$
$(-\infty, a)$	$\{x : x < a\}$
$(-\infty, a]$	$\{x : x \leq a\}$

Density property of the reals

If $a, b \in \mathbb{R}$ and $a < b$, then there is a rational number x and an irrational number y such that

$$a < x < b \text{ and } a < y < b.$$

Upper and lower bounds

Suppose that A is a non-empty subset of $\mathbb{R}$. Then A is **bounded above** if there is a real number M such that

$$x \leq M, \quad \text{for all } x \in A.$$

The number M is called an **upper bound** of A. Clearly, any number bigger than M is also an upper bound of A. A **lower bound** of A is defined similarly.

Among all upper bounds of A, the smallest (which always exists if A is bounded above) is called the **least upper bound** of A, or the **supremum** of A, written $\boldsymbol{\sup A}$. The **greatest lower bound**, or the **infimum**, of A, written $\boldsymbol{\inf A}$, is defined similarly.

If A has infinitely many elements, then $\sup A$ and $\inf A$ may or may not belong to A. In contrast, if A has finitely many elements, then $\sup A$ and $\inf A$ are the largest and the smallest elements of A, respectively, and therefore are members of A.

If $\sup A$ belongs to A, then $\sup A$ is the maximal element in A and we may denote it by $\boldsymbol{\max A}$. Similarly, if $\inf A$ belongs to A, then we may denote it by $\boldsymbol{\min A}$.

Example If A is the interval $[-1, 1)$, then

$$\sup A = 1 \quad \text{and} \quad \inf A = \min A = -1.$$

If $B = \{-1, 0, 1, 2\}$, then

$$\sup B = \max B = 2 \quad \text{and} \quad \inf B = \min B = -1.$$

Inequalities

Rules for rearranging inequalities

For all a, b, $c \in \mathbb{R}$,

(a) $a < b \iff b - a > 0$.

(b) $a < b \iff a + c < b + c$.

(c) If $c > 0$, then $a < b \iff ac < bc$;
if $c < 0$, then $a < b \iff ac > bc$.

(d) If $a, b > 0$, then $a < b \iff \frac{1}{a} > \frac{1}{b}$.

(e) If $a, b \geq 0$ and $p > 0$, then $a < b \iff a^p < b^p$.

Corresponding versions of these inequalities exist with *strict* inequalities replaced by *weak* inequalities; that is, with $a < b$ replaced by $a \leq b$.

The **solution set** of an inequality involving an unknown real number x is the set of values of x for which the inequality holds.

Rules for deducing new inequalities from given ones

(a) For all $a, b, c \in \mathbb{R}$,

$$a < b \text{ and } b < c \implies a < c. \qquad \text{Transitivity Rule}$$

(b) For all $a, b, c, d \in \mathbb{R}$, if $a < b$ and $c < d$ then

$$a + c < b + d, \qquad \text{Sum Rule}$$
$$ac < bd \quad (\text{provided that } a, c \geq 0). \qquad \text{Product Rule}$$

Modulus

If $x \in \mathbb{R}$ then the **modulus**, or **absolute value**, of x is

$$|x| = \begin{cases} x, & \text{if } x \geq 0, \\ -x, & \text{if } x < 0. \end{cases}$$

Thus $|x|$ is the distance from the origin to x, and so

(a) $|x| < a \iff -a < x < a$

(b) $|x| > a \iff x > a$ or $x < -a$

(c) the distance on the real line from a to b is $|b - a| = |a - b|$.

5 Equivalence relations

A **relation** $\sim$ on a set X is a rule such that, for any two elements $x, y \in X$, it is possible to determine whether x is related to y:

if x is related to y, we write $x \sim y$;

if x is not related to y, we write $x \nsim y$.

An **equivalence relation** on a set X is a relation $\sim$ on X that satisfies the following three axioms.

E1 Reflexive For all $x \in X$, $x \sim x$.

E2 Symmetric For all $x, y \in X$, if $x \sim y$, then $y \sim x$.

E3 Transitive For all $x, y, z \in X$, if $x \sim y$ and $y \sim z$, then $x \sim z$.

Let $\sim$ be an equivalence relation defined on a set X; then the **equivalence class** of $x \in X$, denoted by $[x]$, is the set $[x] = \{y \in X : x \sim y\}$.

An equivalence relation on a set X **partitions** X into equivalence classes such that:

- each $x \in X$ is in an equivalence class
- any two equivalence classes are either the same or disjoint (that is, have no elements in common).

6 Functions

Definitions

A **function** f is defined by specifying:

- a set A, the **domain** of f
- a set B, the **codomain** of f
- a **rule** $x \mapsto f(x)$ that associates with each element $x \in A$ a *unique* element $f(x) \in B$.

The element $f(x)$ is the **image** of x under f, and the set

$$f(A) = \{f(x) : x \in A\} \subseteq B$$

is the **image set** of A under f.

The function f is **onto** if the image $f(A)$ is equal to the codomain B.

The function f is **one–one** if each element of the image $f(A)$ is the image of exactly one element of A; that is, if $x_1, x_2 \in A$ and $f(x_1) = f(x_2)$, then $x_1 = x_2$.

Standard functions

Type	Rule	Domain
Polynomial functions	$p(x) = a_0 + a_1x + \cdots + a_nx^n (a_i \in \mathbb{R})$	$\mathbb{R}$
Rational functions	$p(x)/q(x)$, p and q polynomial functions (q not the zero function)	$\mathbb{R} - \{x : q(x) = 0\}$
Trigonometric functions	$\sin x$ $\cos x$ $\tan x$	$\mathbb{R}$ $\mathbb{R}$ $\mathbb{R} - \{(n + \frac{1}{2})\pi : n \in \mathbb{Z}\}$
Exponential functions	e^x a^x, where $a > 0$	$\mathbb{R}$ $\mathbb{R}$
Natural log function	$\log x$	$\{x : x > 0\}$

Remarks

1. Definition of e^x:

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

2. Definition of $\log x$:

$$x = e^y \iff y = \log x;\ \log x = \int_1^x t^{-1}\, dt.$$

3. Definition of a^x (for $a > 0$):

$$a^x = \exp(x \log a).$$

4. Index laws:

$$a^{x+y} = a^x a^y, \quad (a^x)^y = a^{xy}, \quad \sqrt[n]{a^m} = a^{m/n}, \quad a^{-x} = \frac{1}{a^x}.$$

5. Logarithmic identities:

$$\log xy = \log x + \log y, \quad \log\left(\frac{1}{x}\right) = -\log x.$$

The plane $\mathbb{R}^2$

Points in the plane $\mathbb{R}^2$ have two coordinates, and are often denoted using subscript notation such as (x_1, x_2). We also use the notation $\mathbf{x}$ for the point (x_1, x_2) and $\mathbf{0}$ for the point $(0, 0)$. We refer to the point $\mathbf{0}$ as the origin of $\mathbb{R}^2$.

Functions defined on the plane have domain $\mathbb{R}^2$. To specify the rule of a function f with domain $\mathbb{R}^2$, we use the notation $f(x_1, x_2)$ or $f(\mathbf{x})$.

Higher dimensions

The space $\mathbb{R}^n$ has dimension n. Points in $\mathbb{R}^n$ have n coordinates and are denoted by ***n*-tuples** such as $(x_1, x_2, \ldots, x_n)$. As in two dimensions, we use the notation $\mathbf{x}$ for the point $(x_1, x_2, \ldots, x_n)$ and $\mathbf{0}$ for the point $(0, 0, \ldots, 0)$, the origin of $\mathbb{R}^n$.

7 Calculus

Derivatives and integrals of standard functions

Function $f(x)$	Derivative $f'(x)$
x^n	nx^{n-1}
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
$\tan x$	$\sec^2 x$
e^x	e^x
$\log x$	$1/x$

Function $f(x)$	Integral $\int f(x)\,dx$
x^n $(n \neq -1)$	$x^{n+1}/(n+1)$
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
e^x	e^x
x^{-1}	$\log \lvert x \rvert$

Combination rules for differentiation

Let f and g be defined on an interval I, and $c \in I$. Then if f and g are differentiable at c, so are:

Sum Rule $f + g$, and
$(f+g)'(c) = f'(c) + g'(c)$;

Multiple Rule λf, for $\lambda \in \mathbb{R}$, and
$(\lambda f)'(c) = \lambda f'(c)$;

Product Rule fg, and
$(fg)'(c) = f'(c)g(c) + f(c)g'(c)$;

Quotient Rule f/g, provided that $g(c) \neq 0$, and
$$\left(\frac{f}{g}\right)'(c) = \frac{g(c)f'(c) - f(c)g'(c)}{(g(c))^2}.$$

Composition Rule for differentiation

Let f and g be defined on the intervals I and J, respectively, and let $c \in I$ and $f(I) \subseteq J$. If f is differentiable at c, and g is differentiable at $f(c)$, then $g \circ f$ is differentiable at c, and

$$(g \circ f)'(c) = g'(f(c))f'(c).$$

Inverse Function Rule

Let f be a function whose domain contains an interval I on which f is continuous and strictly monotonic, with image $J = f(I)$. If f is differentiable on I, and $f'(x) \neq 0$ for $x \in I$, then f^{-1} is differentiable on J. Also, if $c \in I$ and $d = f(c)$, then

$$\left(f^{-1}\right)'(d) = \frac{1}{f'(c)}.$$

Mean Value Theorem

Let f be continuous on the closed interval $[a, b]$ and differentiable on the open interval (a, b). Then there exists a point c in (a, b) such that

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

Combination rules for integration

If f and g are integrable on $[a, b]$, then so are:

Sum Rule $f + g$, and
$\int_a^b (f + g) = \int_a^b f + \int_a^b g$;

Multiple Rule λf, for $\lambda \in \mathbb{R}$, and
$\int_a^b \lambda f = \lambda \int_a^b f$;

Product Rule fg;

Quotient Rule f/g, provided that $1/g$ is bounded on $[a, b]$.

Fundamental Theorem of Calculus

Let f be a function defined on an interval I. Then the function F is a **primitive** of f on I if F is differentiable on I and

$$F' = f.$$

Let f be integrable on $[a, b]$, and let F be a primitive of f on $[a, b]$. Then

$$\int_a^b f = F(b) - F(a).$$

Sometimes we write $[F(x)]_a^b$ instead of $F(b) - F(a)$.

Inequalities for integrals

Let f be integrable on $[a, b]$. If $f(x) > 0$ on $[a, b]$, then $\int_a^b f > 0$. In particular, if f is continuous and non-negative on $[a, b]$, and if $\int_a^b f = 0$, then $f(x) = 0$ for all $x \in [a, b]$.

Let f and g be integrable on $[a, b]$. If $f(x) \leq g(x)$ for $x \in [a, b]$, then

$$\int_a^b f \leq \int_a^b g.$$

Let f be integrable on $[a, b]$. Then

$$\left| \int_a^b f \right| \leq \int_a^b |f|.$$

Book A: Number theory

Glossary

arithmetic progression The sequence of terms of an arithmetic series. **Chapter 1**, Section 2

arithmetic series A series of the form $a + (a + d) + (a + 2d) + \cdots$ with a common difference d from one term to the next. **Chapter 1**, Section 2

complete set of residues modulo n A set comprising one element from each of the n residue classes. **Chapter 3**, Section 1

cycle (of a decimal fraction) A block of digits that repeats indefinitely in a decimal fraction. The length of the cycle is the number of digits forming the cycle. **Chapter 4**, Section 2

Diophantine equation A polynomial equation in two or more variables for which we seek integer solutions. **Chapter 1**, Section 5

Fibonacci numbers The sequence F_n: $1, 1, 2, 3, 5, 8, 13, 21, 34, \ldots$ defined by $F_1 = 1, \quad F_2 = 1, \quad F_{n+2} = F_{n+1} + F_n, \quad$ for $n \geq 1$. **Chapter 2**, Section 5

geometric series A series of the form $a + ar + ar^2 + ar^3 + \cdots$ in which the ratio of successive terms is constant. **Chapter 1**, Section 3

hexagonal numbers The sequence arising from arrangements of dots forming hexagons: $1, \quad 1 + 5, \quad 1 + 5 + 9, \quad 1 + 5 + 9 + 13, \quad \ldots$. **Chapter 1**, Section 2

integral polynomial A polynomial in which the coefficients are integers: $P(x) = c_m x^m + c_{m-1} x^{m-1} + \cdots + c_1 x + c_0$. **Chapter 3**, Section 3

leading term of a polynomial $P(x)$ The (non-zero) term involving the highest power of x in the polynomial $P(x)$. **Chapter 4**, Section 4

linear congruence A congruence of the form $ax \equiv b \pmod{n}$, where a and b are integers, n is a positive integer and x is a variable. **Chapter 3**, Section 3

logarithmic integral The integral $\mathrm{Li}(x) = \int_2^x (\log t)^{-1} dt$. **Chapter 2**, Section 4

pentagonal numbers The sequence arising from arrangements of dots forming pentagons: $1, \quad 1 + 4, \quad 1 + 4 + 7, \quad 1 + 4 + 7 + 10, \quad \ldots$. **Chapter 1**, Section 2

$\pi(x)$ The number of primes not exceeding x. **Chapter 2**, Section 4

polygonal numbers The sequences of numbers arising from arrangements of dots in the shape of a given polygon, such as triangular numbers, square numbers, pentagonal numbers and so on. $P(k, n)$ denotes the nth term in the sequence of k-gonal numbers. **Chapter 1**, Section 2

polynomial congruence A congruence of the form $P(x) \equiv 0 \pmod{n}$, where $P(x)$ is an integral polynomial. **Chapter 3**, Section 3

prime decomposition The unique representation of an integer as a product of its prime factors, usually with the primes arranged in ascending order and like primes collected together as a power. **Chapter 2**, Section 1

Chapter 2, Section 4 **prime triplet (quartet)** A set of three (four) prime numbers among four (five) consecutive odd numbers.

Chapter 4, Section 1 **pseudoprime** A composite integer n that divides $2^n - 2$.

Chapter 1, Section 2 **pyramidal numbers** The numbers arising as sums of finite sequences of polygonal numbers. $Q(k, n)$ denotes the nth term in the sequence of k-gonal pyramidal numbers.

Chapter 3, Section 1 **residue class** The residue class modulo n of an integer a consists of all integers that are congruent to a modulo n.

Chapter 3, Section 1 **set of least positive residues modulo n** The set of n integers $\{0, 1, 2, \ldots, n-1\}$, each the least positive residue in a residue class modulo n.

Chapter 1, Section 2 **triangular number T_n** The sum of the first n natural numbers: $T_n = 1 + 2 + 3 + \cdots + n = \frac{1}{2}n(n+1)$.

Chapter 2, Section 4 **twin primes** Two consecutive odd numbers, both of which are prime.

Chapter 1: Foundations

Mathematical induction

The Well-Ordering Principle for $\mathbb{N}$

Every non-empty subset of $\mathbb{N}$ has a least member. In other words, if S is a non-empty subset of $\mathbb{N}$ then there exists $b \in S$ such that $b \leq n$ for all $n \in S$.

Theorem 3.1 *Principle of Induction*

If S is a set of natural numbers with the following two properties:

(a) 1 is a member of S

(b) if $k \in S$, then the next integer $k + 1 \in S$

then $S = \mathbb{N}$.

Principle of Mathematical Induction

Let $P(n)$ be a proposition depending on a natural number n. If:

(a) $P(1)$ is true (the *basis for the induction*)

(b) for any integer $k \geq 1$, if $P(k)$ is true then $P(k+1)$ is true (the *induction step* based on the *induction hypothesis*)

then $P(n)$ is true for all $n \in \mathbb{N}$.

Principle of Mathematical Induction (generalised)

Let $P(n)$ be a proposition depending on an integer n. If:

(a) $P(n_0)$ is true

(b) for any integer $k \geq n_0$, if $P(k)$ is true then $P(k+1)$ is true

then $P(n)$ is true for all integers $n \geq n_0$.

Second Principle of Mathematical Induction

Let $P(n)$ be a proposition depending on an integer n. If:

(a) $P(n_0)$ is true

(b′) for any integer $k \geq n_0$, if $P(n_0)$, $P(n_0+1)$, …, $P(k)$ are all true, then $P(k+1)$ is true

then $P(n)$ is true for all integers $n \geq n_0$.

Divisibility

Theorem 4.1 *The Division Algorithm*

For any two integers a and b, where $b > 0$, there exist unique integers q and r such that $a = bq + r$, where $0 \leq r < b$.

Definition 4.3 *Factors and multiples*

An integer a is divisible by the natural number b if there exists some integer q such that $a = bq$. We say that b **divides** a (written $b \mid a$), that b is a **factor** of a or that a is a **multiple** of b.

Theorem 4.4 *Properties of division*

Let a and b be natural numbers and c and d be any integers.

(a) If $a \mid c$ then $a \mid (c + na)$ for any integer n.

(b) If $c \neq 0$ and $a \mid c$ then $a \leq |c|$.

(c) If $a \mid b$ and $b \mid a$ then $a = b$.

(d) If $a \mid b$ and $b \mid c$ then $a \mid c$.

(e) If $a \mid c$ and $a \mid d$ then $a \mid (mc + nd)$ for any integers m and n.

Definition 4.5 *Highest common factor*

The **highest common factor**, $\text{hcf}(a, b)$, of two integers a and b, not both of which are zero, is the natural number n satisfying

(a) $n \mid a$ and $n \mid b$; (b) if $d \mid a$ and $d \mid b$ then $d \leq n$.

Definition 4.6 *Integer combination*

If a and b are integers, then any integer of the form $ma + nb$, with $m, n \in \mathbb{Z}$, is called an **integer combination** of a and b.

Proposition 4.7

Given any integers a and b, not both zero, there exist integers m and n such that $\text{hcf}(a, b) = ma + nb$. In other words, $\text{hcf}(a, b)$ is an integer combination of a and b.

Definition 4.8 *Coprime*

Two integers a and b, not both zero, are **coprime** or **relatively prime** whenever $\text{hcf}(a, b) = 1$.

Lemma 4.9

Integers a and b are coprime if, and only if, there exist integers m and n such that $1 = ma + nb$.

Proposition 4.10

For any integers a and b, not both zero, if $\text{hcf}(a,b) = d$ then $\frac{a}{d}$ and $\frac{b}{d}$ are integers such that $\text{hcf}\left(\frac{a}{d}, \frac{b}{d}\right) = 1$.

That is, the integers $\dfrac{a}{\text{hcf}(a,b)}$ and $\dfrac{b}{\text{hcf}(a,b)}$ are coprime.

Lemma 4.11

If $a \mid c$ and $b \mid c$, with $\text{hcf}(a,b) = 1$, then $ab \mid c$.

Theorem 4.12 *Euclid's Lemma*

If $a \mid bc$, with $\text{hcf}(a,b) = 1$, then $a \mid c$.

Definition 4.14 *Least common multiple*

The **least common multiple**, $\text{lcm}(a,b)$, of the non-zero integers a and b is the natural number n satisfying

(a) $a \mid n$ and $b \mid n$; (b) if $a \mid m$ and $b \mid m$ then $n \leq |m|$.

Proposition 4.15

For any pair of natural numbers a and b, $\text{lcm}(a,b) \times \text{hcf}(a,b) = ab$.

Linear Diophantine equations

Property of highest common factors

If $a = qb + r$, then $\text{hcf}(a,b) = \text{hcf}(b,r)$.

Theorem 5.4 *Solution of linear Diophantine equations*

The linear Diophantine equation $ax + by = c$ has solutions if, and only if, $\text{hcf}(a,b)$ divides c.

If this condition holds with $\text{hcf}(a,b) > 1$ then division by $\text{hcf}(a,b)$ simplifies the equation to $a'x + b'y = c'$, where $\text{hcf}(a',b') = 1$.

If x_0, y_0 is one solution of this equation then the general solution is $x = x_0 + b'k, \quad y = y_0 - a'k, \quad k \in \mathbb{Z}$.

Chapter 2: Prime numbers

The primes

Definition 1.1 *Prime numbers*

An integer n, where $n \geq 2$, is **prime** if it has no positive factor other than itself and 1. Otherwise n is **composite**.

Proposition 1.2

Each integer $n \geq 2$ is divisible by some prime number.

Proposition 1.3

If the integer $n \geq 2$ is composite, then it is divisible by some prime $p \leq \sqrt{n}$.

The value of hcf(n, p)

If p is a prime and n is any integer, then

$$\text{hcf}(n,p) = \begin{cases} p, & \text{if } p \text{ divides } n, \\ 1, & \text{otherwise.} \end{cases}$$

Theorem 1.4 *Euclid's Lemma for prime factors*

If p is prime and p divides the product $a_1a_2\cdots a_n$ then p divides a_i, for some integer i, where $1 \le i \le n$.

Corollary 1.5 *to Euclid's Lemma for prime factors*

If p, p_1, p_2, …, p_n are primes such that $p \mid p_1p_2\cdots p_n$, then $p = p_i$ for some i, where $1 \le i \le n$.

Theorem 1.7 *The Fundamental Theorem of Arithmetic*

Any integer $n \ge 2$ can be written uniquely in the form $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$, where p_i, $i = 1,\ldots,r$, are primes with $p_1 < p_2 < \cdots < p_r$ and each k_i, $i = 1,\ldots,r$, is a natural number.

The prime decomposition of integers

Proposition 2.1 *Factors of an integer*

Let n be an integer greater than 1, with prime decomposition $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$. Then the set of all factors of n is the set of integers $\left\{p_1^{h_1}p_2^{h_2}\cdots p_r^{h_r} : 0 \le h_i \le k_i \text{ for } i = 1,2,\ldots,r\right\}$.

Prime decomposition of hcf(m, n) and lcm(m, n)

If $m = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$ and $n = p_1^{l_1}p_2^{l_2}\cdots p_r^{l_r}$ then

- $\text{hcf}(m,n) = p_1^{\min\{k_1,l_1\}}p_2^{\min\{k_2,l_2\}}\cdots p_r^{\min\{k_r,l_r\}}$
- $\text{lcm}(m,n) = p_1^{\max\{k_1,l_1\}}p_2^{\max\{k_2,l_2\}}\cdots p_r^{\max\{k_r,l_r\}}$.

Definition 2.2 *The τ function*

For any integer $n \ge 1$, $\tau(n)$ is defined to be the number of distinct factors of n, including 1 and n.

Proposition 2.3 *The formula for $\tau(n)$*

For any integer $n \ge 2$, with prime decomposition $n = p_1^{k_1}p_2^{k_2}\cdots p_r^{k_r}$, we have $\tau(n) = (k_1+1)(k_2+1)\cdots(k_r+1)$.

The infinitude of the primes

Theorem 3.1 *Euclid's Theorem*

There are infinitely many primes.

Bertrand's Conjecture

For each integer $n \ge 2$, there exists a prime p such that $n < p < 2n$.

Lemma 3.3

Any integer of the form $4k + 3$ is divisible by a prime of this same form.

Theorem 3.4 *Primes of the form* $4k + 3$

There are infinitely many primes of the form $4k + 3$, where k is an integer.

Theorem 3.5 *Dirichlet's Theorem*

If a and b are positive integers with $\text{hcf}(a, b) = 1$, then there are infinitely many primes of the form $ak + b$, where k is a non-negative integer.

Famous problems concerning primes

The Goldbach Conjecture

Every even integer from 6 onwards can be written as a sum of two odd primes.

The Prime Number Theorem

$$\lim_{x \to \infty} \frac{\pi(x)}{\text{Li}(x)} = 1$$

Fibonacci numbers

Proposition 5.1 *The coprimality of consecutive Fibonacci numbers*

For any natural number n, $\text{hcf}(F_{n+1}, F_n) = 1$.

A useful relationship between Fibonacci numbers

For all integers $m \geq 2$ and $n \geq 1$, $F_{m+n} = F_{m-1}F_n + F_mF_{n+1}$.

Theorem 5.3 *Divisibility property of the Fibonacci numbers*

For all integers $n > 2$, F_m is divisible by F_n if, and only if, m is divisible by n.

Chapter 3: Congruence

Properties of congruence

Definition 1.1 *Congruence modulo* n

Let n be a fixed natural number and let a and b be any integers. Then a is **congruent to** b **modulo** n, written $a \equiv b \,(\text{mod } n)$, if the difference $a - b$ is divisible by n. Otherwise a is **incongruent to** b **modulo** n, written $a \not\equiv b \,(\text{mod } n)$.

Theorem 1.2 *Properties of congruence*

Let n be a fixed natural number and let a, b, c and d be any integers. Then the following properties hold.

(a) $a \equiv a \pmod{n}$.

(b) If $a \equiv b \pmod{n}$ then $b \equiv a \pmod{n}$.

(c) If $a \equiv b \pmod{n}$ and $b \equiv c \pmod{n}$ then $a \equiv c \pmod{n}$.

(d) If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ then $a + c \equiv b + d \pmod{n}$.

(e) If $a \equiv b \pmod{n}$ and $c \equiv d \pmod{n}$ then $ac \equiv bd \pmod{n}$.

(f) If $a \equiv b \pmod{n}$ then $a^r \equiv b^r \pmod{n}$, for any integer $r \geq 1$.

Proposition 1.4 *Congruent integers have the same remainders*

$a \equiv b \pmod{n}$ if, and only if, the integers a and b have the same remainder when divided by n.

Theorem 1.6

If $a \equiv b \pmod{m}$ and $a \equiv b \pmod{n}$, then $a \equiv b \pmod{\operatorname{lcm}(m, n)}$.

Corollary 1.7 *to Theorem 1.6*

If $a \equiv b \pmod{m}$ and $a \equiv b \pmod{n}$, where $\operatorname{hcf}(m, n) = 1$, then $a \equiv b \pmod{mn}$.

Proposition 1.8 *Cancellation Rule*

If $ca \equiv cb \pmod{n}$ then $a \equiv b \pmod{\frac{n}{d}}$, where $d = \operatorname{hcf}(c, n)$.

Corollary 1.9 *to the Cancellation Rule*

If $ca \equiv cb \pmod{n}$, where $\operatorname{hcf}(c, n) = 1$, then $a \equiv b \pmod{n}$.

Theorem 1.10 *Euclid's Lemma for prime factors*

If $ab \equiv 0 \pmod{p}$ then $a \equiv 0 \pmod{p}$ or $b \equiv 0 \pmod{p}$.

Divisibility tests

Proposition 2.1 *Divisibility by 9 and 11*

Let the integer N be written in decimal notation as $N = a_m a_{m-1} \cdots a_2 a_1 a_0$ so that the digits satisfy $0 \leq a_i \leq 9$ with $a_m \neq 0$. Then:

(a) $N \equiv a_0 + a_1 + a_2 + a_3 + \cdots + a_m \pmod{9}$

(b) $N \equiv a_0 - a_1 + a_2 - a_3 + \cdots + (-1)^m a_m \pmod{11}$.

Corollary 2.2 *to Proposition 2.1*

(a) A natural number N is divisible by 9 if, and only if, the sum of its digits is divisible by 9.

(b) A natural number N is divisible by 11 if, and only if, the alternating sum of its digits is divisible by 11, where by 'alternating sum' we mean the sum using alternating signs as given in part (b) of Proposition 2.1.

Proposition 2.3

A natural number N is divisible by 3 if, and only if, the sum of its digits is divisible by 3.

Linear congruences

Proposition 3.1

Let $P(x)$ be an integral polynomial. If $a \equiv b \pmod{n}$ then $P(a) \equiv P(b) \pmod{n}$. In particular, a is a solution of the polynomial congruence $P(x) \equiv 0 \pmod{n}$ if, and only if, b is a solution.

Definition 3.2 *Number of solutions of a polynomial congruence*

Let $S = \{b_1, b_2, \ldots, b_n\}$ be a complete set of residues modulo n. The *number of solutions* of the congruence $P(x) \equiv 0 \pmod{n}$ is the number of integers $b \in S$ for which $P(b) \equiv 0 \pmod{n}$.

Theorem 3.4 *Solution of linear congruences*

Consider the linear congruence $ax \equiv b \pmod{n}$.

(a) The congruence has solutions if, and only if, $\text{hcf}(a, n)$ divides b.

(b) If $\text{hcf}(a, n) = 1$, the congruence has a unique solution.

(c) If $\text{hcf}(a, n) = d$ and d divides b, then the congruence has d solutions, which are given by the unique solution modulo $\frac{n}{d}$ of the congruence $\frac{a}{d}x \equiv \frac{b}{d} \pmod{\frac{n}{d}}$.

Strategy: to solve the linear congruence $ax \equiv b \pmod{n}$

1. Check that $\text{hcf}(a, n)$ divides b. If it does not, the congruence has no solutions. If it does:
2. Cancel any common factors of all three of a, b and n. The resulting congruence has a unique solution modulo the new modulus.

   The resulting coefficients (originally a and b) can then be changed by applying the remaining steps in any order, any number of times, with the goal of reaching a congruence in which the coefficient of x is 1.
3. Cancel any common factor of the coefficients.
4. Replace either coefficient by any congruent number.
5. Multiply through the congruence by any number that is coprime to the modulus.

Simultaneous linear congruences

Theorem 4.2 *Chinese Remainder Theorem*

Let $n_1, n_2, \ldots, n_r$ be natural numbers such that $\text{hcf}(n_i, n_j) = 1$ for $i \neq j$. Then the system of linear congruences

$$x \equiv b_1 \pmod{n_1};\ x \equiv b_2 \pmod{n_2};\ \ldots\ ;\ x \equiv b_r \pmod{n_r}$$

has a simultaneous solution that is unique modulo $n_1 n_2 \cdots n_r$.

Theorem 4.5

The system of linear congruences

$$x \equiv b_1 \pmod{n_1};\ x \equiv b_2 \pmod{n_2}$$

has a simultaneous solution if, and only if, $\text{hcf}(n_1, n_2)$ divides $b_2 - b_1$. If a solution exists then it is unique modulo $\text{lcm}(n_1, n_2)$.

Corollary 4.6

The system of linear congruences

$$x \equiv b_1 \,(\mathrm{mod}\ n_1);\ x \equiv b_2 \,(\mathrm{mod}\ n_2);\ \ldots\ ;\ x \equiv b_r \,(\mathrm{mod}\ n_r)$$

has a simultaneous solution if, and only if, $\mathrm{hcf}(n_i, n_j)$ divides $b_j - b_i$ for each pair i, j of suffixes. If a solution exists it is unique modulo $\mathrm{lcm}(n_1, n_2, \ldots, n_r)$.

Chapter 4: Fermat's and Wilson's Theorems

Fermat's Little Theorem

Theorem 1.1 *Fermat's Little Theorem (FLT)*

If p is a prime and a is any integer with $\mathrm{hcf}(a, p) = 1$, then $a^{p-1} \equiv 1 \,(\mathrm{mod}\ p)$.

Fermat's Little Theorem, an alternative formulation

If p is a prime and a is any integer, $a^p \equiv a \,(\mathrm{mod}\ p)$.

Representation of fractions by decimals

Proposition 2.1 *Terminating decimals*

The decimal of $\frac{1}{n}$ terminates if, and only if, $n = 2^r 5^s$, for some integers $r \geq 0$ and $s \geq 0$.

Definition 2.2 *The order of an integer*

If p is prime and $\mathrm{hcf}(a, p) = 1$, then the *order* of a modulo p is the least positive integer c such that $a^c \equiv 1 \,(\mathrm{mod}\ p)$.

See Book C, Chapter 9, Definition 3.4 for the order of an integer modulo n.

Proposition 2.3 *The cycle length of $\frac{1}{p}$*

The length of the cycle in the decimal of $\frac{1}{p}$ is equal to the order of 10 modulo p, where p is a prime other than 2 or 5.

Proposition 2.4

If a has order c modulo p, where $p \geq 3$ is prime, and $a^k \equiv 1 \,(\mathrm{mod}\ p)$ then c is a factor of k. In particular, c divides $p - 1$.

Wilson's Theorem

Theorem 3.1 *Wilson's Theorem*

If p is prime then $(p-1)! \equiv -1 \,(\mathrm{mod}\ p)$.

Proposition 3.2 *Converse of Wilson's Theorem*

If $n > 1$ is an integer and $(n-1)! \equiv -1 \,(\mathrm{mod}\ n)$ then n is prime.

Polynomial congruences

Definition 4.1 *Polynomial congruences and their solutions*

A polynomial congruence is an expression

$$P(x) = c_r x^r + c_{r-1}x^{r-1} + \cdots + c_1 x + c_0 \equiv 0 \,(\mathrm{mod}\ n),$$

where $P(x)$ is a polynomial of degree $r \geq 0$ with integer coefficients.

An integer a is a solution of the polynomial congruence $P(x) \equiv 0 \,(\mathrm{mod}\ n)$ if, and only if, $P(a) \equiv 0 \,(\mathrm{mod}\ n)$.

The number of solutions of a polynomial congruence is the number of incongruent solutions modulo n.

Definition 4.2 *Degree of a polynomial congruence*

If $P(x) = c_r x^r + c_{r-1}x^{r-1} + \cdots + c_1 x + c_0$ then the polynomial $P(x)$ has degree k modulo n if, after removing each term in which the coefficient $c_i \equiv 0 \,(\mathrm{mod}\ n)$, the leading term remaining is $c_k x^k$.

If every coefficient $c_i \equiv 0 \,(\mathrm{mod}\ n)$ then the polynomial is not assigned a degree modulo n.

The degree of the polynomial congruence $P(x) \equiv 0 \,(\mathrm{mod}\ n)$ is the degree of $P(x)$ modulo n.

Theorem 4.3 *Lagrange's Theorem (Numbers)*

Let p be a prime. A polynomial congruence $P(x) \equiv 0 \,(\mathrm{mod}\ p)$ of degree $k \geq 1$ can have at most k solutions.

Lemma 4.4

Let b be a solution of the polynomial congruence $P(x) \equiv 0 \,(\mathrm{mod}\ p)$ of degree $k \geq 1$. Then $P(x) \equiv (x-b)P_1(x) \,(\mathrm{mod}\ p)$, where $P_1(x)$ is a polynomial of degree $k-1$ modulo p.

Proposition 4.5 *Factorising a polynomial modulo p*

Let p be prime and let $b_1, b_2, \ldots, b_r$ be incongruent solutions of the polynomial congruence $P(x) \equiv 0 \,(\mathrm{mod}\ p)$ of degree $k \geq r$. Then

$$P(x) \equiv (x-b_1)(x-b_2)\cdots(x-b_r)P_r(x) \,(\mathrm{mod}\ p),$$

where the polynomial $P_r(x)$ has degree $k-r$ modulo p.

Corollary 4.7 *to Proposition 4.5*

For any prime p,
$x^{p-1} - 1 \equiv (x-1)(x-2)(x-3)\cdots(x-(p-1)) \,(\mathrm{mod}\ p)$.

Proposition 4.8

If p is prime and d is a factor of $p-1$ then the congruence $x^d - 1 \equiv 0 \,(\mathrm{mod}\ p)$ has exactly d solutions.

Book B: Groups

Glossary

AB The set $\{ab : a \in A,\ b \in B\}$ where A and B are subsets of a group G. **Chapter 5**, Section 2

abelian group A group whose binary operation is commutative; that is, $a \circ b = b \circ a$ for all a, b in the group. **Chapter 5**, Section 1

alternating group A_n The normal subgroup of S_n consisting of all the even permutations. **Chapter 5**, Section 1

automorphism An isomorphism from a group to itself. **Chapter 8**, Section 4

Cartesian product of X and Y The set $X \times Y$ of all ordered pairs (x, y) with $x \in X$ and $y \in Y$. **Chapter 6**, Section 1

Cayley table A table describing the result of applying the group operation to every pair of elements of the group. **Chapter 5**, Section 1

cycle form A permutation written as a composite (product) of disjoint cycles. **Chapter 5**, Section 1

cyclic group $\mathbb{Z}_n$ The group consisting of the set $\{0, \ldots, n-1\}$ with addition modulo n. This is isomorphic to the (standard) cyclic group of order n, which is the quotient group $\mathbb{Z}/n\mathbb{Z}$. **Chapter 5**, Section 1

dicyclic group Dic_n The dicyclic group of order $4n$. **Chapter 7**, Section 1

dihedral group D_n The symmetry group of a regular n-gon; a group of order $2n$. **Chapter 7**, Section 1

even permutation A permutation that can be expressed as the composite of an even number of transpositions. **Chapter 5**, Section 1

G/N The set of left cosets of a normal subgroup N in G. **Chapter 5**, Section 3

general linear group $\mathrm{GL}(2, \mathbb{R})$ The group of 2×2 matrices over $\mathbb{R}$ with non-zero determinants. **Chapter 5**, Section 1

generator One of the group elements g_i where $\langle g_1, \ldots, g_n \mid r_1, \ldots, r_m \rangle$ is a group presentation. **Chapter 7**, Section 1

Klein group V The non-cyclic group of order 4, also denoted by (and isomorphic to) $\mathbb{Z}_2 \times \mathbb{Z}_2$, $\Gamma(\square)$ and D_2. **Chapter 5**, Section 4

odd permutation A permutation that can be expressed as the composite of an odd number of transpositions. **Chapter 5**, Section 1

orthogonal group $\mathrm{O}(2, \mathbb{R})$ The group of all symmetries of the plane $\mathbb{R}^2$; the elements can be represented by certain 2×2 matrices with determinant ± 1. **Chapter 5**, Section 1

permutation A bijective function from the set $\{1, 2, \ldots, n\}$ to itself. **Chapter 5**, Section 1

permutation group A group of permutations with composition of functions as the operation. **Chapter 5**, Section 1

Chapter 5, Section 2 **proper subgroup** A subgroup that is not the whole group.

Chapter 7, Section 1 **quaternion group** The dicyclic group of order 8, Dic_2, also with presentation $\langle \pm 1, \pm i, \pm j, \pm k \mid i^2 = j^2 = k^2 = -1,\ ij = k,\ jk = i,\ ki = j \rangle$.

Chapter 5, Section 1 **relations** Equalities that hold among elements of a group.

Chapter 5, Section 1 **special orthogonal group** $\mathrm{SO}(2, \mathbb{R})$ The group of all rotations of the plane $\mathbb{R}^2$; the elements can be represented by certain 2×2 matrices with determinant 1.

Chapter 5, Section 1 **symmetric group** S_n The group of all permutations of n symbols.

Chapter 5, Section 1 **symmetry group of a rectangle** $\Gamma(\square)$ The group of symmetries of a rectangle, isomorphic to the Klein group V.

Chapter 5, Section 1 **transposition** A permutation cycle of length 2.

Chapter 5, Section 2 **trivial subgroup** The subgroup comprising just the identity.

Chapter 5: Examples of groups

Group axioms

Definition 1.1 *Group axioms*

A set G together with a binary operation $\circ$ defined on G is a **group** if the following axioms are satisfied.

G1 Closure For all pairs x, y of elements in G, $x \circ y$ is also in G.

G2 Identity There is an **identity element** – e, say – in G with the property that, for any element x in G, $e \circ x = x = x \circ e$.

G3 Inverses For each element x in G, there is an **inverse element** x^{-1} in G with the property that $x \circ x^{-1} = e = x^{-1} \circ x$, where e is the identity element of G.

G4 Associativity For all choices of elements x, y, z in G, we have $(x \circ y) \circ z = x \circ (y \circ z)$.

Proposition 1.2 *Elementary consequences of the group axioms*

Let G be a group.

(a) **Uniqueness of identity** The identity element of G is unique.

(b) **Uniqueness of inverses** Let a be an element of G. Then the inverse of a is unique.

(c) **Left cancellation rule** For any elements x, y and a of G, $ax = ay$ implies $x = y$.

(d) **Right cancellation rule** For any elements x, y and a of G, $xa = ya$ implies $x = y$.

(e) **Inverse of a product** For any elements x and y of G, the inverse of the product xy is $y^{-1}x^{-1}$.

Propositions 1.4–1.6

Let G be a group with identity e and let x be an element of G.

- If g is an element of G such that either $gx = x$ or $xg = x$, then $g = e$.
- If y is an element of G such that either $yx = e$ or $xy = e$, then $y = x^{-1}$.

Proposition 1.7 *Generalised associativity rule*

Let G be a group and n be a positive integer. Then the product of n elements of G, *taken in a fixed order*, is the same no matter which way the product is calculated.

Definition 1.8 *Cyclic group*

A group G is **cyclic** if every element has the form g^n for some element g in G. Such an element g is called a **generator** of the group.

Subgroups and cosets

Definition 2.1 *Subgroup*

A subset H of the underlying set of a group G is a **subgroup** of G if it is itself a group under the group operation of G.

We use the notation $H \leq G$ to denote that H is a subgroup of G.

Proposition 2.2 *Subgroup conditions*

If G is a group with binary operation $\circ$ and H is a subset of the underlying set of G, then H is a subgroup of G (with binary operation $\circ$) if all the following conditions are satisfied.

SG1 Closure For all pairs x, y of elements in H, $x \circ y$ is also in H.

SG2 Identity The identity element e of G is in H.

SG3 Inverses For each element x in H, the inverse element x^{-1} in G is also in H.

Conditions SG1, SG2 and SG3 are the *subgroup axioms*.

Proposition 2.3

Let G be a group. Then:

(a) if e is the identity, then $\{e\}$ is a subgroup of G

(b) G is a subgroup of G.

Proposition 2.4 *Subgroup criterion*

Let G be a group and H be a subset of the underlying set of G. Then H is a subgroup of G if, and only if, it satisfies the following two properties:

(a) H is non-empty

(b) $x, y \in H$ implies $x^{-1}y \in H$.

Proposition 2.5 *Subgroup criteria (additive notation)*

1. A subset H of a group G, written using additive notation, is a subgroup if, and only if, it satisfies the following conditions:
   (a) H is non-empty; (b) $x, y \in H$ implies $-x + y \in H$.
2. A subset H of an additive group G is a subgroup if, and only if, it satisfies the following conditions:
   (a) H is non-empty; (b) $x, y \in H$ implies $x - y \in H$.

Propositions 2.6 and 2.7

Let G be a group. The intersection of any non-empty collection of subgroups of G is a subgroup both of G and of each of the subgroups in the collection.

Definition 2.8 *Left coset*

Let G be a group, H be a subgroup of G and a be an element of G. Then the subset $aH = \{ah : h \in H\}$ is called the **left coset** of H by a.

Proposition 2.10

Let G be a group and H be a subgroup of G. Then $aH = H$ if, and only if, a is an element of H.

Proposition 2.11

Let G be a group, H be a subgroup of G and a and b be elements of G. Then the left cosets aH and bH are equal if, and only if, $a^{-1}b \in H$.

Proposition 2.12

Let G be a group, H be a subgroup of G and a and b be elements of G. Then the left cosets aH and bH are either disjoint or equal.

Proposition 2.13

Let G be a group, H be a subgroup of G and a and b be elements of G. Then the following four statements are equivalent:
(a) $a^{-1}b \in H$; (c) a and b are in the same left coset;
(b) $aH = bH$; (d) $b \in aH$.

Definition 2.14 *Finite and infinite groups*

A group is **finite** if its underlying set is finite, that is, it has a finite number of elements. If the underlying set is infinite, the group is an **infinite group**.

Definition 2.15 *Order of a group*

The **order** of a finite group G is the number of elements it contains, and it is denoted by $|G|$. An infinite group is said to have **infinite order**.

Theorem 2.16 *Lagrange's Theorem (Groups)*

Let G be a finite group and H be a subgroup of G. Then the order of H divides the order of G, that is, $|H|$ divides $|G|$.

Normal subgroups and quotient groups

Propositions 3.1 and 3.2

Let G be a group, H be a subgroup of G and a and b be elements of G.

- The right cosets Ha and Hb are equal if, and only if, $ab^{-1} \in H$. (The symmetry in a and b implies that an equivalent condition is $ba^{-1} \in H$.)
- The right cosets Ha and Hb are either disjoint or equal.

Definition 3.4 *Index of a subgroup*

Let G be a group, and H be a subgroup with finitely many cosets. Then the **index** of H in G is the number of left (or right) cosets of H in G.

Proposition 3.5

Let G be a finite group and let H be a subgroup of G. Then the index of H in G is $\frac{|G|}{|H|}$.

Definition 3.6 *Normal subgroup*

Let H be a subgroup of a group G. Then H is a **normal subgroup** of G if $aH = Ha$ for every $a \in G$.

Proposition 3.7

Let H be a subgroup of the group G. Then:

(a) H is a normal subgroup if, and only if, $aHa^{-1} \subseteq H$ for each a in G, where $aHa^{-1} = \{aha^{-1} : h \in H\}$.

(b) H is a normal subgroup if, and only if, the set of left cosets of H is equal to the set of right cosets of H.

Theorem 3.8

Let G be a group and S be a subset of (the underlying set of) G. Then there exists a normal subgroup N of G that contains S, and such that for any other normal subgroup M of G containing S, $N \subseteq M$.

Definition 3.9 *Conjugate elements and subgroups*

Let G be a group and let a and b be elements of G.

(a) The elements a and b are **conjugate elements** in G if there exists $g \in G$ such that $gag^{-1} = b$.

(b) The **conjugacy class** of a in G is the set of elements conjugate to a, that is, the set $\mathrm{Conj}_G(a) = \{gag^{-1} : g \in G\}$.

(c) Two subgroups H and K in G are **conjugate subgroups** in G if there exists $g \in G$ such that $gHg^{-1} = K$.

Proposition 3.10

Let G be a group. Then a subgroup N is normal in G if, and only if, N has no conjugate subgroups apart from N.

Proposition 3.11

Let G be a group.

(a) The relation $\sim$ on elements of G defined by $a \sim b$ if, and only if, a is conjugate to b is an equivalence relation.

(b) The conjugacy classes of G partition the group.

Proposition 3.12

Let G be a group.

(a) Any normal subgroup N of G is a union of conjugacy classes.

(b) Any subgroup H of G that is a union of conjugacy classes is a normal subgroup of G.

Definition 3.13 *Product of cosets*

Let G be a group, N be a normal subgroup of G and let $a, b \in G$. We define the **product** of the cosets aN and bN to be $(aN)(bN) = (ab)N$.

Theorem 3.14

Let G be a group, N be a normal subgroup of G and a, a_1, b and b_1 be elements of G such that $aN = a_1N$ and $bN = b_1N$. Then $(ab)N = (a_1b_1)N$.

Definition 3.15 *Quotient group*

Let N be a normal subgroup of the group G. The **quotient group of G with respect to N** is defined as the group whose underlying set is G/N and whose operation is given by $(aN)(bN) = (ab)N$ for any $aN, bN \in G/N$.

Isomorphisms and homomorphisms

Definition 4.1 *Isomorphism of groups*

The groups $(G, \circ)$ and $(H, *)$ are **isomorphic** if, and only if, there exists a bijection $\phi : G \to H$ such that, for all a and b in G, we have $\phi(a \circ b) = \phi(a) * \phi(b)$. Such a bijection is called an **isomorphism** from G to H.

We use the notation $G \cong H$ to denote that G is isomorphic to H.

Proposition 4.2

Let G and H be groups and let ϕ be an isomorphism from G to H. Then ϕ^{-1} is an isomorphism from H to G.

Theorem 4.3

Let G and H be groups, let e be the identity of G and let ϕ be an isomorphism from G to H. Then $\phi(e)$ is the identity of H.

Definition 4.4 *Order of a group element*

Let a be an element of a group G and let e be the identity in G. Then a has **order** n if, and only if, n is the smallest positive integer such that

$a^n = e$. If no such positive integer exists, then the element a is said to have **infinite order**.

We use the notation $|a|$ to denote the order of an element a of finite order.

Proposition 4.5

Let G be a finite group of order n and let $g \in G$. Then the order of g divides n.

Theorem 4.6

Let G and H be groups and let ϕ be an isomorphism from G to H. Then if $a \in G$ has finite order n, so does $\phi(a) \in H$.

Definition 4.7 *Homomorphism*

Let $(G, \circ)$ and $(H, *)$ be groups. A function $\phi : G \to H$ is a **homomorphism** from G to H if, for all a and b in G, we have $\phi(a \circ b) = \phi(a) * \phi(b)$.

This property is referred to as the **morphism** or **homomorphism property**.

Definitions 4.9 and 4.10 *Kernel and image*

Let $\phi : G \to H$ be a homomorphism and let e be the identity in H. Then the **kernel** of ϕ is the set $\mathrm{Ker}(\phi) = \{x \in G : \ \phi(x) = e\}$, and the **image** of ϕ is the set $\mathrm{Im}(\phi) = \{h \in H : \ h = \phi(x) \text{ for some } x \in G\}$.

Theorem 4.11

Let $\phi : G \to H$ be a homomorphism from G to H. Then:

(a) The kernel of ϕ is a normal subgroup of G.

(b) The image of ϕ is a subgroup of H.

Theorem 4.13 *First Isomorphism Theorem*

Let $\phi : G \to H$ be a homomorphism from the group G to the group H, with kernel K and image I. Then the function $\psi : G/K \to I$ defined by $\psi : gK \mapsto \phi(g)$ is an isomorphism.

Corollary 4.14

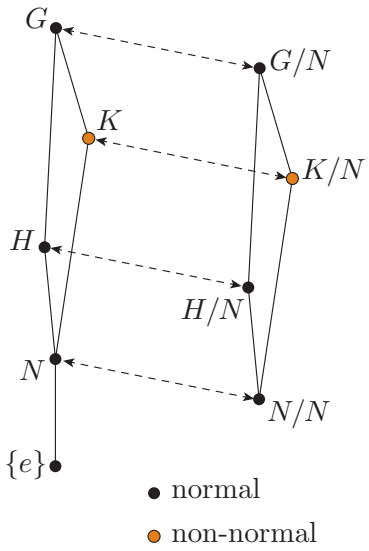
Suppose that $\phi : G \to H$ is a homomorphism from G to H such that $\mathrm{Ker}(\phi) = \{e\}$. Then ϕ is an isomorphism from G to $\phi(G) = \mathrm{Im}(\phi)$.

Theorem 4.15

Let N be a normal subgroup of a group G. Then the function ϕ defined by $\phi(a) = aN$ is a homomorphism from G onto the quotient group G/N. Furthermore, the kernel of ϕ is N.

Definition 4.16 *Natural homomorphism*

Let N be a normal subgroup of a group G. Then the homomorphism $\phi : G \to G/N$ defined by $\phi : a \mapsto aN$ is called the **natural homomorphism** from G onto G/N.



Theorem 4.17 *Correspondence Theorem*

Let G be a group and let N be a normal subgroup of G. Then there is a one–one correspondence between the subgroups of G that contain N and the subgroups of the quotient group G/N. Moreover, this correspondence maps normal subgroups of G to normal subgroups of G/N and vice versa.

Corollary 4.18

Let G be a group, N be a normal subgroup of G and H be a subgroup of G that contains N. If G is finite, the index of H in G is equal to the index of H/N in G/N; that is,

$$\frac{|G|}{|H|} = \frac{|G/N|}{|H/N|}.$$

Chapter 6: Towards classification

Direct products

Definition 1.2 *Direct product of two groups*

Let $(G, \circ)$ and $(H, *)$ be two groups. Their **direct product** $(G \times H, \bullet)$ is the group defined as follows.

Set The underlying set of the group is $G \times H$, the Cartesian product of the underlying sets G and H; that is, $G \times H = \{(g, h) : g \in G, h \in H\}$.

Operation If (g_1, h_1) and (g_2, h_2) are elements of $G \times H$ then $(g_1, h_1) \bullet (g_2, h_2) = (g_1 \circ g_2, h_1 * h_2)$.

Definition 1.4 *Internal direct product*

Let G be a group and let H_1 and H_2 be subgroups of G. Then G is the **internal direct product** of H_1 and H_2 if the map

$$\begin{aligned} \phi &: H_1 \times H_2 \to G \\ \phi &: (h_1, h_2) \mapsto h_1 h_2 \end{aligned}$$

is an isomorphism.

Theorem 1.5 *Internal Direct Product Theorem*

If H_1 and H_2 are subgroups of a group G, then

$$\begin{aligned} \phi &: H_1 \times H_2 \to G \\ \phi &: (h_1, h_2) \mapsto h_1 h_2 \end{aligned}$$

is an isomorphism if, and only if, all three of the following conditions hold:

(a) $G = H_1 H_2$

(b) $H_1 \cap H_2 = \{e\}$

(c) H_1 and H_2 are normal subgroups of G.

In particular, when these three conditions are satisfied, G is the internal direct product of H_1 and H_2.

Theorem 1.6

If A, B and C are groups, then:

(a) $A \times B \cong B \times A$; (b) $A \times (B \times C) \cong (A \times B) \times C$.

Corollary 1.7

If H_i are subgroups of a group G for $i = 1, \ldots, n \in \mathbb{N}$, $n \geq 2$, then

$$\begin{array}{rccl} \phi\colon & H_1 \times H_2 \times \cdots \times H_n & \to & G \\ & \phi : (h_1, h_2, \ldots, h_n) & \mapsto & h_1 h_2 \cdots h_n \end{array}$$

is an isomorphism if, and only if, all three of the following conditions hold:

(a) $G = H_1 H_2 \cdots H_n$

(b) $H_i \cap H_1 H_2 \cdots H_{i-1} H_{i+1} \cdots H_n = \{e\}$ for $1 \leq i \leq n$

(c) $H_1, H_2, \ldots, H_n$ are normal subgroups of G.

In particular, when these three conditions are satisfied, G is the internal direct product of $H_1, H_2, \ldots, H_n$.

Proposition 1.9

Let G be a group with subgroups H and K such that $H \cap K = \{e\}$. Then:

(a) If H and K have finite orders r and s respectively, then HK has rs distinct elements. In particular, if $rs = |G|$ then $G = HK$.

(b) If $kh = hk$ for all $h \in H$ and $k \in K$, then HK is a subgroup of G with $HK \cong H \times K$.

Proposition 1.10

Let H, K be subgroups of G, and H_1, H_2 be subgroups of H. Suppose that $G \cong H \times K$ and $H \cong H_1 \times H_2$. Then $G \cong H_1 \times H_2 \times K$.

Cyclic groups

Theorem 2.1

(a) All infinite cyclic groups are isomorphic to $(\mathbb{Z}, +)$.

(b) A finite cyclic group of order n is isomorphic to the quotient group $\mathbb{Z}/n\mathbb{Z}$, which is $(\mathbb{Z}_n, +_n)$.

Theorem 2.2 *Quotients of cyclic groups*

(a) Let G be a cyclic group and let $\phi : G \to H$ be a homomorphism. Then $\phi(G)$ is cyclic.

(b) Let G be a cyclic group generated by a and let H be a subgroup of G. Then the quotient group G/H is cyclic and generated by the coset aH.

Theorem 2.3

Let G be a cyclic group and let H be a subgroup of G. Then H is cyclic.

Corollary 2.4

The subgroups of $\mathbb{Z}$ are all of the form $n\mathbb{Z}$ for $n \in \mathbb{Z}$, $n \geq 0$.

Theorem 2.5 *Subgroups of cyclic groups*

Let $G = \langle a \rangle$ be a finite cyclic group of order n, so that a has order n. If q is a factor of n with $n = mq$, then G has a unique subgroup of order q that is generated by the element a^m.

Lemma 2.7

Let a and b be elements of finite orders m and n respectively in a group G. Suppose that $ba = ab$ and that $\langle a \rangle \cap \langle b \rangle = \{e\}$ (that is, the only element of G that can be written both as a power of a and as a power of b is e). Then the order of ab is the least common multiple of m and n.

Proposition 2.8

Let G, A, B be groups with $G = A \times B$. Suppose that $a \in A$ has order m and $b \in B$ has order n. Then the order of (a, b) is the least common multiple of m and n.

Proposition 2.9

If $g \in G$ has order n, then the order of g^m is $\frac{l}{m} = \frac{n}{h}$, where l is the least common multiple of m and n and h is the highest common factor of m and n. In particular, if m is a factor of n then the order of g^m is $\frac{n}{m}$.

Proposition 2.10

Let $G = \mathbb{Z}_m \times \mathbb{Z}_n$. Then no element of G can have order greater than the least common multiple of m and n.

Theorem 2.11 *Direct products of cyclic groups*

The direct product $\mathbb{Z}_m \times \mathbb{Z}_n$ of the cyclic groups $\mathbb{Z}_m$ and $\mathbb{Z}_n$ is cyclic if, and only if, m and n are coprime positive integers.

Corollary 2.12

Let m and n be coprime positive integers. Then $\mathbb{Z}_{mn} \cong \mathbb{Z}_m \times \mathbb{Z}_n$.

Theorem 2.13 *Decomposition of finite cyclic groups*

If n is a positive integer with prime decomposition $n = p_1^{k_1} \cdots p_r^{k_r}$, where $p_1 < \cdots < p_r$ are distinct primes and $k_1, \ldots, k_r$ are positive integers, then $\mathbb{Z}_n \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_r}$, where $n_i = p_i^{k_i}$, $i = 1, \ldots, r$.

Group actions

Definition 3.1 *Group action*

Let G be a group and X a non-empty set. A **group action** of G on X is a function $G \times X \to X$
$(g, x) \mapsto g \wedge x$
that satisfies the following conditions:

(a) $e \wedge x = x$, for all $x \in X$, where e is the identity element of G

(b) $(gh) \wedge x = g \wedge (h \wedge x)$, for all g, $h \in G$ and $x \in X$.

We also say that G acts on X by $\wedge$.

Proposition 3.5

In the group D_6, let r denote a rotation through $\frac{\pi}{3}$ and s denote a reflection in an axis of symmetry, so that $r^6 = e$, $s^2 = e$ and $sr = r^5 s = r^{-1}s$. Then every element of D_6 can be written as $r^m s^n$ for $m = 0, \ldots, 5$ and $n = 0, 1$.

Proposition 3.6

Suppose that the action $\wedge$ by D_6 on the set $X = D_6$ is defined as follows: $g \wedge x = gxg^{-1}$ for all $g \in D_6$ and $x \in X$. Then D_6 acts on X by $\wedge$.

Definitions 3.7 and 3.9 *Orbit and stabiliser*

If a group G acts on a set X and x is an element of X, then:

- The **orbit** of x under G is the set $\mathrm{Orb}(x) = \{g \wedge x : \ g \in G\}$; that is, the set of elements of X obtained by acting on x with the elements of G.
- The **stabiliser** of x under G is the set $\mathrm{Stab}(x) = \{g : \ g \in G, \ g \wedge x = x\}$; that is, the set of elements of G that fix x.

Orbits are subsets of X and stabilisers are subgroups of G.

Proposition 3.8

Let G be a group acting on a set X. Then the relation $\sim$ on X defined by

$$x \sim y \quad \text{if, and only if, there is } g \in G \text{ such that } x = g \wedge y$$

is an equivalence relation.

Definition 3.10 $Send_x(y)$

Let G be a group acting on a set X and let $x \in X$. Then for each element y let $\mathrm{Send}_x(y) = \{g \in G : g \wedge x = y\}$. So $\mathrm{Send}_x(y)$ is the set of all elements of G that send x to y.

Lemma 3.11

Let G be a group acting on a set X, let $x \in X$ and $y \in \mathrm{Orb}(x)$. If $g \in G$ is such that $g \wedge x = y$, then $\mathrm{Send}_x(y) = g\mathrm{Stab}(x)$.

Theorem 3.12

Let G be a group acting on a set X, let $x \in X$ and let $H = \mathrm{Stab}(x)$. Let $C(x)$ be the set of left cosets of H in G. Then the function $\sigma_x : \mathrm{Orb}(x) \to C(x)$ defined by $\sigma_x : g \wedge x \mapsto gH$ is a bijection.

Corollary 3.13 *Orbit–Stabiliser Theorem*

Let G be a finite group acting on the set X. Then, for each $x \in X$, $|\mathrm{Orb}(x)| \times |\mathrm{Stab}(x)| = |G|$.

Theorem 3.15 *Cayley's Theorem*

Every group G is isomorphic to a subgroup of S_G, the group of permutations of G. In particular, if G is finite then G is isomorphic to a subgroup of S_n, where $n = |G|$.

Theorem 3.16 *Cauchy's Theorem*

Let G be a group of order n and let p be a prime divisor of n. Then G has an element of order p.

Theorem 3.17

Let G be a finite abelian group of order n, where $n = p_1 p_2 \cdots p_r$ for r distinct primes $p_1, p_2, \ldots, p_r$. Then G is cyclic and of order n. Moreover, $G \cong \mathbb{Z}_{p_1} \times \mathbb{Z}_{p_2} \times \cdots \times \mathbb{Z}_{p_r}$.

Lemma 3.18

Suppose that G is an abelian group of order p^n for some prime p and some positive integer n. Then $G \cong H \times K$, where H is a cyclic subgroup of G of order p^r, for some positive integer r.

Theorem 3.19

Let G be an abelian group of order p^n for some positive integer n and for some prime p. Then G is a direct product of cyclic subgroups.

Chapter 7: Finite groups

Describing groups

Properties of a group presentation

Let G be a group with presentation $G = \langle g_1, g_2, \ldots, g_n \mid r_1, r_2, \ldots, r_m \rangle$.

1. Any element of the group can be written in the form $g_{i_1}^{n_1} g_{i_2}^{n_2} \cdots g_{i_s}^{n_s}$, where $i_1, \ldots, i_s \in \{1, 2, \ldots, n\}$ and $n_i \in \mathbb{Z}$. In this expression there may be repetitions of a generator, and the powers can be both positive and negative.
2. Any equation that holds in the group can be derived from the given relations.

Definition 1.1 *Standard form of elements*

Let G be a group with presentation $G = \langle g_1, g_2, \ldots, g_n \mid r_1, r_2, \ldots, r_m \rangle$. Then elements of G have a **standard form** if every element of G can be written uniquely as a product of powers of the generators according to a specified pattern.

Dihedral group D_n

The dihedral group D_n has presentation

$$D_n = \langle r, s \mid r^n = s^2 = e,\ sr = r^{n-1}s \rangle.$$

Each element can be expressed in the standard form $r^i s^j$ for $i = 0, \ldots, n-1,\ j = 0, 1$ and, in particular, $sr^{-i} = r^i s$.

Dicyclic group Dic_n

The dicyclic group Dic_n has presentation

$$\mathrm{Dic}_n = \langle a, b \mid a^4 = e, a^2 = b^n,\ aba^3b = e \rangle.$$

Each element can be expressed in the standard form $a^r b^s$ for $r = 0, \ldots, 3,\ s = 0, \ldots, n-1$ and, in particular, $ba = a^3 b^{n-1} = ab^{-1}$.

Definition 1.3 *Subgroup generated by a set*

Let G be a group and S be a subset of (the underlying set of) G. We define the **subgroup H of G generated by** S to be a subgroup of G satisfying: (a) $S \subseteq H$; (b) if K is a subgroup of G and $S \subseteq K$ then $H \subseteq K$.

We denote this subgroup by $\langle S \rangle$.

Lemma 1.4

Let G be a group and S be a subset of (the underlying set of) G. Then the group generated by S exists and is unique.

Lemma 1.6

Let G be a group and let S be a non-empty finite subset of G. Then the set $H = \{s_1^{r_1} \cdots s_n^{r_n} : n \in \mathbb{N},\ s_i \in S,\ r_i \in \mathbb{Z}, i = 1, \ldots, n\}$ is a subgroup of G.

Proposition 1.7

Let G be a group, let S be a non-empty finite subset of G and let $H = \{s_1^{r_1} \cdots s_n^{r_n} : n \in \mathbb{N}, s_i \in S, r_i \in \mathbb{Z}, i = 1, \ldots, n\}$. Then the subgroup generated by S is the whole of H, that is, $\langle S \rangle = H$.

Conjugates, centralisers and centres

Proposition 2.1

Let G be a group and let $g \in G$. Then the function $\phi : G \to G$, which sends each element a to its conjugate by g, that is, $\phi : a \mapsto gag^{-1}$, is an isomorphism.

Theorem 2.2

Let G be a group. Then G acts on G by the rule $g \wedge a = gag^{-1}$ for all $a, g \in G$, and we say that G **acts on itself by conjugation**.

Definition 2.3 *Conjugacy class*

Let G be a group and $a \in G$. Then the orbit of a in the action of G on itself by conjugation is written $\mathrm{Conj}_G(a)$ and it is called the **conjugacy class** of a in G. Thus $\mathrm{Conj}_G(a) = \{gag^{-1} : g \in G\}$.

Proposition 2.5

Let G be a group. Then:

(a) for each element a of G, $a \in \mathrm{Conj}_G(a)$

(b) for $a, b \in G$, $\mathrm{Conj}_G(a) = \mathrm{Conj}_G(b)$ if, and only if, b is a conjugate of a

(c) $\mathrm{Conj}_G(e) = \{e\}$ for any group G

(d) all the elements of a conjugacy class have the same order

(e) if G is finite of order n, then the number of elements in a conjugacy class must divide n.

Theorem 2.6

Let G be a finite group and let $C_1, C_2, \ldots, C_k$ be its distinct conjugacy classes. Then $|G| = |C_1| + |C_2| + \cdots + |C_k|$. This equation is usually called the **class equation** for G.

Definition 2.7 *Centraliser*

Let G be a group and $a \in G$. Then the stabiliser of a in the action of G on itself by conjugation is called the **centraliser** of a in G and it is written $\mathrm{Cent}_G(a)$. Thus $\mathrm{Cent}_G(a) = \{g \in G : gag^{-1} = a\}$.

Definition 2.8 *Central elements*

Let G be a group, and let $a \in G$ be an element that commutes with every other element of G. Then a is a **central element** in G. The set of all central elements in G is called the **centre** of G and it is denoted by $Z(G)$. Thus $Z(G) = \{a \in G : ga = ag \text{ for all } g \in G\}$.

Proposition 2.9

Let G be a group. Then the centre $Z(G)$ is a normal subgroup of G.

Proposition 2.10

Let G be a group and let $a \in G$. The following are equivalent ways of saying that a is central in G.

(a) $ga = ag$ for all $g \in G$
(b) $\mathrm{Cent}_G(a) = G$
(c) $a \in \mathrm{Cent}_G(g)$ for all $g \in G$
(d) $gag^{-1} = a$ for all $g \in G$
(e) $\mathrm{Conj}_G(a) = \{a\}$

Theorem 2.11

For any element a of any finite group G, $|\mathrm{Conj}_G(a)| \times |\mathrm{Cent}_G(a)| = |G|$.

Theorem 2.13

Let G be a group of prime power order, that is, $|G| = p^n$ for some prime number p and positive integer n. Then G has a non-trivial centre.

Theorem 2.14

If G is a group such that $G/Z(G)$ is cyclic, then G is abelian.

Theorem 2.15

If p is a prime number, then every group G of order p^2 is abelian.

Groups of small order

Proposition 3.1

Let p be a prime number and G be a group of order p. Then $G \cong \mathbb{Z}_p$.

Theorem 3.2

Let p be a prime number. Then every group of order p^2 is isomorphic to either $\mathbb{Z}_{p^2}$ or $\mathbb{Z}_p \times \mathbb{Z}_p$.

Lemma 3.3

Let G be a group of even order $2n$. Suppose that G has an element a of order n and an element b of order 2 with $b \notin \langle a \rangle$. Then:

(a) $bab^{-1} = a^k$ for some integer k such that $0 < k < n$ and n is a factor of $k^2 - 1$
(b) if $n \geq 3$ and $bab^{-1} = a^{n-1}$, then $G \cong D_n$.

Theorem 3.4

If p is an odd prime, then every group of order $2p$ is isomorphic to either $\mathbb{Z}_{2p}$ or D_p.

Proposition 3.5

Let G be a group of order 8 that is not isomorphic to any of the groups $\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$ and D_4. Then G has exactly one element d of order 2 and six elements $a^{\pm 1}$, $b^{\pm 1}$ and $c^{\pm 1}$ of order 4. Moreover, $G = \{e, a, a^2, a^3, b, ab, a^2b, a^3b\}$ and $ba = a^3b$.

Theorem 3.6

Every group G of order 8 is isomorphic to one of the five groups

$$\mathbb{Z}_8,\ \mathbb{Z}_4 \times \mathbb{Z}_2,\ \mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2,\ D_4,\ \mathrm{Dic}_2.$$

Proposition 3.7

Let G be an abelian group of order 12. Then either $G \cong \mathbb{Z}_6 \times \mathbb{Z}_2$ or $G \cong \mathbb{Z}_{12}$.

Proposition 3.8

Let G be a non-abelian group of order 12, and let a and b be elements of G such that a has order 6, b has order 2 and $b \notin \langle a \rangle$. Then $G \cong D_6$.

Proposition 3.9

Let G be a group of order 12. Then:

(a) If G has a subgroup H of order 3 that is not normal, then $G \cong A_4$.

(b) If G has no element of order 6, then $G \cong A_4$.

Proposition 3.10

Let G be a group of order 12 that is not isomorphic to any of the groups $\mathbb{Z}_{12}$, $\mathbb{Z}_6 \times \mathbb{Z}_2$, D_6, A_4. Then G is not abelian, and it has an element b of order 6 such that every element of order 2 in G is in $\langle b \rangle$. Moreover, G has two elements b and b^{-1} of order 6, two elements of order 3, one element $d = b^3$ of order 2, and six elements of order 4.

Theorem 3.11

Every group G of order 12 is isomorphic to one of the five groups $\mathbb{Z}_{12}$, $\mathbb{Z}_6 \times \mathbb{Z}_2$, A_4, D_6, Dic_3.

Finite p-groups

Definition 4.1 *Finite p-group*

Let p be a prime and n a positive integer. Then a group of order p^n is called a **p-group**.

Lemma 4.2

Let G be an abelian p-group of order p^n for some positive integer n, with $n \geq 2$. Then G has a subgroup of order p^{n-1}.

Theorem 4.3

Let G be a p-group of order p^n for some positive integer n. Then there exist normal subgroups $\{e\} = H_0, H_1, \ldots, H_n = G$ of G such that $\{e\} = H_0 \subset H_1 \subset \cdots \subset H_{n-1} \subset H_n = G$, where $|H_i| = p^i$ for $i = 0, \ldots, n$.

Definition 4.4 *Simple group*

A group G with more than one element is a **simple group** if it has no non-trivial proper normal subgroups.

Definition 4.5 *Composition series*

Let G be a group and suppose that $\{e\} = H_0, H_1, \ldots, H_n = G$ are subgroups of G such that:

(a) $\{e\} = H_0 \subset H_1 \subset \cdots \subset H_{n-1} \subset H_n = G$

(b) H_i is a normal subgroup of H_{i+1} for $i = 1, \ldots, n-1$

(c) H_{i+1}/H_i is a simple group for $i = 0, \ldots, n-1$.

Then $H_0, H_1, \ldots, H_n$ are a **composition series** for G of **length** n.

Definition 4.7 *Soluble group*

Let G be a group and suppose that G has subgroups $\{e\} = H_0, H_1, \ldots, H_n = G$ such that

$$\{e\} = H_0 \subset H_1 \subset \cdots \subset H_{n-1} \subset H_n = G,$$

where H_i is a normal subgroup of H_{i+1} and H_{i+1}/H_i is a cyclic group of prime order for $i = 0, \ldots, n-1$. Then G is a **soluble group**.

Chapter 8: The Sylow Theorems

The Sylow Theorems

Lemma 1.4

Let G be a finite group and let X be the set of all the k-element subsets of G. If G acts by left multiplication on X and if $A \in X$, then:

(a) $|\mathrm{Stab}(A)|$ divides $|A|$.

(b) If k is a power of a prime p, then $\mathrm{Stab}(A)$ is a p-group.

(c) If k is the maximal power of p dividing the order of G and if p does not divide $|\mathrm{Orb}(A)|$, then $\mathrm{Stab}(A)a = A$ for any $a \in A$.

Definition 1.7 *Sylow p-subgroup*

Let G be a finite group of order n and let p be a prime dividing n. Let p^α be the highest power of p dividing n. Then a subgroup of G of order p^α is called a **Sylow p-subgroup** of G.

Lemma 1.8

Let G be a finite group of order $p^\alpha z$, where p is a prime not dividing z. Let X be the set of p^α-element subsets of G. Then, in the action of G on X by left multiplication, the size of every orbit is divisible by z.

Lemma 1.9

If G is a group that has only one subgroup of a particular order, then this subgroup is normal.

Theorem 1.12 *The Sylow Theorems*

Let G be a finite group of order $n = p^\alpha z$, where p is a prime not dividing z. Let m be the number of subgroups of G of order p^α. Then:

- **Sylow's First Theorem**
  G has a Sylow p-subgroup, that is, $m \geq 1$.
- **Sylow's Second Theorem**
  $m \equiv 1 \pmod{p}$, that is, m is of the form $1 + kp$ for some integer k.
- **Sylow's Third Theorem (a)**
  All the Sylow p-subgroups of G are conjugate.
- **Sylow's Third Theorem (b)**
  m divides $|G|$, that is, m divides z.

This result collects together all the Sylow Theorems from Book B: Theorems 1.1, 1.6, 1.10 and Corollary 1.11 (to Theorem 1.10). Note that Corollary 1.2 (to Theorem 1.1) is covered by Theorem 3.1.

Applications of the Sylow Theorems

Lemma 2.1

Let G be a finite abelian group with subgroups $H_1, H_2, \ldots, H_s$ such that $s \geq 2$ and $H_i \cap H_1 H_2 \cdots H_{i-1} H_{i+1} \cdots H_s = \{e\}$ for each i with $1 \leq i \leq s$. Then $H_1 \times H_2 \times \cdots \times H_s$ is isomorphic to a subgroup of G.

Theorem 2.2

Let G be a finite abelian group of order $n = p_1^{a_1} p_2^{a_2} \cdots p_s^{a_s}$, where $p_1, p_2, \ldots, p_s$ are distinct primes. Then $G \cong G_1 \times G_2 \times \cdots \times G_s$, where each G_i is an abelian group of order $p_i^{a_i}$ and hence a direct product of cyclic groups.

Theorem 2.3 *Internal direct product for finite groups*

If H_1 and H_2 are subgroups of a finite group G, then

$$\phi : H_1 \times H_2 \to G$$
$$\phi : (h_1, h_2) \mapsto h_1 h_2$$

is an isomorphism if all three of the following conditions hold:

(a) $|G| = |H_1| \times |H_2|$

(b) $|H_1|$ and $|H_2|$ are coprime

(c) H_1 and H_2 are normal subgroups of G.

Subgroups of prime power order

Theorem 3.1 *Prime Power Subgroups Theorem*

Let G be a finite group and p be a prime. If p^β divides $|G|$ then G has a subgroup of order p^β. If m is the number of distinct subgroups of G of order p^β, then $m \equiv 1 \pmod{p}$.

Appendix 1: Table of small groups

Order of group	Possible groups
1	$\mathbb{Z}_1 = \{e\}$
2	$\mathbb{Z}_2$
3	$\mathbb{Z}_3$
4	$\mathbb{Z}_4$, $\mathbb{Z}_2 \times \mathbb{Z}_2$
5	$\mathbb{Z}_5$
6	$\mathbb{Z}_6$, $D_3 (\cong S_3)$
7	$\mathbb{Z}_7$
8	$\mathbb{Z}_8$, $\mathbb{Z}_4 \times \mathbb{Z}_2$, $\mathbb{Z}_2 \times \mathbb{Z}_2 \times \mathbb{Z}_2$, D_4, Dic_2
9	$\mathbb{Z}_9$, $\mathbb{Z}_3 \times \mathbb{Z}_3$
10	$\mathbb{Z}_{10}$, D_5
11	$\mathbb{Z}_{11}$
12	$\mathbb{Z}_{12}$, $\mathbb{Z}_6 \times \mathbb{Z}_2$, A_4, D_6, Dic_3
13	$\mathbb{Z}_{13}$
14	$\mathbb{Z}_{14}$, D_7

Book C: Numbers and rings

Glossary

abundant integer An integer $n \geq 1$ with $\sigma(n) > 2n$. **Chapter 9**, Section 2

amicable pair A pair of natural numbers m and n that satisfies $\sigma(m) = \sigma(n) = m + n$. **Chapter 9**, Section 2

binomial coefficient The positive integer given by $\binom{n}{k} = \dfrac{n!}{k!(n-k)!}$. **Chapter 11**, Section 4

complex modulus The number $|a + bi| = \sqrt{a^2 + b^2}$. **Chapter 12**, Section 2

cyclotomic polynomial $\Phi_n(x)$ A polynomial defined by $\Phi_n(x) = x^{n-1} + x^{n-2} + \cdots + x + 1 = \dfrac{x^n - 1}{x - 1}$. The roots are all the nth roots of unity, except for 1 itself. **Chapter 11**, Section 4

deficient integer An integer $n \geq 1$ with $\sigma(n) < 2n$. **Chapter 9**, Section 2

discriminant The number $b^2 - 4ac$ of the quadratic equation $ax^2 + bx + c = 0$. **Chapter 10**, Section 1

exceptional prime The prime with an odd exponent in Euler's form for an odd perfect number. **Chapter 9**, Section 2

finite field A field containing only finitely many elements, such as $\mathbb{Z}_p$. **Chapter 11**, Section 1

Gaussian integers The ring $\mathbb{Z}[i] = \{a + ib : a, b \in \mathbb{Z}\}$ with $i = \sqrt{-1}$, and the usual addition and multiplication of complex numbers. **Chapter 11**, Section 1

int(x) The integer part of x, the largest integer that does not exceed x. **Chapter 10**, Section 3

method of infinite descent A method of proof that uses the fact that a strictly decreasing sequence of positive integers must terminate. **Chapter 12**, Section 1

number-theoretic function A function $f : \mathbb{N} \to \mathbb{Z}$. **Chapter 9**, Section 1

parity A pair of integers that are both odd or both even have the *same* parity; otherwise they have *opposite* parity. **Chapter 12**, Section 1

Pell's equation A Diophantine equation $a^2 - db^2 = 1$, where a and b are variables and d is an integer. **Chapter 12**, Section 2

Pythagorean equation A Diophantine equation $x^2 + y^2 = z^2$. **Chapter 12**, Section 1

Pythagorean triangle A right-angled triangle associated with a Pythagorean triple. **Chapter 12**, Section 1

quadratic congruence A congruence of the form $ax^2 + bx + c \equiv 0 \pmod{p}$, where p is an odd prime and $a \not\equiv 0 \pmod{p}$. **Chapter 10**, Section 1

repunit A number comprising n 1s, such as 1, 11, 111, 1111, ... **Chapter 9**, Section 3

square-free An integer that is not divisible by the square of any prime. **Chapter 9**, Section 1

$\mathbb{Z}[x]$ The (commutative) ring of polynomials in x with integer coefficients. **Chapter 11**, Section 4

Chapter 9: Multiplicative functions

Multiplicative functions

Definition 1.1 *The τ function*

For any integer $n \geq 1$, $\tau(n)$ is defined to be the number of distinct factors of n, including 1 and n.

Definition 1.2 *The σ function*

For any integer $n \geq 1$, $\sigma(n)$ is defined to be the sum of the distinct factors of n, including 1 and n.

Factor sum formulas for τ and σ

$$\tau(n) = \sum_{d|n} 1 \quad \text{and} \quad \sigma(n) = \sum_{d|n} d$$

Definition 1.3 *Multiplicative functions*

A number-theoretic function is **multiplicative** if, for any integer with prime decomposition $p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$,

$$f\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) = f\left(p_1^{k_1}\right) f\left(p_2^{k_2}\right) \cdots f\left(p_r^{k_r}\right).$$

Proposition 1.4 *Equivalent definition of multiplicative functions*

The function f is a multiplicative function if, and only if, for every pair, m and n, of coprime natural numbers, $f(mn) = f(m)f(n)$.

Proposition 1.5 *Generating new number-theoretic functions*

If f is a multiplicative function, then the number-theoretic function F defined by $F(n) = \sum_{d|n} f(d)$ is also multiplicative.

Corollary 1.6 *to Proposition 1.5*

The functions τ and σ are multiplicative.

Proposition 1.7 *The prime decomposition formula for σ*

$$\begin{aligned}\sigma\left(p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}\right) &= \left(1 + p_1 + p_1^2 + \cdots + p_1^{k_1}\right) \cdots \left(1 + p_r + p_r^2 + \cdots + p_r^{k_r}\right) \\ &= \prod_{1 \leq i \leq r} \frac{p_i^{k_i+1} - 1}{p_i - 1}\end{aligned}$$

Perfect numbers

Definition 2.1 *Perfect numbers*

A natural number n is **perfect** if it satisfies $\sigma(n) = 2n$.

Theorem 2.2 *Classification of even perfect numbers*

If k is a positive integer such that $2^k - 1$ is prime, then $n = 2^{k-1}\left(2^k - 1\right)$ is perfect. Furthermore, every even perfect number is of this form for some positive integer k.

Definition 2.3 *Mersenne numbers*

The numbers $M_p = 2^p - 1$, where p is prime, are known as **Mersenne numbers**. When M_p is prime it is referred to as a **Mersenne prime**.

Proposition 2.4

If p and $2p + 1$ are primes, where $p \equiv 3 \pmod 4$, then $2p + 1$ divides M_p.

Proposition 2.5

Any prime factor of M_p, where p is an odd prime, is of the form $2kp + 1$ for some positive integer k.

Proposition 2.6 *Euler's form for an odd perfect number*

If n is an odd perfect number then $n = p^k p_1^{2k_1} p_2^{2k_2} \cdots p_r^{2k_r}$, where the p_is are distinct primes and $p \equiv k \equiv 1 \pmod 4$. (The exceptional prime p is listed first but is not necessarily the smallest prime in the decomposition.)

Definition 2.7 *Fermat numbers*

The numbers $F_n = 2^{2^n} + 1$, where n is a non-negative integer, are known as **Fermat numbers**.

Euler's ϕ-function

Definition 3.1 *Euler's ϕ-function*

For each integer $n \geq 1$, define $\phi(n)$ to be the number of natural numbers not exceeding n that are coprime to n.

Definition 3.2 *Reduced set of residues*

A **reduced set of residues modulo** n is a set of integers $\{a_1, a_2, a_3, \ldots, a_{\phi(n)}\}$ each of which is coprime to n and no two of which are congruent modulo n.

The reduced set of residues in which each is a positive integer less than or equal to n is called the **reduced set of least positive residues modulo** n.

Theorem 3.3 *Euler's Theorem – a generalisation of FLT*

If n is a natural number and a is any integer with $\text{hcf}(a, n) = 1$, then $a^{\phi(n)} \equiv 1 \pmod n$.

Definition 3.4 *The order of an integer modulo n*

If $n \geq 1$ and $\text{hcf}(a, n) = 1$, then the **order** of a modulo n is the least positive integer c such that $a^c \equiv 1 \pmod n$.

Proposition 3.5

If the integer a has order c modulo n then $a^k \equiv 1 \pmod n$ if, and only if, k is a multiple of c. In particular, the order c divides $\phi(n)$.

Repunit properties

1. $R_n = \dfrac{10^n - 1}{9}$ 2. $R_{m+n} = R_m \times 10^n + R_n$

Properties of Euler's ϕ-function

Theorem 4.1 *Multiplicativity of ϕ*

The function ϕ is multiplicative.

Proposition 4.2 *Formula for $\phi(n)$*

$\phi(1) = 1$.

If $n > 1$ has prime decomposition $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$ then

$$\begin{aligned}\phi(n) &= p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r} \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_2}\right) \cdots \left(1 - \frac{1}{p_r}\right) \\ &= n \prod_{\substack{p \mid n \\ p \text{ prime}}} \left(1 - \frac{1}{p}\right).\end{aligned}$$

There is one term in the product $\prod$ for each prime that divides n; that is, the product is taken over $p_1, p_2, \ldots, p_r$.

Alternative formula for $\phi(n)$

If $n = p_1^{k_1} p_2^{k_2} \cdots p_r^{k_r}$, then

$$\phi(n) = p_1^{k_1 - 1} p_2^{k_2 - 1} \cdots p_r^{k_r - 1} (p_1 - 1)(p_2 - 1) \cdots (p_r - 1).$$

In this formulation some (possibly all) of the exponents $k_i - 1$ may be 0.

Proposition 4.4

$$\sum_{d \mid n} \phi(d) = n$$

Primitive roots

Definition 5.1 *Primitive root*

If a has order $\phi(n)$ modulo n then a is a **primitive root of** n.

Proposition 5.2

If a is a primitive root of n then $\{a, a^2, a^3, \ldots, a^{\phi(n)}\}$ is a reduced set of residues modulo n.

Theorem 5.3

If a has order c modulo n then, for any $k \geq 1$, a^k has order

$$\frac{c}{\text{hcf}(c, k)} \quad \text{modulo } n.$$

Corollary 5.4

If a is a primitive root of n then a^k is also a primitive root of n if, and only if, k is coprime to $\phi(n)$.

Furthermore, if n has a primitive root, then it has exactly $\phi(\phi(n))$ primitive roots.

Theorem 5.6

Any integer that can be expressed as the product of two factors that are coprime and each exceeding 2 does not have a primitive root.

Corollary 5.7

If an integer n has a primitive root then it must have one of the following forms.

(a) $n = p^k$, p prime, $k \geq 1$

(b) $n = 2p^k$, p an odd prime, $k \geq 1$

Proposition 5.8

For any $k \geq 3$, 2^k does not have a primitive root.

Theorem 5.9 *Integers with primitive roots*

The integer n has a primitive root if, and only if, $n = 2$, 4, p^k, $2p^k$, for p an odd prime and $k \geq 1$.

Chapter 10: Quadratic reciprocity

Euler's Criterion

Solutions of quadratic congruences

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. The congruence $ax^2 + bx + c \equiv 0 \pmod{p}$ has solution(s) if, and only if, $b^2 - 4ac$ is congruent modulo p to a square.

Definition 1.2 *Quadratic residues*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. 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. We refer to whether a is a quadratic residue or quadratic non-residue of p as being the **quadratic character** of a modulo p.

Proposition 1.4 *The quadratic residues of p*

For any odd prime p there are $\frac{p-1}{2}$ quadratic residues and $\frac{p-1}{2}$ quadratic non-residues. The quadratic residues are congruent modulo p to the integers 1^2, 2^2, 3^2, $\ldots$, $\left(\frac{p-1}{2}\right)^2$.

Theorem 1.5 *Euler's Criterion*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. Then a is a quadratic residue of p if, and only if, $a^{(p-1)/2} \equiv 1 \pmod{p}$ and is a quadratic non-residue of p if, and only if, $a^{(p-1)/2} \equiv -1 \pmod{p}$.

The Legendre symbol

Definition 2.1 *Legendre symbol*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. The **Legendre symbol** (a/p) is defined as follows:

$$(a/p) = \begin{cases} 1, & \text{if } a \text{ is a quadratic residue of } p, \\ -1, & \text{if } a \text{ is a quadratic non-residue of } p. \end{cases}$$

Theorem 2.2 *Properties of the Legendre symbol*

Let p be an odd prime and let $a \not\equiv 0 \pmod{p}$ and $b \not\equiv 0 \pmod{p}$. Then the following properties hold.

(a) If $a \equiv b \pmod{p}$ then $(a/p) = (b/p)$.

(b) $(a^2/p) = 1$

(c) $(ab/p) = (a/p)(b/p)$

(d) $(a/p) \equiv a^{(p-1)/2} \pmod{p}$

(e) $(-1/p) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod{4}, \\ -1, & \text{if } p \equiv 3 \pmod{4}. \end{cases}$

Gauss's Lemma

Theorem 3.1 *Gauss's Lemma*

Let p be an odd prime and $a \not\equiv 0 \pmod{p}$. Let S be the set $S = \left\{a, 2a, 3a, \ldots, \frac{p-1}{2}a\right\}$ consisting of the first $\frac{p-1}{2}$ positive multiples of a. If n denotes the number of members of S whose least positive residue modulo p exceeds $\frac{p}{2}$, then $(a/p) = (-1)^n$.

Proposition 3.2 *The quadratic character of 2*

If p is an odd prime then

$$(2/p) = \begin{cases} 1, & \text{if } p \equiv 1 \pmod{8} \text{ or } p \equiv 7 \pmod{8}, \\ -1, & \text{if } p \equiv 3 \pmod{8} \text{ or } p \equiv 5 \pmod{8}. \end{cases}$$

Propositions 3.4 and 3.5

There are infinitely many primes of each of the forms $4k + 1$ and $8k - 1$.

The Law of Quadratic Reciprocity

Theorem 4.1 *The Law of Quadratic Reciprocity (LQR)*

If p and q are distinct odd primes then $(p/q)(q/p) = (-1)^{(p-1)(q-1)/4}$.

Theorem 4.2 *An alternative formulation of LQR*

If p and q are distinct odd primes then

$$(p/q) = \begin{cases} (q/p), & \text{if either } p \equiv 1 \pmod{4} \quad \text{or} \quad q \equiv 1 \pmod{4}, \\ -(q/p), & \text{if } p \equiv q \equiv 3 \pmod{4}. \end{cases}$$

Lemma 4.5

Let p be an odd prime and a an odd integer not divisible by p. Then $(a/p) = (-1)^{\alpha(a,p)}$, where

$$\alpha(a, p) = \sum_{k=1}^{(p-1)/2} \operatorname{int}\left(\frac{ka}{p}\right).$$

Proposition 4.7 *The quadratic character of 3*

If $p > 3$ is prime then

$$(3/p) = \begin{cases} 1, & \text{if } p \equiv \pm 1 \pmod{12}, \\ -1, & \text{if } p \equiv \pm 5 \pmod{12}. \end{cases}$$

Definition 4.8 *The Jacobi symbol*

Let m and n be integers, with n odd, $n \geq 3$ and $\text{hcf}(m, n) = 1$. If we write $n = p_1 p_2 \cdots p_r$ as a product of (not necessarily distinct) primes, then the Jacobi symbol (m/n) is defined by $(m/n) = (m/p_1)(m/p_2)\cdots(m/p_r)$, where the symbols on the right-hand side are Legendre symbols.

Proposition 4.9

If m and n are odd integers both exceeding 1 with $\text{hcf}(m, n) = 1$, then:

(a) $(-1/n) = (-1)^{(n-1)/2}$, Euler's Criterion

(b) $(2/n) = (-1)^{(n^2-1)/8}$, Proposition 3.2

(c) $(m/n)(n/m) = (-1)^{(m-1)(n-1)/4}$, LQR.

Chapter 11: Rings and polynomials

Rings

Definition 1.1 *Ring axioms*

Let R be a set and let $+$ and $\cdot$ be binary operations defined on R. Then $(R, +, \cdot)$ is a **ring** if the following axioms hold.

Axioms for addition:

R1 Closure For all $a, b \in R$, $a + b \in R$.

R2 Associativity For all $a, b, c \in R$, $a + (b + c) = (a + b) + c$.

R3 Additive identity There exists an additive identity $0 \in R$ such that, for all $a \in R$, $a + 0 = a = 0 + a$.

R4 Additive inverses For each $a \in R$, there exists an additive inverse $-a \in R$ such that $a + (-a) = 0 = (-a) + a$.

R5 Commutativity For all $a, b \in R$, $a + b = b + a$.

Axioms for multiplication:

R6 Closure For all $a, b \in R$, $a \cdot b \in R$.

R7 Associativity For all $a, b, c \in R$, $a \cdot (b \cdot c) = (a \cdot b) \cdot c$.

R8 Multiplicative identity There exists a multiplicative identity $1 \in R$ such that, for all $a \in R$, $a \cdot 1 = a = 1 \cdot a$.

Axioms combining addition and multiplication:

R9 Distributive laws For all $a, b, c \in R$, multiplication is left and right distributive over addition in R:
$a \cdot (b + c) = (a \cdot b) + (a \cdot c)$ and $(a + b) \cdot c = (a \cdot c) + (b \cdot c)$.

$(R, +, \cdot)$ is a **commutative ring** if the following extra axiom holds.

R10 Commutativity For all $a, b \in R$, $a \cdot b = b \cdot a$.

Lemma 1.2

Let $(R, +, \cdot)$ be a ring. Then the additive identity $0 \in R$ is unique. Furthermore, for every $a \in R$ the additive inverse $(-a)$ is unique.

Proposition 1.5 *Basic properties of rings*

Let $(R, +, \cdot)$ be a ring. Then:

(a) $0 \cdot a = a \cdot 0 = 0$ for all $a \in R$

(b) $-(-a) = a$ for all $a \in R$

(c) $a \cdot (-b) = (-a) \cdot b = -(a \cdot b)$ for all $a, b \in R$

(d) $(-a) \cdot (-b) = a \cdot b$ for all $a, b \in R$.

Definition 1.6 *Subring*

Let R be a ring. We say that S is a **subring** of R if:

(a) the set S is a subset of R

(b) S is a ring when equipped with the same binary operations as R

(c) the ring S has the same multiplicative identity as R.

Lemma 1.7

Let R be a ring, and S a subring of R. Then the additive identity of S is the same as the additive identity of R.

Lemma 1.8 *Subring criterion*

Let R be a ring and let S be a subset of R. Then S is a subring of R if:

SR1 for all $s, t \in S$, $s - t \in S$

SR2 for all $s, t \in S$, $st \in S$

SR3 the multiplicative identity of R is in S.

Definition 1.10 *Zero divisor*

Let R be a ring. A non-zero element $a \in R$ is a **zero divisor** in R if there is a non-zero element $b \in R$ for which $ab = 0$ or $ba = 0$.

Remark

From now on, all rings are considered to be commutative.

Definition 1.13 *Unit*

An element a of a ring R is a **unit** if there is an element $a^{-1} \in R$ for which $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Definition 1.14 *Field*

A ring R is a **field** if the additive and multiplicative identities are distinct, and every non-zero element is a unit. In other words, R has distinct elements 0 and 1, axioms R1–R10 hold, and so does:

R11 Multiplicative inverses For every non-zero $a \in R$, there exists $a^{-1} \in R$ such that $a \cdot a^{-1} = a^{-1} \cdot a = 1$.

Lemma 1.17

Let F be a field. Then F has no zero divisors.

Lemma 1.18

Let $a \in \mathbb{Z}_n$ be non-zero, where $n \in \mathbb{N}$. Then a is either a zero divisor or a unit.

Corollary 1.19

The ring $\mathbb{Z}_n$ is a field if, and only if, n is a prime number.

Definition 1.20 *Subfield*

A subset S of a field F that is a field with the same binary operations and multiplicative identity as F is known as a **subfield** of F.

Lemma 1.21 *Subfield criterion*

Let F be a field, and let S be any subset of F. Then S is a subfield of F if:

(a) S is a subring of F

(b) every non-zero element of S has a multiplicative inverse in S.

Polynomials over fields

Definition 2.1 *Polynomial ring over a field*

Let F be a field. Then a **polynomial over** F with the variable x is a polynomial of the form $f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$, where $a_0, a_1, a_2, \ldots, a_n \in F$ and $n \geq 0$, where x^0 is defined to be 1.

The **polynomial ring over** F is

$$F[x] = \{a_0 + a_1x + \cdots + a_nx^n : a_0, a_1, \ldots, a_n \in F, n \geq 0\},$$

with addition defined by

$$\left(\sum_{i=0}^{n} a_i x^i\right) + \left(\sum_{i=0}^{m} b_i x^i\right) = \sum_{i=0}^{\max(m,n)} (a_i + b_i)\, x^i$$

and multiplication defined by

$$\left(\sum_{i=0}^{n} a_i x^i\right) \cdot \left(\sum_{j=0}^{m} b_j x^j\right) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_i \cdot b_j\right) x^k.$$

Note that this is the 'usual' multiplication of polynomials.

Definition 2.4 *Degree of a polynomial*

Let F be a field, and let $f(x) = a_0 + a_1x + \cdots + a_nx^n$ be a polynomial in $F[x]$. The **degree** of $f(x)$, $\deg(f)$, is the largest $k \geq 0$ for which $a_k \neq 0$. This non-zero coefficient a_k is known as the **leading coefficient**.

The degree of the polynomial 0_x is set to be $-\infty$; in other words, we can choose the degree of 0_x to be as largely negative as we need.

Lemmas 2.6 and 2.7

Let F be a field, and let $f(x)$ and $g(x)$ be polynomials in $F[x]$. Then:

(a) $\deg(fg) = \deg(f) + \deg(g)$

(b) $\deg(f + g) \leq \max(\deg(f), \deg(g))$.

Lemma 2.8

Let F be a field, and let $f(x), g(x) \in F[x]$ be polynomials such that $f(x)g(x) = 0_x$. Then either $f(x) = 0_x$, or $g(x) = 0_x$ (or both).

Lemma 2.9

An element of a polynomial ring $F[x]$ over a field F has a multiplicative inverse (that is, the element is a unit) if, and only if, it has degree 0.

Lemma 2.10 *Cancellation of polynomials*

If $f(x)$ is a non-zero polynomial and $f(x)g(x) = f(x)h(x)$ in $F[x]$, then $g(x) = h(x)$.

Definition 2.11 *Associate and monic*

Let F be a field, and $f(x), g(x) \in F[x]$.

(a) We say that $f(x)$ is an **associate** of $g(x)$ if $f(x) = ug(x)$ where u is a unit in $F[x]$.

(b) The polynomial $f(x)$ is **monic** if the leading coefficient is equal to 1.

Divisibility of polynomials

Theorem 3.1 *Division Algorithm for polynomials*

Let F be a field and let f and g be polynomials in $F[x]$, with $g \neq 0_x$. Then there exist unique polynomials q and r in $F[x]$ such that

(a) $\deg(r) < \deg(g)$ and (b) $f = qg + r$.

The polynomials q and r are known as the **quotient** and **remainder** when dividing f by g.

Definition 3.3 *Factors of a polynomial*

Let F be a field, and f, g polynomials in $F[x]$. We say that g **divides** f (or g is a **factor** of f) if there is some polynomial $h \in F[x]$, such that $f = gh$.

If f is non-zero and neither g nor h are units then g and h are **proper factors** of f.

Definition 3.4 *Highest common factor of two polynomials*

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then the **highest common factor** of f and g, written $\text{hcf}(f, g)$, is a monic polynomial of largest degree satisfying the following:

(a) $\text{hcf}(f, g)$ divides both f and g.

(b) Any polynomial $d \in F[x]$ that divides both f and g must also divide $\text{hcf}(f, g)$.

Theorem 3.5

Let F be a field, and f, g two polynomials in $F[x]$, not both equal to 0_x. Then $\text{hcf}(f, g)$ is unique, and there exist polynomials $a, b \in F[x]$ such that $\text{hcf}(f, g) = af + bg$.

Definition 3.6 *Coprime*

Let F be a field, and let f, g be polynomials in $F[x]$, not both equal to 0_x. If $\text{hcf}(f, g) = 1_x$ then f and g are **coprime**.

Corollary 3.7

The polynomials $f, g \in F[x]$, not both equal to 0_x, are coprime if, and only if, there exist polynomials $a, b \in F[x]$ such that $af + bg = 1_x$.

Strategy: Euclidean Algorithm to find the hcf of two polynomials

Given a field F, and two polynomials $f, g \in F[x]$, not both equal to 0_x, let $f^* = f$ and $g^* = g$.

1. Apply the Division Algorithm (Theorem 3.1) to f^*, g^* to find $q, r \in F[x]$ for which $f^* = qg^* + r$.
2. If $r = 0_x$ then stop: $\text{hcf}(f, g) = a^{-1}g^*$, where $a \in F$ is the coefficient of the highest power of x in g^*.
3. Otherwise, $r \neq 0_x$. Replace f^* by g^*, and g^* by r, and go back to step 1.

Definition 3.10 *Least common multiple of two polynomials*

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then the **least common multiple** of f and g, $\text{lcm}(f, g)$, is the monic polynomial $\ell \in F[x]$ satisfying the following:

(a) f and g both divide ℓ.

(b) For any polynomial $h \in F[x]$ that is divisible by both f and g, we have $\deg(h) \geq \deg(\ell)$.

Lemma 3.11

Let F be a field, and f, g two non-zero polynomials in $F[x]$. Then $\ell = \text{lcm}(f, g)$ is unique.

Proposition 3.12

Let $f(x), g(x)$ be non-zero monic polynomials with coefficients from a field F. Then $\text{lcm}(f, g) \cdot \text{hcf}(f, g) = f \cdot g$.

Factorising polynomials

Definition 4.1 *Root*

Let F be a field and let $f(x)$ be a non-zero polynomial in $F[x]$. Then $a \in F$ is called a **root** of $f(x)$ if $f(a) = 0$.

Theorem 4.2 *Factor Theorem*

Let F be a field, $a \in F$ and $f \in F[x]$ with $f \neq 0_x$. Then a is a root of f if, and only if, $f(x) = (x - a)q(x)$ for some $q(x) \in F[x]$, with $\deg(q) = \deg(f) - 1$.

Proposition 4.3

Let $f(x)$ be a non-zero polynomial of degree n in $F[x]$. Then the equation $f(x) = 0$ has at most n distinct roots in F.

Definition 4.5 *Irreducibility*

Let F be a field, and $f \in F[x]$ a non-zero polynomial over F. We say that f is **irreducible** if it is not a unit and it has no proper factors. We say that f is **reducible** if it is not a unit but it does have proper factors.

Lemma 4.6

Let f, g be polynomials in $\mathbb{Q}[x]$, such that f and g are associates. Then f is irreducible if, and only if, g is irreducible.

Theorem 4.7 *Rational Root Test*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$ of degree $n \geq 1$, with a_n and a_0 non-zero. Then any rational root $\frac{p}{q} \in \mathbb{Q}$ of $f(x)$ with $\text{hcf}(p, q) = 1$ must satisfy the following:

(a) p is an integer factor of a_0

(b) q is an integer factor of a_n.

Definition 4.9 *Irreducibility over* $\mathbb{Z}$

Let f, g be polynomials in $\mathbb{Z}[x]$, with $f \neq 0_x$.

(a) We say that g **divides** f (or is a **factor** of f) if there exists some $h \in \mathbb{Z}[x]$ such that $f = gh$.

(b) If $f = gh$ and neither g nor h is equal to ± 1, then g, h are **proper factors**.

(c) We say that f is **irreducible** over $\mathbb{Z}$ if $f \neq \pm 1$ and it has no proper factors in $\mathbb{Z}[x]$.

Definition 4.11 *Content of a polynomial*

Let $f(x) = \sum_{i=0}^{n} a_i x^i$ be a polynomial in $\mathbb{Z}[x]$.

(a) The **content** of f is the highest common factor of all its coefficients: $c_f = \text{hcf}(a_0, a_1, \ldots, a_n)$.

(b) $f(x)$ is **primitive** if $c_f = 1$, or, in other words, if its coefficients have no non-trivial common factor.

Lemma 4.12

The polynomials f and g are primitive in $\mathbb{Z}[x]$ if, and only if, their product fg is a primitive polynomial in $\mathbb{Z}[x]$.

Theorem 4.13 *Gauss's Lemma*

Let $f(x)$ be a primitive polynomial in $\mathbb{Z}[x]$. If we can write $f = gh$ with $g, h \in \mathbb{Q}[x]$, then we can write $f = GH$ with $G, H \in \mathbb{Z}[x]$, $\deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Theorem 4.15 *Eisenstein's Criterion*

Let $f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$ be a polynomial in $\mathbb{Z}[x]$. If there is a prime $p \in \mathbb{Z}$ such that

(a) p divides a_i for $i = 0, \ldots, n-1$, but p does not divide a_n, and

(b) p^2 does not divide a_0,

then $f(x)$ is irreducible in $\mathbb{Q}[x]$.

Lemma 4.17

Let $f(x)$ be a polynomial in $\mathbb{Z}[x]$. Then $f(x)$ is irreducible in $\mathbb{Q}[x]$ if, and only if, $f(x - b)$ is irreducible in $\mathbb{Q}[x]$ for any $b \in \mathbb{Q}$.

Lemma 4.19

Let f be a polynomial in $\mathbb{Z}[x]$ and let p be a prime that does not divide the leading coefficient of f. If f is irreducible in $\mathbb{Z}_p[x]$, then f is irreducible in $\mathbb{Z}[x]$.

Chapter 12: Fermat's Last Theorem and unique factorisation

Fermat's Last Theorem and Diophantine equations

Definition 1.1 *Pythagorean triples*

A **Pythagorean triple** is a triple (x, y, z) of positive integers such that $x^2 + y^2 = z^2$. The triple is **primitive** if $\text{hcf}(x, y) = 1$.

Convention for Pythagorean triples

If (x, y, z) is any primitive Pythagorean triple, then x is even while y and z are both odd.

Theorem 1.2 *Primitive solutions of the Pythagorean equation*

The primitive Pythagorean triples are the triples $(2mn, m^2 - n^2, m^2 + n^2)$, where m and n are coprime natural numbers of opposite parity with $m > n$.

Corollary 1.4 *All solutions of the Pythagorean equation*

The Pythagorean triples (x, y, z) are given by $x = 2kmn$, $y = k(m^2 - n^2)$ and $z = k(m^2 + n^2)$, where $k \geq 1$ is any integer and m and n are coprime natural numbers with opposite parity and $m > n$.

Fermat's method of infinite descent: to show that a Diophantine equation has no positive solutions

1. Assume that the Diophantine equation has some positive solution (the *basis*).
2. Show that from *any* solution a 'smaller' positive solution can be found (the *descent step*).

Since an infinite chain of decreasing positive solutions cannot exist, there is no positive solution.

Theorem 1.6, Corollary 1.7, Theorem 1.9, Corollary 1.10

The following Diophantine equations have no positive solutions:

$$x^4 + y^4 = z^2, \qquad x^4 + y^4 = z^4, \qquad x^4 + 4y^4 = z^2, \qquad x^4 - y^4 = z^2.$$

Integral domains

Lemma 2.2

Let R be a commutative ring, and u a unit of R. Then u^k is a unit of R for all $k \geq 1$.

Definition 2.3 *Associate*

Let R be a commutative ring. An **associate** of $a \in R$ is any element that can be written as ua where u is a unit in R.

Definition 2.4 *Factor*

Let R be a commutative ring. We say that $a \in R$ is a **factor** or **divisor** of $c \in R$ (or a **divides** c) if there is $b \in R$, such that $c = ab$. If c is non-zero and neither a nor b are units then they are **proper factors** of c.

Definition 2.5 *Irreducible*

Let R be a commutative ring. We say that $a \in R$ is **irreducible** if it is not a unit, it is not zero and whenever $a = bc$ for $b, c \in R$, then one of b or c is a unit.

Definition 2.6 *Prime*

Let R be a commutative ring. A non-zero, non-unit element p of R is **prime** if whenever $p \mid ab$ for $a, b \in R$, then either $p \mid a$ or $p \mid b$.

Definition 2.8 *Integral domain*

A commutative ring is an **integral domain** if it has no zero divisors. Equivalently, a ring R is an integral domain if for $a, b \in R$, whenever $a \cdot b = 0$ then either $a = 0$ or $b = 0$.

Lemma 2.11 *Cancellation in an integral domain*

Let R be a commutative ring. Then R is an integral domain if, and only if, for $a, b, c \in R$ with c non-zero, $ac = bc$ implies $a = b$.

Theorem 2.12

Let R be an integral domain and let p be a prime element of R. Then p is irreducible.

Lemma 2.13 *Properties of division in an integral domain*

Let R be an integral domain and let a, b, c be non-zero elements in R.

(a) If $a \mid c$, then $a \mid (c + na)$ for any $n \in R$.

(b) If $a \mid b$ and $b \mid a$, then a is an associate of b.

(c) If $a \mid b$ and $b \mid c$, then $a \mid c$.

(d) If $a \mid b$ and $a \mid c$, then $a \mid (mb + nc)$ for any $m, n \in R$.

(e) Every element of R divides 0.

(f) If u is a unit in R, then $u \mid n$ for every $n \in R$.

Definition 2.14 *Norm*

Let R be an integral domain. A **norm** is a function $N : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$, mapping the non-zero elements of R to the non-negative integers.

N is a **Euclidean norm** if for non-zero $a, b \in R$ we have

(a) $N(a) \leq N(ab)$, and

(b) there exist $q, r \in R$ such that $a = bq + r$, and either $r = 0$ or $N(r) < N(b)$.

N is a **multiplicative norm** if for non-zero $a, b \in R$ we have $N(ab) = N(a)N(b)$.

Euclidean domains

Definition 3.1 *Euclidean domain*

Let R be an integral domain. If there exists a Euclidean norm $N : R \setminus \{0\} \to \mathbb{N} \cup \{0\}$ on R, then R is a **Euclidean domain**.

Theorem 3.3 *Division Algorithm for Euclidean domains*

Let R be a Euclidean domain with Euclidean norm N and let $a, b \in R$ with b non-zero. Then there exist $q, r \in R$ such that $a = bq + r$ with $N(r) < N(b)$ or $r = 0$.

Lemma 3.5

Let R be a Euclidean domain with norm N. Then:

(a) $N(1)$ is the smallest value taken by the function N over all non-zero elements of R.

(b) $N(u) = N(1)$ if, and only if, u is a unit in R.

(c) If a, b are associates in R then $N(a) = N(b)$.

(d) For any non-zero $a \in R$, if $b \neq 0$ is not a unit then $N(a) < N(ab)$.

Definition 3.6 *Highest common factor*

Let a and b be elements of an integral domain R. An element h is a **highest common factor** (**hcf**) of a and b if it satisfies the following:

(a) $h \mid a$ and $h \mid b$

(b) if $d \mid a$ and $d \mid b$, then $d \mid h$.

Theorem 3.8

Let R be a Euclidean domain with norm N, and let $a, b \in R$ with b non-zero. Then a and b have a highest common factor. If $h = \text{hcf}(a, b)$, then there are elements $s, t \in R$ so that $h = sa + tb$.

Lemma 3.10

Suppose that p is an irreducible element of a Euclidean domain R, that a is a non-zero element of R and that p does not divide a. Then there exist $s, t \in R$ with $ps + at = 1$.

Proposition 3.11

Let R be a Euclidean domain with norm N, and let p be an irreducible element in R. Then p is prime in R.

Important identity for two squares

$$(a^2 + b^2)(c^2 + d^2) = (ac + bd)^2 + (ad - bc)^2$$

Theorem 3.13

A prime number p can be expressed as the sum of two squares if, and only if, it is either even, or odd of the form $4k + 1$ for some integer k.

Unique factorisation domains

Definition 4.1 *Unique factorisation domain*

An integral domain R is a **unique factorisation domain** (**UFD**) if

(a) every non-zero non-unit element of R is expressible as a product of irreducible elements of R, and

(b) if $a \in R$ can be written as a product of irreducibles in two different ways, $a = p_1 p_2 \cdots p_m = q_1 q_2 \cdots q_n$, then $m = n$ and the p_i, q_j pair off as associates.

Theorem 4.2

Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a UFD if, and only if, every irreducible element in R is prime.

Corollary 4.3

In a unique factorisation domain, every irreducible element is prime.

Definition 4.4 *Least common multiple*

Let a and b be elements of an integral domain R. An element ℓ is a **least common multiple** (**lcm**) of a and b if it satisfies the following:

(a) a divides ℓ and b divides ℓ

(b) for any $h \in R$ such that a divides h and b divides h, then ℓ also divides h.

Theorem 4.5

Every Euclidean domain is a unique factorisation domain.

Proposition 4.7

Let R be a unique factorisation domain and $a, b \in R$, not both of which are zero. Then $\mathrm{hcf}(a, b)$ exists in R.

Lemma 4.8

Every polynomial in $\mathbb{Z}[x]$ has a factorisation into irreducible polynomials in $\mathbb{Z}[x]$.

Proposition 4.9

Let f be an irreducible polynomial in $\mathbb{Z}[x]$. Then f is prime in $\mathbb{Z}[x]$.

Lemma 4.10

Let $p \in \mathbb{Z}$. If p is prime in $\mathbb{Z}$, then it is prime in $\mathbb{Z}[x]$.

Theorem 4.11

The ring $\mathbb{Z}[x]$ is a unique factorisation domain.

Theorem 4.12 *Uniqueness of representation*

The expression of a prime p of the form $4k + 1$ as a sum of two squares is unique except for the order of the two summands.

Proposition 4.13 *Sums of two squares*

A positive integer n can be expressed as a sum of two squares if, and only if, each of its prime divisors of the form $4k + 3$, if any, occurs to an even power.

Proposition 4.14

The equation $a^2 + 2 = b^3$ has exactly one positive integer solution, namely $a = 5$, $b = 3$.

Appendix 2: Values of $\tau(n)$, $\sigma(n)$ and $\phi(n)$ for small n

n	1	2	3	4	5	6	7	8	9	10
$\tau(n)$	1	2	2	3	2	4	2	4	3	4
$\sigma(n)$	1	3	4	7	6	12	8	15	13	18
$\phi(n)$	1	1	2	2	4	2	6	4	6	4

Appendix 3: Table of primes

The following table lists all the odd integers up to 3999, excluding the multiples of 5. The entry 'P' in the table indicates that the number is prime. Otherwise the smallest prime factor of the number is given. The numbers are arranged in blocks of 100, with the units digit at the head of the column and the number of 10s in the left-hand column.

	1	3	7	9		1	3	7	9		1	3	7	9		1	3	7	9
0		P	P	3	50	3	P	3	P	100	7	17	19	P	150	19	3	11	3
1	P	P	P	P	51	7	3	11	3	101	3	P	3	P	151	P	17	37	7
2	3	P	3	P	52	P	P	17	23	102	P	3	13	3	152	3	P	3	11
3	P	3	P	3	53	3	13	3	7	103	P	P	17	P	153	P	3	29	3
4	P	P	P	7	54	P	3	P	3	104	3	7	3	P	154	23	P	7	P
5	3	P	3	P	55	19	7	P	13	105	P	3	7	3	155	3	P	3	P
6	P	3	P	3	56	3	P	3	P	106	P	P	11	P	156	7	3	P	3
7	P	P	7	P	57	P	3	P	3	107	3	29	3	13	157	P	11	19	P
8	3	P	3	P	58	7	11	P	19	108	23	3	P	3	158	3	P	3	7
9	7	3	P	3	59	3	P	3	P	109	P	P	P	7	159	37	3	P	3
10	P	P	P	P	60	P	3	P	3	110	3	P	3	P	160	P	7	P	P
11	3	P	3	7	61	13	P	P	P	111	11	3	P	3	161	3	P	3	P
12	11	3	P	3	62	3	7	3	17	112	19	P	7	P	162	P	3	P	3
13	P	7	P	P	63	P	3	7	3	113	3	11	3	17	163	7	23	P	11
14	3	11	3	P	64	P	P	P	11	114	7	3	31	3	164	3	31	3	17
15	P	3	P	3	65	3	P	3	P	115	P	P	13	19	165	13	3	P	3
16	7	P	P	13	66	P	3	23	3	116	3	P	3	7	166	11	P	P	P
17	3	P	3	P	67	11	P	P	7	117	P	3	11	3	167	3	7	3	23
18	P	3	11	3	68	3	P	3	13	118	P	7	P	29	168	41	3	7	3
19	P	P	P	P	69	P	3	17	3	119	3	P	3	11	169	19	P	P	P
20	3	7	3	11	70	P	19	7	P	120	P	3	17	3	170	3	13	3	P
21	P	3	7	3	71	3	23	3	P	121	7	P	P	23	171	29	3	17	3
22	13	P	P	P	72	7	3	P	3	122	3	P	3	P	172	P	P	11	7
23	3	P	3	P	73	17	P	11	P	123	P	3	P	3	173	3	P	3	37
24	P	3	13	3	74	3	P	3	7	124	17	11	29	P	174	P	3	P	3
25	P	11	P	7	75	P	3	P	3	125	3	7	3	P	175	17	P	7	P
26	3	P	3	P	76	P	7	13	P	126	13	3	7	3	176	3	41	3	29
27	P	3	P	3	77	3	P	3	19	127	31	19	P	P	177	7	3	P	3
28	P	P	7	17	78	11	3	P	3	128	3	P	3	P	178	13	P	P	P
29	3	P	3	13	79	7	13	P	17	129	P	3	P	3	179	3	11	3	7
30	7	3	P	3	80	3	11	3	P	130	P	P	P	7	180	P	3	13	3
31	P	P	P	11	81	P	3	19	3	131	3	13	3	P	181	P	7	23	17
32	3	17	3	7	82	P	P	P	P	132	P	3	P	3	182	3	P	3	31
33	P	3	P	3	83	3	7	3	P	133	11	31	7	13	183	P	3	11	3
34	11	7	P	P	84	29	3	7	3	134	3	17	3	19	184	7	19	P	43
35	3	P	3	P	85	23	P	P	P	135	7	3	23	3	185	3	17	3	11
36	19	3	P	3	86	3	P	3	11	136	P	29	P	37	186	P	3	P	3
37	7	P	13	P	87	13	3	P	3	137	3	P	3	7	187	P	P	P	P
38	3	P	3	P	88	P	P	P	7	138	P	3	19	3	188	3	7	3	P
39	17	3	P	3	89	3	19	3	29	139	13	7	11	P	189	31	3	7	3
40	P	13	11	P	90	17	3	P	3	140	3	23	3	P	190	P	11	P	23
41	3	7	3	P	91	P	11	7	P	141	17	3	13	3	191	3	P	3	19
42	P	3	7	3	92	3	13	3	P	142	7	P	P	P	192	17	3	41	3
43	P	P	19	P	93	7	3	P	3	143	3	P	3	P	193	P	P	13	7
44	3	P	3	P	94	P	23	P	13	144	11	3	P	3	194	3	29	3	P
45	11	3	P	3	95	3	P	3	7	145	P	P	31	P	195	P	3	19	3
46	P	P	P	7	96	31	3	P	3	146	3	7	3	13	196	37	13	7	11
47	3	11	3	P	97	P	7	P	11	147	P	3	7	3	197	3	P	3	P
48	13	3	P	3	98	3	P	3	23	148	P	P	P	P	198	7	3	P	3
49	P	17	7	P	99	P	3	P	3	149	3	P	3	P	199	11	P	P	P

	1	3	7	9		1	3	7	9		1	3	7	9		1	3	7	9
200	3	P	3	7	250	41	P	23	13	300	P	3	31	3	350	3	31	3	11
201	P	3	P	3	251	3	7	3	11	301	P	23	7	P	351	P	3	P	3
202	43	7	P	P	252	P	3	7	3	302	3	P	3	13	352	7	13	P	P
203	3	19	3	P	253	P	17	43	P	303	7	3	P	3	353	3	P	3	P
204	13	3	23	3	254	3	P	3	P	304	P	17	11	P	354	P	3	P	3
205	7	P	11	29	255	P	3	P	3	305	3	43	3	7	355	53	11	P	P
206	3	P	3	P	256	13	11	17	7	306	P	3	P	3	356	3	7	3	43
207	19	3	31	3	257	3	31	3	P	307	37	7	17	P	357	P	3	7	3
208	P	P	P	P	258	29	3	13	3	308	3	P	3	P	358	P	P	17	37
209	3	7	3	P	259	P	P	7	23	309	11	3	19	3	359	3	P	3	59
210	11	3	7	3	260	3	19	3	P	310	7	29	13	P	360	13	3	P	3
211	P	P	29	13	261	7	3	P	3	311	3	11	3	P	361	23	P	P	7
212	3	11	3	P	262	P	43	37	11	312	P	3	53	3	362	3	P	3	19
213	P	3	P	3	263	3	P	3	7	313	31	13	P	43	363	P	3	P	3
214	P	P	19	7	264	19	3	P	3	314	3	7	3	47	364	11	P	7	41
215	3	P	3	17	265	11	7	P	P	315	23	3	7	3	365	3	13	3	P
216	P	3	11	3	266	3	P	3	17	316	29	P	P	P	366	7	3	19	3
217	13	41	7	P	267	P	3	P	3	317	3	19	3	11	367	P	P	P	13
218	3	37	3	11	268	7	P	P	P	318	P	3	P	3	368	3	29	3	7
219	7	3	13	3	269	3	P	3	P	319	P	31	23	7	369	P	3	P	3
220	31	P	P	47	270	37	3	P	3	320	3	P	3	P	370	P	7	11	P
221	3	P	3	7	271	P	P	11	P	321	13	3	P	3	371	3	47	3	P
222	P	3	17	3	272	3	7	3	P	322	P	11	7	P	372	61	3	P	3
223	23	7	P	P	273	P	3	7	3	323	3	53	3	41	373	7	P	37	P
224	3	P	3	13	274	P	13	41	P	324	7	3	17	3	374	3	19	3	23
225	P	3	37	3	275	3	P	3	31	325	P	P	P	P	375	11	3	13	3
226	7	31	P	P	276	11	3	P	3	326	3	13	3	7	376	P	53	P	P
227	3	P	3	43	277	17	47	P	7	327	P	3	29	3	377	3	7	3	P
228	P	3	P	3	278	3	11	3	P	328	17	7	19	11	378	19	3	7	3
229	29	P	P	11	279	P	3	P	3	329	3	37	3	P	379	17	P	P	29
230	3	7	3	P	280	P	P	7	53	330	P	3	P	3	380	3	P	3	13
231	P	3	7	3	281	3	29	3	P	331	7	P	31	P	381	37	3	11	3
232	11	23	13	17	282	7	3	11	3	332	3	P	3	P	382	P	P	43	7
233	3	P	3	P	283	19	P	P	17	333	P	3	47	3	383	3	P	3	11
234	P	3	P	3	284	3	P	3	7	334	13	P	P	17	384	23	3	P	3
235	P	13	P	7	285	P	3	P	3	335	3	7	3	P	385	P	P	7	17
236	3	17	3	23	286	P	7	47	19	336	P	3	7	3	386	3	P	3	53
237	P	3	P	3	287	3	13	3	P	337	P	P	11	31	387	7	3	P	3
238	P	P	7	P	288	43	3	P	3	338	3	17	3	P	388	P	11	13	P
239	3	P	3	P	289	7	11	P	13	339	P	3	43	3	389	3	17	3	7
240	7	3	29	3	290	3	P	3	P	340	19	41	P	7	390	47	3	P	3
241	P	19	P	41	291	41	3	P	3	341	3	P	3	13	391	P	7	P	P
242	3	P	3	7	292	23	37	P	29	342	11	3	23	3	392	3	P	3	P
243	11	3	P	3	293	3	7	3	P	343	47	P	7	19	393	P	3	31	3
244	P	7	P	31	294	17	3	7	3	344	3	11	3	P	394	7	P	P	11
245	3	11	3	P	295	13	P	P	11	345	7	3	P	3	395	3	59	3	37
246	23	3	P	3	296	3	P	3	P	346	P	P	P	P	396	17	3	P	3
247	7	P	P	37	297	P	3	13	3	347	3	23	3	7	397	11	29	41	23
248	3	13	3	19	298	11	19	29	7	348	59	3	11	3	398	3	7	3	P
249	47	3	11	3	299	3	41	3	P	349	P	7	13	P	399	13	3	7	3

Book D: Metric spaces 1

Glossary

Chapter 15, Section 2 **Cantor metric** The Cantor distance, which is a metric.

Chapter 14, Section 1 **Cartesian product** The product set $A_1 \times A_2 \times \cdots \times A_k$ of the sets $A_1, A_2, \ldots, A_k$.

Chapter 13, Section 2 **closed bounded interval** An interval of the form $[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$ for real numbers a and b.

Chapter 14, Section 1 **coordinate sequences** The n individual sequences $(a_{j,k})$ for $j = 1, 2, \ldots, n$ of a sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$.

Chapter 13, Section 1 **constant sequence** A sequence all of whose terms are the same.

Chapter 14, Section 2 **d-closed ball** A closed ball in a metric space with metric d.

Chapter 14, Section 2 **d-open ball** An open ball in a metric space with metric d.

Chapter 14, Section 2 **d-sphere** A sphere in a metric space with metric d.

Chapter 14, Section 2 **distance function** A function that defines a notion of distance between pairs of points of a space X. It need not be a metric.

Chapter 13, Section 2 **essential discontinuity** A discontinuity of a function f where it is not possible to change the definition of f in such a way as to make it continuous at that point.

Chapter 14, Section 1 **Euclidean distance in $\mathbb{R}^n$** The distance in $\mathbb{R}^n$ given by $d^{(n)}(\mathbf{x}, \mathbf{y}) = \sqrt{(y_1 - x_1)^2 + (y_2 - x_2)^2 + \cdots + (y_n - x_n)^2}$.

Chapter 14, Section 2 **Euclidean n-space** The metric space $(\mathbb{R}^n, d^{(n)})$, where $d^{(n)}$ denotes the usual Euclidean distance between points.

Chapter 14, Section 2 **Euclidean metric for $\mathbb{R}^n$** The Euclidean distance function $d^{(n)}$ for $\mathbb{R}^n$, for each $n \in \mathbb{N}$.

Chapter 15, Section 2 **eventually zero** A sequence of real numbers (a_n) where there is an $N \in \mathbb{N}$ such that $a_n = 0$ for each $n > N$.

Chapter 13, Section 2 **existence result** A result that a particular thing (point, function etc.) exists, but does not necessarily give a method for finding it.

Chapter 15, Section 1 **Hamming distance d_H** The metric defined on the words of length n over a set of symbols A that defines the distance between two distinct words as being the number of places at which they differ.

Chapter 15, Section 1 **Hamming metric** The Hamming distance, which is a metric.

Chapter 15, Section 5 **integral metric** The metric defined on $C[0, 1]$ that defines distance between the functions f and g as

$$d(f, g) = \int_0^1 |g(x) - f(x)| \, dx.$$

ℓ_0 The set consisting of all infinite sequences of real numbers whose terms are eventually zero. **Chapter 15**, Section 2

listable set A countable set (finite or infinite). **Chapter 16**, Section 6

magnitude of a The distance between a point $\mathbf{a}$ and the origin in $\mathbb{R}^n$. **Chapter 13**, Section 3

max metric on the plane The metric defined on the plane that defines distance between two points $\mathbf{x}$ and $\mathbf{y}$ as $e_\infty(\mathbf{x}, \mathbf{y}) = \max\{|y_1 - x_1|, |y_2 - x_2|\}$. **Chapter 15**, Section 1

metrisable topology A topology arising from a metric. **Chapter 16**, Section 7

null sequence A sequence in $\mathbb{R}^n$ $(n \geq 1)$ that converges to $\mathbf{0}$. **Chapter 14**, Section 1

origin The point $\mathbf{0} = (0, 0, \ldots, 0)$ in $\mathbb{R}^n$; in $\mathbb{R}^2$, the point $\mathbf{0} = (0, 0)$. **Chapter 13**, Section 3

product metric A metric constructed from the metrics d_1 and d_2 on the product set $X = X_1 \times X_2$, where (X_1, d_1) and (X_2, d_2) are metric spaces. **Chapter 15**, Section 1

$\mathbb{R}^\infty$ The product of infinitely many copies of $\mathbb{R}$. It is the set of all infinite sequences of real numbers. **Chapter 14**, Section 1

removable discontinuity A point of discontinuity of a function f for which an appropriate change to the definition of f at that point yields a continuous function. **Chapter 13**, Section 2

$\tan^{-1}$ metric The metric defined on the plane that defines distance between two points $\mathbf{x}$ and $\mathbf{y}$ as $d(x, y) = |\tan^{-1} y - \tan^{-1} x|$. **Chapter 15**, Section 1

taxicab metric on the plane The metric defined on the plane that defines distance between two points $\mathbf{x}$ and $\mathbf{y}$ as $e_1(\mathbf{x}, \mathbf{y}) = |y_1 - x_1| + |y_2 - x_2|$. **Chapter 14**, Section 2

Triangle Inequality The inequality that expresses the fact that, in a triangle, the length of any one side cannot exceed the sum of the lengths of the other two sides. **Chapter 13**, Section 3

words of length n over the set of symbols A The elements of the set $X = \underbrace{A \times \cdots \times A}_{n} = A^n$, for some fixed $n \in \mathbb{N}$. **Chapter 15**, Section 1

Chapter 13: Distance and continuity

Functions and sequences

Definition 1.1 *Function between two arbitrary sets*

A **function** f is defined by specifying:

- a set A, the **domain** of f
- a set B, the **codomain** of f
- a **rule** $x \mapsto f(x)$ that associates with each element $x \in A$ a *unique* element $f(x) \in B$.

The element $f(x)$ is called the **image** of x under f.

For $C \subseteq A$, the set $f(C) = \{f(x) : x \in C\} \subseteq B$ is called the **image set** of C under f, and for $D \subseteq B$, the set $f^{-1}(D) = \{x \in A : f(x) \in D\}$ is called the **preimage set** of D under f.

Proposition 1.2 *Solution sets of inequalities 1*

- For $c > 0$, the following inequalities are equivalent:
  (a) $|x| < c$, (b) $-c < x < c$, (c) $-c < x$ and $x < c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| < c\} = \begin{cases} (-c, c), & \text{if } c > 0, \\ \emptyset, & \text{if } c \leq 0. \end{cases}$$
- For $c \geq 0$, the following inequalities are equivalent:
  (a) $|x| \leq c$, (b) $-c \leq x \leq c$, (c) $-c \leq x$ and $x \leq c$.
  Moreover, for $c \in \mathbb{R}$,
  $$\{x \in \mathbb{R} : |x| \leq c\} = \begin{cases} [-c, c], & \text{if } c \geq 0, \\ \emptyset, & \text{if } c < 0. \end{cases}$$

Proposition 1.5 *Solution sets of inequalities 2*

- For $b, c \in \mathbb{R}$ with $c > 0$,
  $$\{x \in \mathbb{R} : |x - b| < c\} = (b - c, b + c).$$
- For $b, c \in \mathbb{R}$ with $c \geq 0$,
  $$\{x \in \mathbb{R} : |x - b| \leq c\} = [b - c, b + c].$$

Proposition 1.6 *Triangle Inequality in* $\mathbb{R}$

For all real numbers a and b,

$$|a + b| \leq |a| + |b|.$$

Corollary 1.7

For all real numbers a, b and c,

$$|b - a| \leq |b - c| + |c - a|.$$

Proposition 1.8 *Reverse Triangle Inequality in* $\mathbb{R}$

For all real numbers a and b,

$$|b-a| \geq \big||b|-|a|\big|.$$

Definition 1.9 *Sequence of real numbers*

A **(real) sequence** is an unending ordered list of real numbers

$$a_1, a_2, a_3, \ldots.$$

The real number a_n ($n \in \mathbb{N}$) is the ***n*th term** of the sequence, and the whole sequence is denoted by (a_n).

A sequence (a_n) is **monotonic increasing** if for each $n \in \mathbb{N}$, $a_n \leq a_{n+1}$. It is **strictly monotonic increasing** if this inequality is strict for each n, $a_n < a_{n+1}$.

A sequence (a_n) is **monotonic decreasing** if for each $n \in \mathbb{N}$, $a_n \geq a_{n+1}$. It is **strictly monotonic decreasing** if this inequality is strict for each n, $a_n > a_{n+1}$.

For any subset A of $\mathbb{R}$, we say that a sequence (a_n) is **in** A if $a_n \in A$ for each n.

Definition 1.10 *Null sequence*

A real sequence (a_n) is a **null sequence** if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n| < \varepsilon \quad \text{whenever } n > N.$$

We say that a null sequence **converges to 0**, and write

$$a_n \to 0 \text{ as } n \to \infty, \quad \text{or simply } a_n \to 0.$$

Theorem 1.13 *Basic null sequences*

The following sequences are null:

- $(1/n^p)$, for any constant $p > 0$
- $(n^p c^n)$, for any constants $p \in \mathbb{R}$ and $|c| < 1$
- $(c^n/n!)$, for any constant $c \in \mathbb{R}$
- $(n^p/n!)$, for any constant $p \in \mathbb{R}$.

Theorem 1.14 *Combination rules for null sequences*

Sum Rule If (a_n) and (b_n) are null, then $(a_n + b_n)$ is null.

Multiple Rule If (a_n) is null, then (λa_n) is null for each $\lambda \in \mathbb{R}$.

Product Rule If (a_n) and (b_n) are null, then $(a_n b_n)$ is null.

Power Rule If (a_n) is null, where $a_n \geq 0$ for each $n \in \mathbb{N}$, and $p > 0$, then (a_n^p) is null.

Lemma 1.15 *Squeeze Rule*

If (a_n) is null and if there is an $M \in \mathbb{N}$ such that $|b_n| \leq |a_n|$ for each $n > M$, then (b_n) is null.

Definition 1.17 *Convergence and divergence*

A real sequence (a_n) **converges** to $l \in \mathbb{R}$ if the sequence $(a_n - l)$ is a null sequence: that is if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$|a_n - l| < \varepsilon \quad \text{whenever } n > N.$$

We say that l is the **limit** of the sequence (a_n) and write $a_n \to l$ as $n \to \infty$, or simply $a_n \to l$.

A sequence that does not converge to any real number is **divergent**.

We write $a_n \not\to l$ as $n \to \infty$ if either (a_n) is divergent or it converges to some number other than l.

Convergent real sequences have unique limits.

Divergent sequences

- (a_n) is a *divergent sequence* if for each $l \in \mathbb{R}$, there is an $\varepsilon > 0$ such that for each $N \in \mathbb{N}$, there is an $n > N$ with $|a_n - l| \geq \varepsilon$.
- If (a_n) is a real null sequence, then $(1/a_n)$ is divergent.

Theorem 1.19 *Combination rules for convergent sequences in* $\mathbb{R}$

If $a_n \to l$ and $b_n \to m$ as $n \to \infty$, then:

Sum Rule	$a_n + b_n \to l + m$ as $n \to \infty$
Multiple Rule	$\lambda a_n \to \lambda l$ as $n \to \infty$, for any $\lambda \in \mathbb{R}$
Product Rule	$a_n b_n \to lm$ as $n \to \infty$
Quotient Rule	$a_n/b_n \to l/m$ as $n \to \infty$, provided that $m \neq 0$
Squeeze Rule	if $l = m$ and $a_n \leq c_n \leq b_n$ for each n, then $c_n \to l$ as $n \to \infty$.

Definition 1.21 *Upper and lower bounds*

Let A be a subset of real numbers. Then $b_l \in \mathbb{R}$ is a **lower bound** of A if for each $a \in A$, $b_l \leq a$.

Similarly, $b_u \in \mathbb{R}$ is an **upper bound** of A if for each $a \in A$, $a \leq b_u$.

Theorem 1.22 *Monotone Convergence Theorem*

If (a_n) is a monotone increasing sequence that is bounded above by a real number b_u, then (a_n) is convergent with limit $l \leq b_u$.

If (a_n) is a monotone decreasing sequence that is bounded below by a real number b_l, then (a_n) is convergent with limit $l \geq b_l$.

Definition 1.24 *Subsequence*

The sequence $(a_{n_k})_{k=1}^{\infty}$ is a **subsequence** of the sequence (a_n) if $(n_k)_{k=1}^{\infty}$ is a strictly increasing sequence of positive integers, i.e. if $1 \leq n_1 < n_2 < n_3 < \cdots$.

In particular:

- when $n_k = 2k$, we have $(a_{2k})_{k=1}^{\infty}$, which is the **even subsequence** of (a_n)
- when $n_k = 2k-1$, we have $(a_{2k-1})_{k=1}^{\infty}$, which is the **odd subsequence** of (a_n).

Proposition 1.25 *Convergence of subsequences*

If (a_n) is a convergent sequence with limit l, then every subsequence (a_{n_k}) is convergent with the same limit l.

Continuity on the real line

Definition 2.1 *Continuity at a point*

Let $A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $a \in A$ if:

> whenever (x_n) is a sequence in A for which $x_n \to a$ as $n \to \infty$,
> then the sequence $(f(x_n))$ converges to $f(a)$.

If f does not satisfy this condition at $a \in A$ – that is, there is a sequence (x_n) in A for which $x_n \to a$ as $n \to \infty$ but $f(x_n)$ does not converge to $f(a)$ – then we say that f is **discontinuous** at a.

We write $f(x_n) \to f(a)$ as $x_n \to a$ if the sequence $(f(x_n))$ converges to $f(a)$ for a *particular* sequence (x_n) that converges to a (in the domain A), and $f(x) \to f(a)$ as $x \to a$ if the sequence $(f(x_n))$ converges to $f(a)$ for *every* sequence (x_n) that converges to a (in the domain A).

Definition 2.2 *Continuity on a set*

Let $S \subseteq A \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Combination rules for continuous functions from $\mathbb{R}$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}$ and suppose $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ are continuous at $a \in A$.

Then the following functions are also continuous at a:

Sum Rule	$f+g\colon A \to \mathbb{R}$, defined by $(f+g)(x) = f(x) + g(x)$
Multiple Rule	$\lambda f\colon A \to \mathbb{R}$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$
Product Rule	$fg\colon A \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$
Quotient Rule	$f/g\colon A - \{x : g(x) = 0\} \to \mathbb{R}$, defined by $(f/g)(x) = f(x)/g(x)$, provided $g(a) \neq 0$.

Definition 2.7 *Restricted function*

If the domain of a function $f\colon A \to B$ is restricted to a set C, where $C \subseteq A$, then the resulting **restricted function**, known as the **restriction of f to C**, is denoted by $f|_C\colon C \to B$, and given by $f|_C(x) = f(x)$ for each $x \in C$.

Proposition 2.8 *Restriction Rule for continuous functions on $\mathbb{R}$*

Let $A \subseteq \mathbb{R}$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $b \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at b.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

Definition 2.9 *Composition*

If $f\colon A \to C$ and $g\colon B \to D$ are functions with $f(A) \subseteq B$, then the **composed function** or **composition** $g \circ f\colon A \to D$ is given by $(g \circ f)(x) = g(f(x))$.

Proposition 2.10 *Composition Rule*

Let A, $B \subseteq \mathbb{R}$ and let $f\colon A \to \mathbb{R}$ and $g\colon B \to \mathbb{R}$ be continuous on A and B respectively, with $f(A) \subseteq B$. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Basic continuous functions

The following functions are continuous:

- polynomials and rational functions
- $f(x) = |x|$
- $f(x) = \sqrt{x}$
- the trigonometric functions (sine, cosine and tangent) and their inverses (restricted to appropriate domains)
- the logarithmic function
- the exponential function.

Theorem 2.12 *Intermediate Value Theorem*

Let $f\colon [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ with $f(a) \neq f(b)$ and let k be any number lying between $f(a)$ and $f(b)$. Then there exists a number $c \in (a, b)$ such that $f(c) = k$.

Definition 2.14 *Boundedness*

A set $A \subset \mathbb{R}$ is **bounded** if there exists a real number M such that $|x| \leq M$ for each $x \in A$.

A function $f\colon A \to \mathbb{R}$ is **bounded** on A if there is a real number M such that $|f(x)| \leq M$ for each $x \in A$.

Theorem 2.15 *Boundedness Theorem*

Let $f\colon [a, b] \to \mathbb{R}$ be a continuous function on $[a, b]$. Then f is bounded on $[a, b]$. That is, there is a real number M such that $|f(x)| \leq M$ for each $x \in [a, b]$.

Theorem 2.16 *Extreme Value Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a continuous function on $[a,b]$. Then there are numbers $c, d \in [a,b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for each } x \in [a,b].$$

The *minimum* and *maximum* value of the function in the interval are $f(c)$ and $f(d)$, respectively.

Continuity on the plane

Definition 3.1 *Sequence in* $\mathbb{R}^2$

A **sequence in** $\mathbb{R}^2$ is an unending ordered list of points in $\mathbb{R}^2$:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots.$$

The point $\mathbf{a}_n = (a_{1,n}, a_{2,n})$ is the ***n*th term** of the sequence, and the whole sequence is denoted by $(\mathbf{a}_n)$, $(\mathbf{a}_n)_{n=1}^{\infty}$ or $(\mathbf{a}_n)_{n\in\mathbb{N}}$.

Definition 3.3 *Euclidean distance in* $\mathbb{R}^2$

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2)$ and $\mathbf{b} = (b_1, b_2)$ in $\mathbb{R}^2$ is given by the formula

$$d^{(2)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2}.$$

Definition 3.4 *Convergence in* $\mathbb{R}^2$

A sequence $(\mathbf{a}_n)$ in $\mathbb{R}^2$ **converges** to $\boldsymbol{\ell} \in \mathbb{R}^2$ if $\left(d^{(2)}(\mathbf{a}_n, \boldsymbol{\ell})\right)$ is a (real) null sequence.

We say that $\boldsymbol{\ell}$ is the **limit** of the sequence $(\mathbf{a}_n)$ and write $\mathbf{a}_n \to \boldsymbol{\ell}$ as $n \to \infty$, $\lim_{n\to\infty} \mathbf{a}_n = \boldsymbol{\ell}$ or simply $\mathbf{a}_n \to \boldsymbol{\ell}$.

A sequence in the plane that does not converge to any point in $\mathbb{R}^2$ is said to be **divergent**.

Note that $\mathbf{a}_n \to \boldsymbol{\ell}$ is equivalent to $(\mathbf{a}_n - \boldsymbol{\ell}) \to \mathbf{0}$.

Theorem 3.7

Let $(\mathbf{a}_n) = ((a_{1,n}, a_{2,n}))$ be a sequence in $\mathbb{R}^2$ and let $\boldsymbol{\ell} = (\ell_1, \ell_2) \in \mathbb{R}^2$. Then

$$\mathbf{a}_n \to \boldsymbol{\ell} \quad \text{if, and only if, both } a_{1,n} \to \ell_1 \text{ and } a_{2,n} \to \ell_2 \text{ as } n \to \infty.$$

A sequence $(\mathbf{a}_n)$ in the plane is divergent if, and only if, at least one of the component sequences $(\mathbf{a}_{1,n})$ and $(\mathbf{a}_{2,n})$ is divergent.

Definition 3.8 *Continuity at a point in the plane*

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

Then f is **continuous** at $\mathbf{a} \in A$ if:

whenever $(\mathbf{x}_n)$ is a sequence in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$,
then the (real) sequence $(f(\mathbf{x}_n))$ converges to $f(\mathbf{a})$.

If f does not satisfy this condition at $\mathbf{a} \in A$ – that is, there is a sequence $(\mathbf{x}_n)$ in A for which $\mathbf{x}_n \to \mathbf{a}$ as $n \to \infty$ but $f(\mathbf{x}_n)$ does not converge to $f(\mathbf{a})$ – then we say that f is **discontinuous** at $\mathbf{a}$.

Definition 3.9 *Continuity on a subset in the plane*

Let $S \subseteq A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Definition 3.10 *Projection functions in the plane*

The two functions $p_1\colon \mathbb{R}^2 \to \mathbb{R}$ and $p_2\colon \mathbb{R}^2 \to \mathbb{R}$ given by

$$p_1(x_1, x_2) = x_1, \quad p_2(x_1, x_2) = x_2,$$

are known as **projection functions**.

Continuity of projection functions

The projection functions p_1 and p_2 are continuous on $\mathbb{R}^2$.

Combination rules for continuous functions from $\mathbb{R}^2$ to $\mathbb{R}$

Let $A \subseteq \mathbb{R}^2$ and let $f\colon A \to \mathbb{R}$ and $g\colon A \to \mathbb{R}$ be functions that are continuous at $\mathbf{a} \in A$.

Then the following functions are continuous at $\mathbf{a}$:

Sum Rule	$f + g$, defined by $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$
Multiple Rule	λf where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$
Product Rule	fg, defined by $(fg)(\mathbf{x}) = f(\mathbf{x})g(\mathbf{x})$
Quotient Rule	f/g, defined by $(f/g)(\mathbf{x}) = f(\mathbf{x})/g(\mathbf{x})$, provided $g(\mathbf{a}) \neq 0$.

Proposition 3.15 *Restriction Rule*

Let $A \subseteq \mathbb{R}^2$, let $f\colon A \to \mathbb{R}$ and let $B \subseteq A$. If f is continuous at $\mathbf{b} \in B$, then the restricted function $f|_B\colon B \to \mathbb{R}$ is continuous at $\mathbf{b}$.

In particular, if f is continuous on A, then $f|_B$ is continuous on B.

Proposition 3.16 *Composition Rule*

Let $A \subseteq \mathbb{R}^2$ and suppose that $f\colon A \to \mathbb{R}$ is continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}$ and suppose $g\colon B \to \mathbb{R}$ is continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}$ is continuous on A.

Properties of Euclidean distance on $\mathbb{R}^2$

For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^2$:

non-negativity	(M1)	$d^{(2)}(\mathbf{a}, \mathbf{b}) \geq 0$, with $d^{(2)}(\mathbf{a}, \mathbf{b}) = 0$ if, and only if, $\mathbf{a} = \mathbf{b}$
symmetry	(M2)	$d^{(2)}(\mathbf{a}, \mathbf{b}) = d^{(2)}(\mathbf{b}, \mathbf{a})$
Triangle Inequality	(M3)	$d^{(2)}(\mathbf{a}, \mathbf{c}) \leq d^{(2)}(\mathbf{a}, \mathbf{b}) + d^{(2)}(\mathbf{b}, \mathbf{c})$.

Properties of Euclidean distance on $\mathbb{R}$

For each $a, b, c \in \mathbb{R}$:

(M1) $|b-a| \geq 0$, with $|b-a| = 0$ if, and only if, $b = a$ — non-negativity

(M2) $|b-a| = |a-b|$ — symmetry

(M3) $|c-a| \leq |b-a| + |c-b|$. — Triangle Inequality

Chapter 14: Metric spaces and continuity 1

Continuity of functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Definition 1.1 *Product set*

Let $A_1, A_2, \ldots, A_k$ be sets. The **product set** $A_1 \times A_2 \times \cdots \times A_k$ is defined to be the set consisting of all ordered k-tuples $(a_1, a_2, \ldots, a_k)$ where $a_1 \in A_1$, $a_2 \in A_2$, …, $a_k \in A_k$. Thus

$$A_1 \times A_2 \times \cdots \times A_k = \{(a_1, a_2, \ldots, a_k) : a_i \in A_i \text{ for } i = 1, \ldots, k\}.$$

If $A_1 = A_2 = \cdots = A_k = A$, then the product set $A_1 \times A_2 \times \cdots \times A_k$ is written simply as A^k and

$$A^k = \{(a_1, a_2, \ldots, a_k) : a_i \in A \text{ for } i = 1, \ldots, k\}.$$

We denote an element of a product set $A_1 \times A_2 \times \cdots \times A_k$ as $\mathbf{a} = (a_1, a_2, \ldots, a_k)$.

Two elements $\mathbf{a} = (a_1, a_2, \ldots, a_k)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_k)$ are equal if $a_i = b_i$ for $i = 1, 2, \ldots, k$.

Definition 1.2 *Euclidean distance on $\mathbb{R}^n$*

The **Euclidean distance** between points $\mathbf{a} = (a_1, a_2, \ldots, a_n)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_n)$ in $\mathbb{R}^n$ is given by the formula

$$d^{(n)}(\mathbf{a}, \mathbf{b}) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \cdots + (b_n - a_n)^2}.$$

When it is clear from the context which dimension of Euclidean space we are working with, we may simply write $d(\mathbf{a}, \mathbf{b})$ for this distance.

Note that $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{a} - \mathbf{b},\ \mathbf{0}) = d^{(n)}(\mathbf{b} - \mathbf{a}, \mathbf{0})$.

Proposition 1.3 *Properties of Euclidean distance on $\mathbb{R}^n$*

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following properties.

For each $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$:

(M1) $d^{(n)}(\mathbf{a}, \mathbf{b}) \geq 0$, with equality holding if, and only if, $\mathbf{a} = \mathbf{b}$ — non-negativity

(M2) $d^{(n)}(\mathbf{a}, \mathbf{b}) = d^{(n)}(\mathbf{b}, \mathbf{a})$ — symmetry

(M3) $d^{(n)}(\mathbf{a}, \mathbf{c}) \leq d^{(n)}(\mathbf{a}, \mathbf{b}) + d^{(n)}(\mathbf{b}, \mathbf{c})$. — Triangle Inequality

Proposition 1.4 *Reverse Triangle Inequality for* $\mathbb{R}^n$

The Euclidean distance function $d^{(n)}$ on $\mathbb{R}^n$ has the following property.

(M3a) If $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{R}^n$, then $d^{(n)}(\mathbf{b}, \mathbf{c}) \geq \left| d^{(n)}(\mathbf{a}, \mathbf{c}) - d^{(n)}(\mathbf{a}, \mathbf{b}) \right|$.

Definition 1.5 *Sequence in* $\mathbb{R}^n$

A **sequence** in $\mathbb{R}^n$ is an unending ordered list of points in $\mathbb{R}^n$:

$$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \ldots.$$

The point $\mathbf{a}_k = (a_{1,k}, a_{2,k}, \ldots, a_{n,k})$ is the ***k*th term** of the sequence, and the whole sequence is denoted by $(\mathbf{a}_k)$, $(\mathbf{a}_k)_{k=1}^{\infty}$ or $(\mathbf{a}_k)_{k \in \mathbb{N}}$.

Definition 1.6 *Convergence in* $\mathbb{R}^n$

A sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ **converges** to $\boldsymbol{\ell} \in \mathbb{R}^n$ if $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)$ is a (real) null sequence.

We say that $\boldsymbol{\ell}$ is the **limit** of the sequence $(\mathbf{a}_k)$ and write $\mathbf{a}_k \to \boldsymbol{\ell}$ as $k \to \infty$, $\lim_{k \to \infty} \mathbf{a}_k = \boldsymbol{\ell}$ or simply $\mathbf{a}_k \to \boldsymbol{\ell}$.

A sequence in $\mathbb{R}^n$ that does not converge to any point in $\mathbb{R}^n$ is said to be **divergent**.

Note that $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)_{k \in \mathbb{N}}$ is a sequence of real numbers. Hence, testing whether a given sequence $(\mathbf{a}_k)$ in $\mathbb{R}^n$ converges to $\boldsymbol{\ell} \in \mathbb{R}^n$ reduces to testing whether $\left(d^{(n)}(\mathbf{a}_k, \boldsymbol{\ell})\right)$ is a *real* null sequence.

Lemma 1.7 *Uniqueness of limits in* $\mathbb{R}^n$

Let $(\mathbf{a}_k)$ be a convergent sequence in $\mathbb{R}^n$. If $\mathbf{a}_k \to \boldsymbol{r}$ and $\mathbf{a}_k \to \boldsymbol{s}$ as $k \to \infty$, then $\boldsymbol{r} = \boldsymbol{s}$.

Proposition 1.9

Let $(\mathbf{a}_k)$ be a sequence in $\mathbb{R}^n$, with $\mathbf{a}_k = (a_{1,k}, \ldots, a_{n,k})$, and let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_n) \in \mathbb{R}^n$. Then

$$\mathbf{a}_k \to \boldsymbol{\ell} \quad \text{as } k \to \infty$$

if, and only if,

$$a_{j,k} \to \ell_j \quad \text{as } k \to \infty, \text{ for each } j = 1, 2, \ldots, n.$$

Combination rules for convergent sequences in $\mathbb{R}^n$

If $\mathbf{a}_k \to \boldsymbol{\ell}$ and $\mathbf{b}_k \to \mathbf{m}$ as $k \to \infty$, then:

Sum Rule $\mathbf{a}_k + \mathbf{b}_k \to \boldsymbol{\ell} + \mathbf{m}$ as $k \to \infty$

Multiple Rule $\lambda \mathbf{a}_k \to \lambda \boldsymbol{\ell}$ as $k \to \infty$, for any $\lambda \in \mathbb{R}$.

Definition 1.11 *Continuity from $\mathbb{R}^n$ to $\mathbb{R}^m$*

Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a function.

Then f is **continuous** at $\mathbf{a} \in A$ if:

> whenever $(\mathbf{x}_k)$ is a sequence in A for which $\mathbf{x}_k \to \mathbf{a}$ as $k \to \infty$,
> then the sequence $(f(\mathbf{x}_k))$ converges to $f(\mathbf{a})$.

If f does not satisfy this condition at $\mathbf{a} \in A$ – that is, there is a sequence $(\mathbf{x}_k)$ in A for which $\mathbf{x}_k \to \mathbf{a}$ as $k \to \infty$ but $f(\mathbf{x}_k)$ does not converge to $f(\mathbf{a})$ – then we say that f is **discontinuous** at $\mathbf{a}$.

Definition 1.12 *Continuity on a set in $\mathbb{R}^n$*

Let $S \subseteq A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be a function.

We say that f is **continuous** on S when f is continuous at each point in S.

We say that f is **discontinuous** on S if f is discontinuous at at least one point in S.

Combination rules for continuous functions from $\mathbb{R}^n$ to $\mathbb{R}^m$

Let $A \subseteq \mathbb{R}^n$, and let $f\colon A \to \mathbb{R}^m$ and $g\colon A \to \mathbb{R}^m$ be continuous on A.

Then the following functions are continuous on A:

Sum Rule $f + g\colon A \to \mathbb{R}^m$, defined by $(f+g)(\mathbf{x}) = f(\mathbf{x}) + g(\mathbf{x})$

Multiple Rule $\lambda f\colon A \to \mathbb{R}^m$ where $\lambda \in \mathbb{R}$, defined by $(\lambda f)(\mathbf{x}) = \lambda \times f(\mathbf{x})$.

Proposition 1.16 *Restriction Rule*

Let $A \subseteq \mathbb{R}^n$, let $f\colon A \to \mathbb{R}^m$ be continuous on A and let $B \subseteq A$. Then the restricted function $f|_B\colon B \to \mathbb{R}^m$ defined by $f|_B(\mathbf{x}) = f(\mathbf{x})$ for $\mathbf{x} \in B$ is continuous on B.

Proposition 1.17 *Composition Rule*

Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$ be continuous on A. Let $f(A) \subseteq B \subseteq \mathbb{R}^m$ and let $g\colon B \to \mathbb{R}^k$ be continuous on B. Then the composed function $g \circ f\colon A \to \mathbb{R}^k$ is continuous on A.

Definition 1.18 *Projection functions on $\mathbb{R}^n$*

The n functions $p_j\colon \mathbb{R}^n \to \mathbb{R}$ given by

$$p_j(x_1, x_2, \ldots, x_n) = x_j, \quad \text{for } j = 1, 2, \ldots, n,$$

are known as **projection functions**.

Continuity of projection functions

The projection functions $p_1, p_2, \ldots, p_n$ are continuous on $\mathbb{R}^n$.

Proposition 1.19

Let $A \subseteq \mathbb{R}^n$ and let $f\colon A \to \mathbb{R}^m$.

The function f is continuous at $\mathbf{a} \in A$ if, and only if, for each $j = 1, 2, \ldots, m$, the function $p_j \circ f\colon A \to \mathbb{R}$ is continuous at $\mathbf{a}$, where $p_j\colon \mathbb{R}^m \to \mathbb{R}$ is the jth projection function.

Theorem 1.21 *Cauchy–Schwarz Inequality*

Let $(r_1, r_2, \ldots, r_n)$ and $(s_1, s_2, \ldots, s_n)$ be points in $\mathbb{R}^n$. Then

$$\left(\sum_{j=1}^{n} r_j s_j\right)^2 \leq \sum_{j=1}^{n} r_j^2 \sum_{j=1}^{n} s_j^2,$$

with equality if, and only if, there is $\lambda \in \mathbb{R}$ for which $s_j = \lambda r_j$ for each j.

The square-rooted version of this result is also known as the Cauchy–Schwarz Inequality.

Introducing metric spaces

Definition 2.1 *Metric*

Let X be a set. A **metric** on X is a function $d\colon X \times X \to \mathbb{R}$ that satisfies the following three conditions.

For each $a, b, c \in X$:

non-negativity (M1) $d(a, b) \geq 0$, with equality holding if, and only if, $a = b$

symmetry (M2) $d(a, b) = d(b, a)$

Triangle Inequality (M3) $d(a, c) \leq d(a, b) + d(b, c)$.

The set X, together with a metric d on X, is called a **metric space**, and is denoted by (X, d).

Proposition 2.2 *Reverse Triangle Inequality for metric spaces*

Let (X, d) be a metric space. For each $a, b, c \in X$,
(M3a) $d(b, c) \geq |d(a, c) - d(a, b)|$.

Definition 2.3 *Discrete metric*

Let X be a set. The **discrete metric** on X is the function $d_0\colon X \times X \to \mathbb{R}$ defined by

$$d_0(a, b) = \begin{cases} 0, & \text{if } a = b, \\ 1, & \text{if } a \neq b. \end{cases}$$

Proposition 2.4

Let X be a set. Then (X, d_0) is a metric space.

Definition 2.6 *Open and closed balls*

Let (X, d) be a metric space, and let $a \in X$ and $r \geq 0$.

The **open ball** of **radius** r **with centre** a is the set

$$B_d(a, r) = \{x \in X : d(a, x) < r\}.$$

The **closed ball** of **radius** r **with centre** a is the set

$$B_d[a, r] = \{x \in X : d(a, x) \leq r\}.$$

The **sphere** of **radius** r **with centre** a is the set

$$S_d(a, r) = \{x \in X : d(a, x) = r\}.$$

When $r = 1$, these sets are called respectively the **unit open ball** with **centre** a, the **unit closed ball** with **centre** a and the **unit sphere** with **centre** a.

Sequences in metric spaces

Definition 3.1 *Sequence in a metric space*

Let X be a set. A **sequence** in X is an unending ordered list of elements of X:

$$a_1, a_2, a_3, \ldots.$$

The element a_k is the ***k*th term** of the sequence, and the whole sequence is denoted by (a_k), $(a_k)_{k=1}^{\infty}$ or $(a_k)_{k\in\mathbb{N}}$.

Definition 3.2 *Convergence in a metric space*

Let (X, d) be a metric space. A sequence (a_k) in X **d-converges** to $a \in X$ if the sequence of real numbers $(d(a_k, a))$ is a null sequence.

We write $a_k \xrightarrow{d} a$ as $k \to \infty$, or simply $a_k \to a$ if the context is clear.

We say that the sequence (a_k) is **convergent** in (X, d) with **limit** a.

A sequence that does not converge (with respect to the metric d) to any point in X is said to be **d-divergent**.

Theorem 3.3 *Uniqueness of limits in a metric space*

Let (X, d) be a metric space and let $a, b \in X$. If (a_k) is a sequence in X that d-converges to both a and b, then $a = b$.

Definition 3.4 *Eventually constant*

Let (a_k) be a sequence in a set X. We say that (a_k) is **eventually constant** if there is $a \in X$ and $N \in \mathbb{N}$ such that $a_k = a$ whenever $k > N$.

Eventually constant sequences are always convergent, no matter the choice of metric.

The definition of continuity in metric spaces

Definition 4.1 *Continuity for metric spaces*

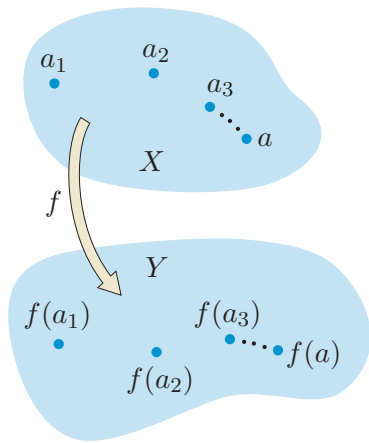
Let (X, d) and (Y, e) be metric spaces and let $f\colon X \to Y$ be a function.

Then f is **(d, e)-continuous** at $a \in X$ if:

> whenever (a_k) is a sequence in X for which $a_k \xrightarrow{d} a$ as $k \to \infty$,
> then the sequence $f(a_k) \xrightarrow{e} f(a)$ as $k \to \infty$.

If f does not satisfy this condition at some $a \in X$ – that is, there is a sequence (x_k) in X for which $x_k \to a$ as $k \to \infty$ but $f(x_k)$ does not converge to $f(a)$ – then we say that f is **(d, e)-discontinuous** at a.

A function that is continuous at all points of X is said to be **(d, e)-continuous** on X (or simply **continuous**, if no ambiguity is possible).



Proposition 4.4 *Composition Rule*

Let (X, d_X), (Y, d_Y) and (Z, d_Z) be metric spaces. Let $f\colon X \to Y$ be (d_X, d_Y)-continuous and let $g\colon Y \to Z$ be (d_Y, d_Z)-continuous. Then the composed function $g \circ f\colon X \to Z$ is (d_X, d_Z)-continuous.

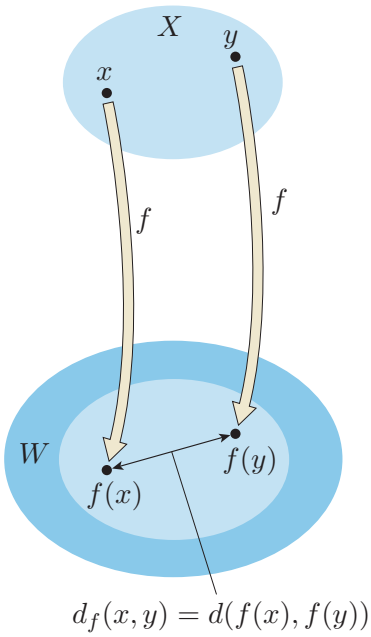
Chapter 15: Metric spaces and continuity 2

New metrics from old

Definition 1.1 *Pull-back*

Let $f\colon X \to W$ be a function and let d be a metric on W. The **pull-back of d by f** is the function $d_f\colon X \times X \to \mathbb{R}$ given by

$$d_f(x, y) = d(f(x), f(y)), \text{ for each } x, y \in X.$$



Defining the pull-back

Theorem 1.2

Let $f\colon X \to W$ be a one–one function and let d be a metric on W. Then the pull-back of d by f, d_f, is a metric on X.

Definition 1.4 *Induced metric*

Let (X, d) be a metric space and let $A \subseteq X$.

The function $d_A\colon A \times A \to \mathbb{R}$ given by

$$d_A(a, b) = d(a, b), \quad \text{for each } a, b \in A,$$

is a metric on A and is called the metric on A **induced by** d, often referred to as the **induced** (or **subspace**) **metric** on A.

The metric space (A, d_A) is called a **metric subspace** of (X, d), and often referred to as the **induced subspace**.

Theorem 1.5

Let (X, d) be a metric space and suppose that $A \subseteq X$. Let d_A denote the metric on A induced by d. Then a sequence $(a_n)_{n\in\mathbb{N}}$ in A is d_A-convergent to a if, and only if, $a \in A$ and the sequence (a_n) is d-convergent to a (when viewed as a sequence in X).

Theorem 1.6 *Restriction Rule*

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$ is a function. Let $A \subseteq X$ and let d_A be the metric induced on A by d.

If $a \in A$ and f is (d, e)-continuous at a, then $f|_A$ is (d_A, e)-continuous at a.

(Here the **restriction of f to A**, $f|_A\colon A \to Y$, is given by $f|_A(a) = f(a)$ for each $a \in A$.)

Theorem 1.7

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. The functions e_1, e_2 and e_∞ from $X \times X$ to $\mathbb{R}$, given by

$$\begin{aligned} e_1(\mathbf{x}, \mathbf{y}) &= d_1(x_1, y_1) + d_2(x_2, y_2), \\ e_2(\mathbf{x}, \mathbf{y}) &= \sqrt{(d_1(x_1, y_1))^2 + (d_2(x_2, y_2))^2}, \\ e_\infty(\mathbf{x}, \mathbf{y}) &= \max\{d_1(x_1, y_1), d_2(x_2, y_2)\}, \end{aligned}$$

for each $\mathbf{x} = (x_1, x_2),\ \mathbf{y} = (y_1, y_2) \in X$, are metrics on X.

More generally, $(e_p(\mathbf{x}, \mathbf{y}) = (d_1(x_1, y_1)^p + d_2(x_2, y_2)^p)^{1/p}$, is a metric on X for $p \geq 1$.

These ideas extend naturally to n-fold products of metric spaces.

Hamming metric

Let A be a finite set, called the **set of symbols**, and let d_0 be the discrete metric for A.

Fix $n \in \mathbb{N}$ and let $X = A^n$, **the words of length n over the set of symbols A.**

Define the Hamming distance $d_H \colon X \times X \to \mathbb{R}$ for $\mathbf{a}, \mathbf{b} \in X$ by

$$\begin{aligned} d_H(\mathbf{a}, \mathbf{b}) &= \text{the number of places at which } \mathbf{a} \text{ and } \mathbf{b} \text{ differ} \\ &= d_0(a_1, b_1) + d_0(a_2, b_2) + \cdots + d_0(a_n, b_n). \end{aligned}$$

Then d_H defines a metric on X, called the **Hamming metric**.

The Cantor metric

Definition 2.1 *Cantor space*

The **Cantor space $\mathbf{C}$** consists of all *infinite* sequences of zeros and ones:

$$\mathbf{C} = \{(a_n) : a_n \in \{0, 1\} \text{ for each } n \in \mathbb{N}\}.$$

Definition 2.2 *Cantor distance*

Let $\mathbf{x}, \mathbf{y} \in \mathbf{C}$. The **Cantor distance** $d_\mathbf{C} \colon \mathbf{C} \times \mathbf{C} \to \mathbb{R}$ between $\mathbf{x}$ and $\mathbf{y}$ is given by

$$d_\mathbf{C}(\mathbf{x}, \mathbf{y}) = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n}, & \text{if } \mathbf{x} \text{ and } \mathbf{y} \text{ first differ at their } n\text{th term.} \end{cases}$$

Theorem 2.3

The Cantor distance is a metric on $\mathbf{C}$.

Balls for the Cantor metric

The open ball around $\mathbf{a}$ of radius 2^{-n} for the Cantor metric d is

$$B_d(\mathbf{a}, 2^{-n}) = \{(a_1, a_2, a_3, \ldots, a_n, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\}.$$

The closed ball around $\mathbf{a}$ of radius 2^{-n} for the Cantor metric d is

$$B_d[\mathbf{a}, 2^{-n}] = \begin{cases} \mathbf{C}, & \text{if } n = 1, \\ \{(a_1, a_2, a_3, \ldots, a_{n-1}, c_1, c_2, c_3, \ldots) : (c_i) \in \mathbf{C}\}, & \text{if } n \geq 2. \end{cases}$$

Definition 2.5 *Shift map*

The **shift map (on C)** is the map $\sigma\colon \mathbf{C} \to \mathbf{C}$ given by

$$\sigma((x_1, x_2, x_3, x_4, \ldots)) = (x_2, x_3, x_4, x_5, \ldots).$$

The shift map is $(d_{\mathbf{C}}, d_{\mathbf{C}})$-continuous on $\mathbf{C}$.

Equivalent metrics

Definition 3.1 *Lipschitz function*

Let (X, d) and (Y, e) be metric spaces.

A function $f\colon X \to Y$ is a **Lipschitz function** if there is a non-negative real number M such that, for each $a, b \in X$,

$$e(f(a), f(b)) \leq M d(a, b).$$

The real number M is known as a **Lipschitz constant for** f.

When we wish to emphasise the metrics d and e on X and Y, we say that f is a (d, e)-Lipschitz function.

Theorem 3.2

Let (X, d) and (Y, e) be metric spaces.

If a function $f\colon X \to Y$ is (d, e)-Lipschitz, then it is (d, e)-continuous.

Definition 3.5 *Equivalent metrics*

Let d_1 and d_2 be two metrics defined on a set X. The metrics d_1 and d_2 are (metrically) **equivalent** if there are positive real numbers m and M such that for each $x, y \in X$,

$$m d_1(x, y) \leq d_2(x, y) \leq M d_1(x, y).$$

Theorem 3.6

Let (X_1, d_1) and (X_2, d_2) be metric spaces, and let $X = X_1 \times X_2$. Then the three product metrics e_1, e_2 and e_∞ formed from d_1 and d_2 are equivalent.

Proposition 3.7

Let d_1 and d_2 be equivalent metrics defined on the set X. Then a sequence (x_n) converges with limit x in (X, d_1) if, and only if, the same sequence (x_n) converges with the same limit x in (X, d_2).

Theorem 3.8

Let (Y, e) be a metric space and let d_1 and d_2 be metrically equivalent metrics on a set X.

Then

(a) a function $f\colon X \to Y$ is (d_1, e)-continuous at a point $x \in X$ if, and only if, f is (d_2, e)-continuous at x; and

(b) a function $g\colon Y \to X$ is (e, d_1)-continuous at a point $y \in Y$ if, and only if, g is (e, d_2)-continuous at y.

Theorem 3.9

Let (X_1, d_1), (X_2, d_2) and (Y, e) be metric spaces, let $X = X_1 \times X_2$ and let e_1, e_2 and e_∞ be the three different product metrics on X (determined by d_1 and d_2).

If $f\colon X \to Y$ then, for $j, k \in \{1, 2, \infty\}$, f is (e_j, e)-continuous at $x \in X$ if, and only if, f is (e_k, e)-continuous at x.

If $g\colon Y \to X$ then, for $j, k \in \{1, 2, \infty\}$, g is (e, e_j)-continuous at $y \in Y$ if, and only if, g is (e, e_k)-continuous at y.

A similar result holds for the n-fold product of metric spaces.

Definition 3.10 *Projection functions on a product space*

Let X_1 and X_2 be sets and let $X = X_1 \times X_2$. The **projection functions** are the functions $p_1\colon X \to X_1$ and $p_2\colon X \to X_2$ given by

$$p_1(x_1, x_2) = x_1 \text{ and } p_2(x_1, x_2) = x_2.$$

Theorem 3.11

Let (X_1, d_1) and (X_2, d_2) be metric spaces and let e be one of the product metrics e_1, e_2 or e_∞ for $X_1 \times X_2$.

Then p_1 is (e, d_1)-continuous and p_2 is (e, d_2)-continuous, where p_1 and p_2 are the projection functions from X to X_1 and X_2 respectively.

Theorem 3.12

Let (Y, d), (X_1, d_1) and (X_2, d_2) be metric spaces, and let (X, e) be the metric space where $X = X_1 \times X_2$ and e is one of the three product metrics e_1, e_2 or e_∞ formed from d_1 and d_2.

Then a function $f\colon Y \to X_1 \times X_2$ is (d, e)-continuous at $a \in Y$ if, and only if, $p_1 \circ f\colon Y \to X_1$ is (d, d_1)-continuous at a and $p_2 \circ f\colon Y \to X_2$ is (d, d_2)-continuous at a.

A similar result holds for the n-fold product of metric spaces.

Definition 3.14 *Component sequences*

Let X and Y be sets and let $((x_n, y_n))$ be a sequence in $X \times Y$. The sequences (x_n) in X and (y_n) in Y are the **component sequences** or **coordinate sequences** of the **product sequence** $((x_n, y_n))$.

Theorem 3.15

Let (X, d_1) and (Y, d_2) be metric spaces and let e be one of the three product metrics e_1, e_2 and e_∞ for $X \times Y$.

Suppose that $(x, y) \in X \times Y$ and $((x_n, y_n))$ is a sequence in $X \times Y$. Then

$$(x_n, y_n) \xrightarrow{e} (x, y) \quad \text{if, and only if,} \quad x_n \xrightarrow{d_1} x \text{ and } y_n \xrightarrow{d_2} y \text{ as } n \to \infty.$$

Spaces of functions

Definition 4.1 *Pointwise convergence*

Let A be a set.

A sequence (f_n) of functions $f_n\colon A \to \mathbb{R}$ **converges pointwise** to the function $f\colon A \to \mathbb{R}$ if, for each $x \in A$, the sequence of real numbers $(f_n(x))$ converges to $f(x)$.

If a sequence (f_n) of functions converges pointwise to a function f, then f is called the **pointwise limit** of the sequence (f_n).

We write $f_n \to f$ pointwise as $n \to \infty$, $f_n \to_p f$ as $n \to \infty$ or simply $f_n \to f$ pointwise.

The pointwise limit of a sequence of continuous functions need not be continuous.

Definition 4.3 *Bounded function*

Let A be a set. A function $f\colon A \to \mathbb{R}$ is **bounded** on A if there is an $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in A$.

Definition 4.4 *Uniform convergence*

Let A be a set. A sequence (f_n) of functions $f_n\colon A \to \mathbb{R}$ **converges uniformly** on A to the function $f\colon A \to \mathbb{R}$ if

(a) the function $f_n - f$ is bounded for each $n \in \mathbb{N}$, and

(b) the sequence (M_n) defined by

$$M_n = \sup\{|f_n(x) - f(x)| : x \in A\}$$

is a (real) null sequence.

Such a function f is called the **uniform limit** of the sequence (f_n).

We write either $f_n \to f$ uniformly as $n \to \infty$, $f_n \rightrightarrows f$ as $n \to \infty$ or simply $f_n \to f$ uniformly.

Lemma 4.5

Let A be a set and let (f_n) be a sequence of functions $f_n\colon A \to \mathbb{R}$.

If the sequence (f_n) converges uniformly on A to the function $f\colon A \to \mathbb{R}$, then (f_n) also converges pointwise to f.

Even when a sequence converges pointwise to a *continuous* function, the convergence need not be uniform.

Theorem 4.7

Let (X, d) be a metric space and fix $a \in X$.

For each $n \in \mathbb{N}$, suppose that $f_n\colon X \to \mathbb{R}$ is $(d, d^{(1)})$-continuous at a.

If the sequence of functions (f_n) converges uniformly on X to a function $f\colon X \to \mathbb{R}$, then f is $(d, d^{(1)})$-continuous at a.

The max metric on $C[0,1]$

Notation

We let $C[0,1]$ denote the set consisting of all functions $f\colon [0,1] \to \mathbb{R}$ that are continuous on $[0,1]$ with respect to the Euclidean metrics on $[0,1]$ and $\mathbb{R}$.

We define the set $C[a,b]$ of all continuous functions from the closed interval $[a,b]$ to $\mathbb{R}$ in a similar way.

Properties of functions in $C[0,1]$

Let f and g be real-valued continuous functions with domain $[0,1]$, i.e. $f, g \in C[0,1]$, and let $\lambda \in \mathbb{R}$. Then the following real-valued functions are all continuous on $[0,1]$, i.e. they are all members of $C[0,1]$:

Sum Rule $f+g\colon [0,1] \to \mathbb{R}$, defined by $(f+g)(x) = f(x) + g(x)$

Multiple Rule $\lambda f\colon [0,1] \to \mathbb{R}$, defined by $(\lambda f)(x) = \lambda \times f(x)$

Product Rule $fg\colon [0,1] \to \mathbb{R}$, defined by $(fg)(x) = f(x)g(x)$

Modulus Rule $|f|\colon [0,1] \to \mathbb{R}$, defined by $|f|(x) = |f(x)|$.

Definition 5.1 *Max metric on $C[0,1]$*

The **max metric** on $C[0,1]$ is the function $d_{\max}\colon C[0,1] \times C[0,1] \to \mathbb{R}$ defined by

$$d_{\max}(f,g) = \max\{|g(x) - f(x)| : x \in [0,1]\}.$$

Proposition 5.2

$(C[0,1], d_{\max})$ is a metric space.

Examples of continuous functions on $C[0,1]$

For each $h \in C[0,1]$, the functions $F_h\colon C[0,1] \to C[0,1]$ and $G_h\colon C[0,1] \to C[0,1]$ given by $F_h(f)(x) = f(x) + h(x)$ and $G_h(x) = h(x)f(x)$ are both $(d_{\max}, d_{\max})$-continuous on $C[0,1]$.

The function $H\colon C[0,1] \to C[0,1]$ given by $H(f)(t) = \int_0^t f(x)\,dx$ is $(d_{\max}, d_{\max})$-continuous on $C[0,1]$.

The function $I\colon C[0,1] \to \mathbb{R}$ given by $I(f) = \int_0^1 f(x)\,dx$ is $(d_{\max}, d^{(1)})$-continuous on $C[0,1]$.

Theorem 5.8 *Uniform Convergence Theorem*

Let (f_n) be a sequence of functions in $C[0,1]$.

If (f_n) converges uniformly on $[0,1]$ to the function $f\colon [0,1] \to \mathbb{R}$, then f is continuous on $[0,1]$ (and so is in $C[0,1]$) and $d_{\max}(f_n, f) \to 0$ as $n \to \infty$.

Conversely, if $f \in C[0,1]$ and $d_{\max}(f_n, f) \to 0$ as $n \to \infty$, then (f_n) converges uniformly to f on $[0,1]$.

Strategy: to determine convergence of sequences of functions in $C[0,1]$

1. Identify the possible limit function $f \in C[0,1]$ by finding the pointwise limit of (f_n).

   If there is no pointwise limit, then Lemma 4.5 implies that there is no uniform limit, and so Theorem 5.8 tells us that (f_n) is not convergent with respect to $d_{\max}$.

2. If the pointwise limit of (f_n) exists, determine whether it is continuous.

   If it is not continuous, then Lemma 4.5 tells us that the only possible uniform limit is not continuous, and Theorems 4.7 and 5.8 together then tell us that (f_n) is not convergent with respect to $d_{\max}$.

3. If f is the pointwise limit of (f_n) and is continuous, determine whether $(d_{\max}(f_n, f))$ is a null sequence.

   If it is, then by definition the sequence (f_n) converges to f for $d_{\max}$. Otherwise it does not converge in $C[0,1]$ for $d_{\max}$.

Chapter 16: Open and closed sets

Closed sets

Definition 1.1 *Closed set*

Let (X, d) be a metric space and let $A \subseteq X$.

We say that A is a **d-closed set** if for each d-convergent sequence with terms in A, the limit of the sequence is also in A. That is, whenever there is a d-convergent sequence (a_n) with $a_n \in A$ for each n then $\lim_{n\to\infty} a_n \in A$.

Lemma 1.2

Let (X, d) be a metric space. The following subsets of X are always d-closed:

1. the empty set, $\emptyset$
2. the whole space, X
3. sets consisting of a single point $\{x\}$ for some $x \in X$.

Proposition 1.3

Let (X, d) be a metric space, let $x \in X$ and let $r \geq 0$. Then the closed ball $B_d[x, r]$ is a d-closed set.

Definition 1.5 *Complement of a set*

Let A be a subset of a set X. The **complement of A with respect to X** is $A^c = X - A$, the set of elements that are in X but not in A.

The complement of the 'open' ball $B_d(x, r)$, $X - B_d(x, r)$, is d-closed.

Lemma 1.8

Let (a_n) be a convergent sequence of real numbers with limit a (for the usual Euclidean metric). If $m, M \in \mathbb{R}$ are such that for each $n \in \mathbb{N}$, $m \leq a_n \leq M$, then

$$m \leq a \leq M.$$

Lemma 1.13

Let (X, d) be a metric space and $(a_n)_{n=1}^{\infty}$ be a sequence in X. If (a_n) is d-convergent, then any subsequence of (a_n) is also d-convergent and has the same limit as the original sequence.

Proposition 1.14

Let (X, d) be a metric space and let A and B be d-closed subsets of X. Then $A \cup B$ is also d-closed.

Definition 1.16 *Intersection*

Let $\{A_i : i \in I\}$ be a collection of subsets of a given set X.

The **intersection** of the collection of sets, written $\bigcap_{i \in I} A_i$ or $\cap\{A_i : i \in I\}$, is defined by

$$\bigcap_{i \in I} A_i = \{x \in X : x \in A_i \text{ for each } i \in I\}.$$

Thus the intersection consists of exactly those points that the sets A_i have in common.

The set I is called an **index set** for the collection.

Proposition 1.17

Let (X, d) be a metric space and let $\{A_i : i \in I\}$ be a collection of d-closed sets. Then $\bigcap_{i \in I} A_i$ is d-closed.

Theorem 1.18

Let (X, d) be a metric space, let $a \in X$ and let $r \geq 0$. Then the sphere $S_d(a, r) = \{x \in X : d(a, x) = r\}$ is a d-closed set.

Open sets

Definition 2.1 *Open set*

Let (X, d) be a metric space and let $A \subseteq X$. Then A is said to be an **open set for the metric** d if the complement of A in X, $X - A$, is a d-closed set.

We say that A is a d-**open set** or that the set A is **open** in (X, d).

Lemma 2.2

Let (X, d) be a metric space. The following subsets of X are d-open:

1. the empty set, $\emptyset$
2. the whole space, X.

Proposition 2.3

Let (X, d) be a metric space, let $a \in X$ and let $r > 0$. Then the (open) ball $B_d(a, r) = \{x \in X : d(x, a) < r\}$ is a d-open set.

Theorem 2.7

Let (X, d) be a metric space, let $a \in X$ and $r > 0$, and consider the open ball $B_d(a, r)$. Then, whenever $x \in B_d(a, r)$, we can find an $s > 0$ such that

$$B_d(x, s) \subseteq B_d(a, r).$$

Theorem 2.9

Let (X, d) be a metric space and let A be a subset of X. Then A is d-open if, and only if, for each $a \in A$ there is $r > 0$ so that $B_d(a, r) \subseteq A$.

Theorem 2.12 *De Morgan's Laws*

Let $\{A_i : i \in I\}$ be a family of subsets of X. Then:

First law

$$X - \left(\bigcup_{i\in I} A_i\right) = \left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c = \bigcap_{i\in I}(X - A_i);$$

Second law

$$X - \left(\bigcap_{i\in I} A_i\right) = \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c = \bigcup_{i\in I}(X - A_i).$$

Theorem 2.13

In a metric space, the intersection of any two open sets is an open set.

Definition 2.14 *Union*

Let $\{A_i : i \in I\}$ be a collection of subsets of a given set X.

The **union** of the collection of sets, written $\bigcup_{i\in I} A_i$ or $\cup\{A_i : i \in I\}$, is defined by

$$\bigcup_{i\in I} A_i = \{x \in X : \text{there is } i \in I \text{ for which } x \in A_i\}.$$

Thus the union consists of exactly those points that appear in at least one of the sets A_i.

Theorem 2.15

In a metric space, the union of any collection of open sets is open.

Strategy: to determine whether sets are open or closed

Let (X, d) be a metric space and let $A \subseteq X$. We list the various techniques that you have seen for showing whether A is open, closed or neither. Remember that A can be both open and closed or could be neither.

- To show A is closed:
  - show that the limits of every convergent sequence with terms all lying in A lie in A; or
  - show that A is the union of a finite number of known closed sets; or
  - show that A is the intersection of an arbitrary collection of known closed sets; or
  - show that the complement of A is open.
- To show A is open:
  - show that the complement of A (in X) is closed; or
  - show that for each $a \in A$, there is $r > 0$ so that $B_d(a, r) \subseteq A$; or
  - show that A is the intersection of a finite number of known open sets; or
  - show that A is the union of an arbitrary collection of known open sets.
- To show A is not closed:
  - show that there is a convergent sequence (a_n) with $a_n \in A$ for each n and $\lim_{n\to\infty} x_n \notin A$; or
  - show that the complement of A is not open.
- To show A is not open:
  - show that there is a point $a \in A$ with no open ball $B_d(a, \varepsilon) \subseteq A$ – that is, show that for each open ball $B_d(a, \varepsilon)$, we have $B_d(a, \varepsilon) \cap (X - A) \neq \emptyset$; or
  - show that the complement of A is not closed.

Closure

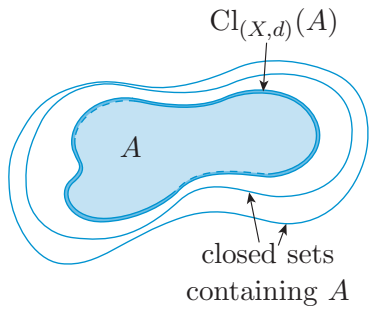
Definition 3.1 *Closure of a set*

Let (X, d) be a metric space and let $A \subseteq X$.

We define the **closure of A in (X, d)**, $\mathrm{Cl}_{(X,d)}(A)$, to be the smallest d-closed set in X that contains A. That is,

$$\mathrm{Cl}_{(X,d)}(A) = \bigcap \{E \subseteq X : A \subseteq E \text{ and } E \text{ is } d\text{-closed in } X\}.$$

We simply write $\mathrm{Cl}(A)$ when the context is clear.



The closure of A

Basic properties of closure

The closure of a set is closed.

It is always the case that $A \subseteq \mathrm{Cl}_{(X,d)}(A)$.

If F is a d-closed set and $A \subseteq F$, then $\mathrm{Cl}_{(X,d)}(A) \subseteq F$.

Theorem 3.2

Let (X, d) be a metric space and let $A, B \subseteq X$ with $A \subseteq B$. Then

$$\mathrm{Cl}_{(X,d)}(A) \subseteq \mathrm{Cl}_{(X,d)}(B).$$

Theorem 3.3

Let (X, d) be a metric space and let $A \subseteq X$. Then A is closed if, and only if, $\mathrm{Cl}_{(X,d)}(A) = A$.

In particular, $\mathrm{Cl}(\mathrm{Cl}(A)) = \mathrm{Cl}(A)$.

Theorem 3.5

Let (X, d) be a metric space and let $A, B \subseteq X$. Then:
(a) $\mathrm{Cl}_{(X,d)}(A \cup B) = \mathrm{Cl}_{(X,d)}(A) \cup \mathrm{Cl}_{(X,d)}(B)$
(b) $\mathrm{Cl}_{(X,d)}(A \cap B) \subseteq \mathrm{Cl}_{(X,d)}(A) \cap \mathrm{Cl}_{(X,d)}(B)$.

Definition 3.6 *Closure point*

Let A be a subset of the metric space (X, d).

A point $x \in X$ is a **d-closure point** of A in X if for each $r > 0$, $B_d(x, r)$ meets the set A; that is, $B_d(x, r) \cap A \neq \emptyset$.

Theorem 3.8

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Cl}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-closure point of } A \text{ in } X\}.$$

Corollary 3.11

Let (X, d) be a metric space and let A be a subset of X. A point a belongs to the closure $\mathrm{Cl}_{(X,d)}(A)$ of A if, and only if, there is a sequence (a_n) in A that converges to a.

Theorem 3.12

Let (X, d) be a metric space and let $A \subseteq X$.

A set $E \subseteq A$ is d_A-closed if, and only if, there is a set F in X that is d-closed and for which

$$E = A \cap F.$$

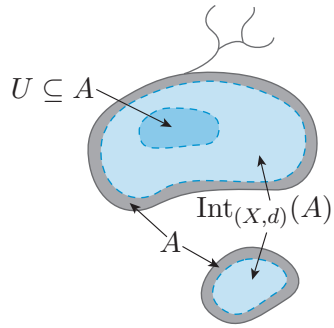
Interior

Definition 4.1 *Interior of a set*

Let (X, d) be a metric space and suppose $A \subseteq X$. We define the **interior of A in (X, d)**, $\text{Int}_{(X,d)}(A)$, to be the largest d-open set in X that is contained in A. That is,

$$\text{Int}_{(X,d)}(A) = \bigcup \{U : U \subseteq A \text{ and } U \text{ is } d\text{-open in } X\}.$$

We simply write $\text{Int}(A)$ when the context is clear.



The interior of A

Basic properties of interior

The interior of a set is always an open set.

It is always the case that $\text{Int}(A) \subseteq A$.

Theorem 4.3

Let (X, d) be a metric space and let $A, B \subseteq X$. Then:

(a) $\text{Int}_{(X,d)}(A) \cup \text{Int}_{(X,d)}(B) \subseteq \text{Int}_{(X,d)}(A \cup B)$

(b) $\text{Int}_{(X,d)}(A) \cap \text{Int}_{(X,d)}(B) = \text{Int}_{(X,d)}(A \cap B)$

Theorem 4.4

Let (X, d) be a metric space and let $A \subseteq X$.

The interior of A is given by

$$\text{Int}_{(X,d)}(A) = \{x \in X : \text{there is } r > 0 \text{ so that } B_d(x, r) \subseteq A\}.$$

Definition 4.5 *Interior point*

Let (X, d) be a metric space and let $A \subseteq X$.

We say $x \in X$ is a **d-interior point of A** if there is an $r > 0$ so that $B_d(x, r) \subseteq A$.

Theorem 4.4 then tells us that

$$\text{Int}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-interior point of } A\}.$$

Theorem 4.6

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\text{Int}_{(X,d)}(A) = \left[\text{Cl}_{(X,d)}(A^c)\right]^c = X - \text{Cl}_{(X,d)}(X - A).$$

In particular, A is d-open if, and only if, $\text{Int}_{(X,d)}(A) = A$.

Boundaries of sets

Definition 5.1 *Boundary*

Let (X, d) be a metric space and let $A \subseteq X$.

The **d-boundary of A in X** is the set

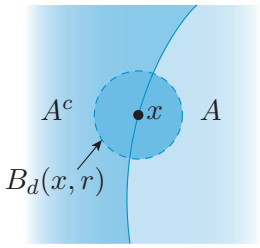
$$\text{Bd}_{(X,d)}(A) = \text{Cl}_{(X,d)}(A) - \text{Int}_{(X,d)}(A).$$

Theorem 5.2

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = \mathrm{Cl}_{(X,d)}(A) \cap \mathrm{Cl}_{(X,d)}(A^c).$$

In particular, $\mathrm{Bd}_{(X,d)}(A)$ is a closed set.



A boundary point of A

Definition 5.3 *Boundary point*

Let (X, d) be a metric space and let $A \subseteq X$.

A point $x \in X$ is a **d-boundary point of A** if for each $r > 0$, $B_d(x, r)$ meets both A and A^c.

Theorem 5.4

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = \{x \in X : x \text{ is a } d\text{-boundary point of } A\}.$$

Theorem 5.5

Let (X, d) be a metric space and let $A \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A) = (\mathrm{Int}_{(X,d)}(A^c) \cup \mathrm{Int}_{(X,d)}(A))^c.$$

Theorem 5.6

Let (X, d) be a metric space and let $A, B \subseteq X$. Then

$$\mathrm{Bd}_{(X,d)}(A \cup B) \subseteq \mathrm{Bd}_{(X,d)}(A) \cup \mathrm{Bd}_{(X,d)}(B).$$

The size of sets

Definition 6.1 *Dense and nowhere dense*

Let (X, d) be a metric space and let $A \subseteq X$.

The set A is **d-dense in** X if $\mathrm{Cl}_{(X,d)}(A) = X$.

The set A is **nowhere dense in** X if $\mathrm{Int}_{(X,d)}(\mathrm{Cl}_{(X,d)}(A)) = \emptyset$.

The underlying set X of a metric space is always dense in itself and the empty set is nowhere dense.

Basic properties of dense and nowhere dense sets

1. Let (X, d) be a metric space and let $A \subseteq X$ be dense in X. If $A \subseteq B$ then B is also dense in X.
2. Let (X, d) be a metric space and let $A \subseteq X$ be nowhere dense in X. If $B \subseteq A$, then B is nowhere dense in X.
3. Let (X, d) be a metric space and let U and V be d-open subsets of X that are d-dense in X. Then $U \cap V$ is also (a d-open subset of X that is) d-dense in X.

Theorem 6.3

Let (X, d) be a metric space and let $A \subseteq X$. Let d_A denote the subspace metric for A.

A set B contained in A is d_A-dense in A if, and only if, $A \subseteq \mathrm{Cl}_{(X,d)}(B)$.

Definition 6.4 *Countable and uncountable*

A set A is **countable** if there exists a one–one map $f\colon A \to \mathbb{N}$, and is **uncountable** otherwise.

A countable set with infinitely many elements is said to be **countably infinite**.

Some countable and uncountable sets

Set	Size	Chapter reference
$\mathbb{N}$	countable	Definition 6.4
$\mathbb{Z}$	countable	Exercise 6.3
subsets of countable sets	countable	Worked Exercise 6.5
$\mathbb{Q}$	countable	Theorem 6.6
$(0,1)$	uncountable	Theorem 6.7
$\mathbb{R}$	uncountable	Corollary 6.8
(a,b) with $a<b$	uncountable	Exercise 6.4
$A=\bigcup_{n=1}^{\infty} A_n$ with A_n countable	countable	Theorem 6.9
union of finitely many countable sets	countable	Corollary 6.10
$\mathbb{Q}\times\mathbb{Q}$	countable	Worked Exercise 6.11

Continuity and open and closed sets

Theorem 7.1

Let (X,d) and (Y,e) be metric spaces and suppose that $f\colon X\to Y$. Then f is (d,e)-continuous on X if, and only if, for each e-closed set $E\subseteq Y$, $f^{-1}(E)$ is d-closed.

Corollary 7.2

Let (X,d) and (Y,e) be metric spaces and suppose that $f\colon X\to Y$. Then f is (d,e)-continuous on X if, and only if, for each e-open set $U\subseteq Y$, $f^{-1}(U)$ is d-open.

Properties of open sets

If (X,d) is a metric space, then the collection, $\mathcal{T}$ say, of *d-open sets* has the following properties:

(T1) Both $\emptyset$ and X are in $\mathcal{T}$.

(T2) If U and V are in $\mathcal{T}$, then so is $U\cap V$.

(T3) If $\{U_i : i\in I\}$ is a collection of sets that are all in $\mathcal{T}$, then

$$\bigcup_{i\in I} U_i \text{ is in } \mathcal{T}.$$

In addition, if $f\colon X\to Y$ and (Y,e) is a metric space, then f is (d,e)-continuous on X if, and only if, for each e-open set U in Y, $f^{-1}(U)$ is d-open in X.

Definition 7.3 *Topology*

We define a **topology** for a set X to be any collection $\mathcal{T}$ of subsets of X that satisfy (T1), (T2) and (T3). We call the sets that are in $\mathcal{T}$ the *open sets*.

A **topological space** is a set X together with a collection of subsets of X, $\mathcal{T}$ say, that form a topology for X, denoted $(X, \mathcal{T})$.

Definition 7.4 *Continuity in a topological space*

Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces and let $f\colon X \to Y$. We say that f is $(\mathcal{T}_X, \mathcal{T}_Y)$-continuous if for each $U \in \mathcal{T}_Y$, $f^{-1}(U) \in \mathcal{T}_X$.

Appendix 4: Examples of metric spaces

Metric space	Name of metric	Definition of metric
(X, d_0), X any set	discrete metric	$d_0(a,b) = \begin{cases} 0, & \text{if } a = b, \\ 1, & \text{if } a \neq b. \end{cases}$
$(\mathbb{R}^n, d^{(n)})$	Euclidean metric	$d^{(n)}(\mathbf{x},\mathbf{y}) = \sqrt{(y_1 - x_1)^2 + \cdots + (y_n - x_n)^2}$
(A^n, d_H), A a finite set	Hamming distance	$d_H(\mathbf{a},\mathbf{b})$ = number of places at which $\mathbf{a}$ and $\mathbf{b}$ differ
$(\mathbf{C}, d_\mathbf{C})$	Cantor distance	$d_\mathbf{C}(\mathbf{x},\mathbf{y}) = \begin{cases} 0, & \text{if } \mathbf{x} = \mathbf{y}, \\ 2^{-n}, & \text{if } \mathbf{x} \text{ and } \mathbf{y} \text{ first differ at their } n\text{th term.} \end{cases}$
$(C[0,1], d_{\max})$	max metric for $C[0,1]$	$d_{\max}(f,g) = \max\{\lvert g(x) - f(x)\rvert : x \in [0,1]\}$
(l_0, d)	metric for eventually zero sequences	$d(\mathbf{a},\mathbf{b}) = \sum_{n=1}^{\infty} \lvert b_n - a_n\rvert$
$(C[0,1], d)$	integral (or L^1) metric for $C[0,1]$	$d(f,g) = \int_0^1 \lvert g(x) - f(x)\rvert \, dx$
(A, d_A) where A is a subset of the metric space (X, d)	induced metric	$d_A(x,y) = d(x,y)$ for $x, y \in A$
$(X_1 \times X_2, e)$ where (X_1, d_1) and (X_2, d_2) are metric spaces	product metrics	$e_1(\mathbf{x},\mathbf{y}) = d_1(x_1,y_1) + d_2(x_2,y_2)$ $e_2(\mathbf{x},\mathbf{y}) = \sqrt{d_1(x_1,y_1)^2 + d_2(x_2,y_2)^2}$ $e_\infty(\mathbf{x},\mathbf{y}) = \max\{d_1(x_1,y_1), d_2(x_2,y_2)\}$

Appendix 5: Examples of open and closed sets in metric spaces

Space	Examples of open sets	Examples of closed sets
any metric space (X, d)	$\emptyset$, X, $B_d(x, r)$	$\emptyset$, X, $B_d[x, r]$, $S_d(x, r)$, $X - B_d(x, r)$, singletons: $\{x\}$
$(\mathbb{R}, d^{(1)})$	open intervals: (a, b), $(-\infty, a)$, (b, ∞)	the middle-third Cantor set, closed intervals: $[a, b]$, $(-\infty, a]$, $[b, \infty)$
(X, d_0), discrete metric	all sets	all sets
$(\mathbf{C}, d_{\mathbf{C}})$	$B_{d_{\mathbf{C}}}(\mathbf{a}, r)$, $B_{d_{\mathbf{C}}}[\mathbf{a}, r]$	$B_{d_{\mathbf{C}}}(\mathbf{a}, r)$, $B_{d_{\mathbf{C}}}[\mathbf{a}, r]$
$(C[0,1], d_{\max})$	$\{f \in C[0,1] : \lvert f(0)\rvert \neq M\}$, $\{f \in C[0,1] :$ there is $x \in [0,1]$ for which $\lvert f(x)\rvert > M\}$	$\{f \in C[0,1] : \lvert f(a)\rvert \leq M\}$, $\{f \in C[0,1] : \lvert f(0)\rvert = M\}$

Book E: Rings and fields

Glossary

Chapter 18, Section 2 — **adjoining α to F** The process of obtaining the algebraic extension $F(\alpha)$ of F.

Chapter 17, Section 4 — **canonical homomorphism** The homomorphism $\hat{f}_p : \mathbb{Z}[x] \to \mathbb{Z}_p[x]$ that takes each polynomial with integer coefficients, and reduces its coefficients modulo some prime number p; that is,

$$\hat{f}_p(a_0 + a_1x + \cdots + a_nx^n) = \hat{a}_0 + \hat{a}_1x + \cdots + \hat{a}_nx^n,$$

where $a_i \equiv \hat{a}_i \pmod{p}$, for $i = 1, 2, \ldots, n$.

Chapter 17, Section 2 — **chain of ideals** A sequence of ideals (possibly infinite) of a ring in which each ideal is a subset of its successor: $I_0 \subseteq I_1 \subseteq I_2 \subseteq \cdots$.

Chapter 20, Introduction — **cryptosystem** A procedure for implementing a particular method of encryption and decryption; also called a **cryptographic system**.

Chapter 18, Section 4 — **derivation** A map from a polynomial ring to a polynomial ring that satisfies $d(f(x) + g(x)) = d(f(x)) + d(g(x))$, and $d(f(x)g(x)) = d(f(x))g(x) + f(x)d(g(x))$; a generalisation of the familiar operation of differentiation.

Chapter 19, Section 2 — **duplication of the cube** The problem of constructing a cube of twice the volume of a given cube.

Chapter 18, Section 1 — **embedding** A map where the image of the domain is isomorphic to the domain.

Chapter 18, Section 3 — F^* The multiplicative group whose elements are the non-zero elements of a finite field F.

Chapter 20, Section 1 — **Fast Euclidean Algorithm** A faster variant of the Euclidean Algorithm procedure that requires pre-checking the input at each step.

Chapter 17, Section 1 — **field of Gaussian numbers** The field $\mathbb{Q}[i] = \{a + bi : a, b \in \mathbb{Q}\}$ where $i = \sqrt{-1}$, which is isomorphic to the field of fractions of the Gaussian integers $\mathbb{Z}[i]$.

Chapter 17, Section 1 — **fraction in its lowest terms** A $\frac{c}{d} \in F_D$ where the two elements c and d of the unique factorisation domain D are such that $\text{hcf}(c, d)$ is a unit.

Chapter 18, Section 1 — **Galois field GF_p** The only field (up to isomorphism) of p elements, also denoted by $\mathbb{Z}_p$.

Chapter 20, Section 1 — **general number sieve algorithm** Currently the best factorisation algorithm available: much faster than the naive algorithm that tests divisibility by primes up to the square root of the number.

Chapter 18, Section 1 — **prime field of F** The smallest subfield of a field F of prime or zero characteristic.

public-key cryptosystem A cryptosystem with two keys: a **public key** used to encrypt the **plaintext** message, and a **private key** used to decrypt the **ciphertext** to obtain the original message. The two keys, usually in the form of some positive integers, are algebraically linked. **Chapter 20**, Introduction

quadrature of the circle The problem of constructing a square of the same area as a given circle. **Chapter 19**, Section 2

ring of Laurent polynomials over F The ring $F[x, x^{-1}] = \{a_m x^m + \cdots + a_n x^n : a_i \in F \text{ for } i = m, \ldots, n \text{ with } m, n \in \mathbb{Z}, m < n\}$ where F is a field; it is a Euclidean domain and hence a UFD. **Chapter 17**, Section 1

square modulo p A non-zero element b of GF_p such that there exists an $x \in GF_p$ with $b = x^2$; also called a **quadratic residue of** p. **Chapter 20**, Section 1

square root modulo p A non-zero element x of GF_p is a square root of b modulo p if $b = x^2$. **Chapter 20**, Section 1

squaring method A quick method for computing residue classes of high powers by writing the exponent as a sum of powers of 2, finding these powers by repeated squaring then multiplying them together. **Chapter 20**, Section 1

trisection of an angle The problem of obtaining an angle equal to one third of a given angle. **Chapter 19**, Section 2

Chapter 17: Rings and homomorphisms

Fields of fractions

Proposition 1.1

Let $(D, +, \cdot)$ be an integral domain and let

$$S_D = \{(a, b) : a, b \in D,\ b \neq 0\}.$$

Then the relation $\sim$ on the set S_D defined by

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc$$

is an equivalence relation.

Lemma 1.2

Let D be an integral domain and let $S_D = \{(a, b) : a, b \in D,\ b \neq 0\}$. Let $F_D = \{[a, b] : (a, b) \in S_D\}$ denote the set of equivalence classes on S_D induced by the relation $\sim$ defined by

$$(a, b) \sim (c, d) \text{ if, and only if, } ad = bc.$$

(a) If $a, b, c \in D$ with $b \neq 0$ and $c \neq 0$, then $(a, b) \sim (ac, bc)$.

(b) If $a, b, d \in D$ with $b \neq 0$ and $d \neq 0$, then $(a, b) \sim (0, d)$ if, and only if, $a = 0$.

Proposition 1.3

Let D be an integral domain and let S_D and F_D be as in Lemma 1.2. Let $+$ and $\cdot$ be binary operations on F_D defined by

$$\begin{aligned} [a, b] + [c, d] &= [ad + bc, bd] \\ [a, b] \cdot [c, d] &= [ac, bd]. \end{aligned}$$

Then F_D is a field with the property that every element can be written in the form

$$[a, 1][b, 1]^{-1}$$

for some $a, b \in D$, $b \neq 0$.

Definition 1.4 *Ring isomorphism*

Let R and S be rings. Then we say that R and S are **isomorphic** if there is a bijection $\theta : R \to S$ such that, for all $a, b \in R$,

- $\theta(a + b) = \theta(a) + \theta(b)$
- $\theta(ab) = \theta(a)\theta(b)$
- $\theta(1_R) = 1_S$.

The bijection θ is said to be a **ring isomorphism**.

These properties of addition and multiplication are known as the **homomorphism** or **morphism properties**.

Proposition 1.5

Let E be a field, and let D be a subring of E such that

1. D is an integral domain
2. every element of E can be written as ab^{-1} with $a, b \in D$.

Then E is isomorphic to F_D.

Definition 1.6 *Field of fractions*

Let D be an integral domain. Then a field F_D is called the **field of fractions** of D if D is isomorphic to a subring A of F_D and every element of F_D can be written as ab^{-1} where $a, b \in A$ and $b \neq 0$.

Definition 1.9 *Primitive polynomial*

Let R be a unique factorisation domain, and let

$$f(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$$

be a polynomial in $R[x]$.

(a) The **content** of f is a highest common factor of all of its coefficients in R: $c_f = \text{hcf}(a_0, a_1, \ldots, a_n)$.

(b) $f(x)$ is **primitive** if $c_f = u$ for some unit in $u \in R$. In other words, $f(x)$ is primitive if the only common factors of its coefficients are units of R.

Lemma 1.10

Let R be a unique factorisation domain. Then f and g are primitive polynomials in $R[x]$ if, and only if, their product fg is primitive.

Lemma 1.11 *Gauss's Lemma for UFDs*

Let R be a UFD and let F be the field of fractions of R. Let $f(x)$ be a primitive polynomial in $R[x]$. If we can write $f = gh$ with $g, h \in F[x]$, then we can write $f = GH$ with $G, H \in R[x], \deg(g) = \deg(G)$ and $\deg(h) = \deg(H)$.

Theorem 1.12 *Theorem 4.2 of Chapter 12*

Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a UFD if, and only if, every irreducible element in R is prime.

Lemma 1.13

Let R be a UFD. Then every polynomial in $R[x]$ has a factorisation into irreducible polynomials in $R[x]$.

Lemma 1.14

Let R be an integral domain and let $R[x]$ be the ring of polynomials in one variable over R. If $p \in R$ is prime in R then it is prime in $R[x]$.

Lemma 1.15

Let R be a UFD, and f be an irreducible element in $R[x]$. Then f is prime in $R[x]$.

Theorem 1.16

Let R be a UFD. Then the polynomial ring $R[x]$ is a UFD.

Ideals

Definition 2.1 *Ideal*

Let $(R, +, \cdot)$ be a commutative ring and let I be a non-empty subset of R. Then I is an **ideal** of R if

I1 for all $a, b \in I$, $a - b \in I$

I2 for all $a \in I$ and $r \in R$, $ra \in I$.

Definition 2.2 *Proper and non-trivial ideal*

An ideal I of a ring R is said to be **proper** if it is not equal to R. It is said to be **non-trivial** if it is not equal to $\{0\}$.

Properties of ideals

Let I be an ideal of the ring R. Then:

- $0 \in I$
- I is an additive subgroup
- I is a proper ideal if, and only if, it does not contain the multiplicative identity 1.

Definition 2.5 *Principal ideal*

Let R be a commutative ring, and let $a \in R$. Then the set

$$aR = \{ar : r \in R\}$$

is the **principal ideal** of R generated by a. If the ring R is clear from the context, we often write $\langle a \rangle$ instead of aR.

Lemma 2.7

Let R be a commutative ring, and let $a, b \in R$. Then the following statements are equivalent.

(a) $aR \subseteq bR$

(b) $b \mid a$

(c) $a \in bR$

Theorem 2.8

Every ideal in $\mathbb{Z}$ is a principal ideal. Every non-zero ideal in $\mathbb{Z}$ is generated by the smallest positive integer that it contains.

Theorem 2.9

Every ideal I in a Euclidean domain R is a principal ideal generated by an element with the least norm in I, or by 0 in the case $I = \{0\}$.

Definition 2.10 *Principal ideal domain*

A **principal ideal domain** (**PID**) is an integral domain in which every ideal is principal.

Corollary 2.11 *to Theorem 2.9*

Every Euclidean domain is a principal ideal domain.

Theorem 2.13 *Highest common factors in PIDs*

Let R be a principal ideal domain, and let $a, b \in R$. Then a and b have a highest common factor $\text{hcf}(a, b)$, that is, a common factor that is a multiple of any other common factor.

Moreover, if $h = \text{hcf}(a, b)$ then there are elements $s, t \in R$ such that $h = sa + tb$.

Theorem 2.15

Let R be a principal ideal domain. Then p is prime in R if, and only if, p is an irreducible element of R.

Theorem 2.16 *Unique factorisation in PIDs*

Every principal ideal domain R is a unique factorisation domain.

Summary of results about different types of domain

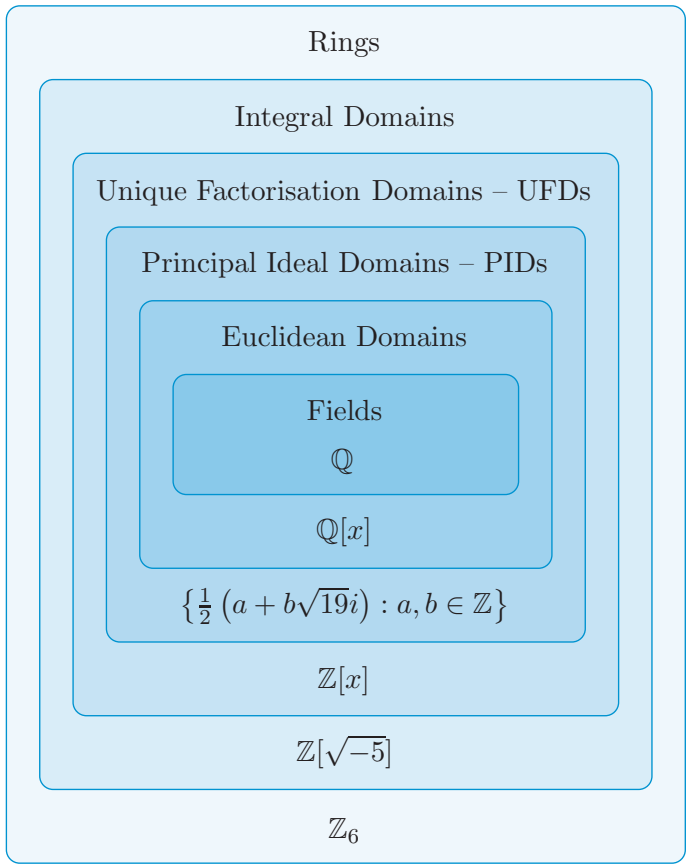
- A polynomial ring in one variable over a field is a Euclidean domain (Book C, Chapter 11, Theorem 3.1).
- Every Euclidean domain is a principal ideal domain (Corollary 2.11 of Chapter 17).
- Every principal ideal domain is a unique factorisation domain (Theorem 2.16 of Chapter 17).
- If R is a unique factorisation domain then $R[x]$ is a unique factorisation domain (Theorem 1.16 of Chapter 17).
- Unique factorisation domains need not be Euclidean domains (Book C, Chapter 12, Example 4.5).
- Not all principal ideal domains are Euclidean domains.
- Not all unique factorisation domains are PIDs. For example, $\mathbb{Z}[x]$ is a unique factorisation domain that is not a principal ideal domain.

Summary of results about irreducible elements

- Let R be an integral domain and p a prime element in R. Then p is irreducible (Book C, Chapter 12, Theorem 2.12).
- Let R be a unique factorisation domain. Then p is prime in R if, and only if, p is an irreducible element of R (Book C, Chapter 12, Theorem 2.12 and Corollary 4.3).
- Let R be an integral domain in which every non-zero non-unit element has a factorisation into irreducible elements. Then R is a unique factorisation domain if, and only if, every irreducible element in R is prime (Book C, Chapter 12, Theorem 4.2).

- Let R be a unique factorisation domain and let $a, b \in R$. Then a and b have a highest common factor $\text{hcf}(a, b)$ (Book C, Chapter 12, Proposition 4.7).
- Let R be a principal ideal domain and let $a, b \in R$. Then a and b have a highest common factor $\text{hcf}(a, b)$. If $h = \text{hcf}(a, b)$, then there are elements $s, t \in R$ such that $h = sa + tb$ (Theorem 2.13 of Chapter 17).

The relationships between different types of commutative rings



Operations on ideals

Lemma 3.1

Let A and B be ideals of a ring R. Then the set

$$A + B = \{a + b : a \in A, b \in B\}$$

is an ideal of R.

Definition 3.2 *Sum of two ideals*

The **sum of two ideals** A and B of a ring R is the ideal

$$A + B = \{a + b : a \in A, b \in B\}.$$

Proposition 3.3 *Basic additive properties of ideals*

Let A, B and C be ideals of a commutative ring R. Then:
(a) $A \subseteq (A + B)$ and $B \subseteq (A + B)$
(b) $A + B = B + A$
(c) $(A + B) + C = A + (B + C)$
(d) $A + \{0\} = A$
(e) $A + R = R$.

Definition 3.4 *Product of two ideals*

The **product of two ideals** A and B of a ring R is the ideal

$$AB = \{a_1b_1 + a_2b_2 + \cdots + a_nb_n : n \in \mathbb{N} \text{ and } a_i \in A, b_i \in B\}.$$

Proposition 3.5 *Basic multiplicative properties of ideals*

Let A, B and C be ideals of a commutative ring R. Then:
(a) $AB \subseteq A$ and $AB \subseteq B$
(b) $AB = BA$
(c) $(AB)C = A(BC)$
(d) $A\{0\} = \{0\}$
(e) $AR = A$.

Definition 3.6 *Ideal generated by a set*

Let R be a commutative ring. Then the ideal of R generated by a finite subset $S = \{a_1, \ldots, a_n\}$ of R is the set $\langle a_1, \ldots, a_n \rangle$ defined by

$$\langle a_1, \ldots, a_n \rangle = \{a_1r_1 + \cdots + a_nr_n : r_1, \ldots, r_n \in R\}.$$

The notation $\langle a_1, \ldots, a_n \rangle$ is sometimes abbreviated to $\langle S \rangle$.

Lemma 3.7

Let R be a PID, and let $a_1, a_2, \ldots, a_n \in R$. Then

$$\langle a_1, a_2, \ldots, a_n \rangle = \langle \text{hcf}(a_1, a_2, \ldots, a_n) \rangle.$$

Proposition 3.8

Let R be a commutative ring and suppose that $a_1, a_2, \ldots, a_n \in R$.
Let A be an ideal of R. Then

$$\langle a_1, a_2, \ldots, a_n \rangle \subseteq A \text{ if, and only if, } a_1, a_2, \ldots, a_n \in A.$$

Lemma 3.9

The product of the ideals $A = \langle a_1, a_2 \rangle$ and $B = \langle b_1, b_2 \rangle$ of a ring R is the ideal

$$I = \langle a_1b_1, a_1b_2, a_2b_1, a_2b_2 \rangle.$$

Quotients

Definition 4.1 *Cosets of an ideal*

Let I be an ideal of a ring R, and let $a \in R$. Then

$$a + I = \{a + r : r \in I\}$$

is the **coset** of I by a.

Proposition 4.2

Let R be a ring and I an ideal of R. Then the cosets of I partition R into disjoint subsets, and the following four statements are equivalent.

(a) $b - a \in I$

(b) $a + I = b + I$

(c) a and b are in the same coset of I

(d) $b \in a + I$

Theorem 4.4

Let I be an ideal of a ring R. Then the set of cosets of I in R, with the operations $+$ and $\cdot$ defined by

$$\begin{aligned}(a + I) + (b + I) &= (a + b) + I \quad \text{and} \\ (a + I) \cdot (b + I) &= ab + I\end{aligned}$$

is a ring, called the **quotient ring** of R by I, and denoted by R/I.

The additive identity of R/I is $0 + I$, the multiplicative identity is $1 + I$, and the additive inverse of $a + I$ is $-a + I$.

Definition 4.7 *Ring homomorphism*

Let R and S be rings. A **ring homomorphism** is a function $f : R \to S$ such that, for all $a, b \in R$:

$$f(a + b) = f(a) + f(b);$$
$$f(ab) = f(a)f(b);$$
$$f(1_R) = 1_S,$$

where 1_R and 1_S are the multiplicative identities in R and S respectively.

These properties of addition and multiplication are known as the **homomorphism** or **morphism properties**.

Proposition 4.9 *Basic properties of homomorphisms*

Let $f : R \to S$ be a ring homomorphism. Then:

(a) $f(0_R) = 0_S$, where 0_R and 0_S are the additive identities of R and S respectively

(b) $f(-a) = -f(a)$ for all $a \in R$

(c) if u is a unit in R, then $f(u)$ is a unit in S.

Definition 4.10 *Ring automorphism*

A ring isomorphism $f : R \to R$ whose domain and codomain are the same ring R is known as a ring **automorphism**.

Definition 4.11 *Image and kernel*

Let R and S be rings, and $f : R \to S$ a ring homomorphism.
(a) The **image** of f is the set $\mathrm{Im}(f) = \{f(a) : a \in R\}$.
(b) The **kernel** of f is the set $\mathrm{Ker}(f) = \{a \in R : f(a) = 0\}$.

Proposition 4.12

Let $f : R \to S$ be a ring homomorphism. Then:
(a) f is onto if, and only if, $\mathrm{Im}(f) = S$
(b) f is one–one if, and only if, $\mathrm{Ker}(f) = \{0\}$.

Proposition 4.14

Let R and S be rings, and $f : R \to S$ a ring homomorphism. Then:
(a) $\mathrm{Im}(f)$ is a subring of S
(b) $\mathrm{Ker}(f)$ is an ideal of R.

Proposition 4.15

Let I be an ideal of a ring R. Then there is a homomorphism $f : R \to R/I$ such that $\mathrm{Ker}(f) = I$.

Definition 4.16 *Natural homomorphism*

Let I be an ideal of a ring R. Then the **natural homomorphism** from R to R/I is the function

$$\begin{aligned} f \; : \; & R \to R/I \\ & a \mapsto a + I. \end{aligned}$$

Theorem 4.17 *First Isomorphism Theorem*

Let $f : R \to S$ be a ring homomorphism from a ring R to a ring S. Then the function $\psi : R/\mathrm{Ker}(f) \to \mathrm{Im}(f)$ defined by

$$\psi : a + \mathrm{Ker}(f) \mapsto f(a)$$

is an isomorphism.

Definition 4.19 *Evaluation homomorphism*

Let $\alpha \in \mathbb{C}$. Then the function $f_\alpha : \mathbb{Q}[x] \to \mathbb{C}$ defined by

$$f_\alpha(a_0 + a_1 x + \cdots + a_n x^n) = a_0 + a_1\alpha + \cdots + a_n\alpha^n$$

is called an **evaluation homomorphism**.

Definition 4.20 *Maximal ideal*

A proper ideal I of a ring R is a **maximal ideal** if whenever J is an ideal of R such that $I \subseteq J \subseteq R$, then either $I = J$ or $J = R$.

Theorem 4.22

Let R be a PID, and let a be a non-zero, non-unit element of R. Then $\langle a \rangle$ is a maximal ideal of R if, and only if, a is irreducible in R.

Theorem 4.23

Let I be an ideal of a ring R. Then I is a maximal ideal if, and only if, R/I is a field.

Definition 4.25 *Prime ideal*

A proper ideal P in a ring R is **prime** if whenever $a, b \in R$ are such that $ab \in P$, then either $a \in P$ or $b \in P$.

Theorem 4.26

Let R be a commutative ring and P be an ideal of R. Then P is a prime ideal if, and only if, R/P is an integral domain.

Proposition 4.27

Let R be a PID and let P be a proper non-trivial ideal in R. If P is prime, then:

(a) $P = \langle p \rangle$ for some prime element $p \in P$

(b) P is maximal in R.

Lemma 4.28

Let P be a prime ideal in a ring R. Then if $a_1 a_2 \cdots a_n \in P$ there is an $i \in \{1, \ldots, n\}$ with $a_i \in P$.

Proposition 4.29

Let R be a UFD and suppose that P is a non-zero prime ideal in R. Then P contains an irreducible element q.

Chapter 18: Fields and polynomials

Preliminaries

Proposition 1.1

Let F_1 and F_2 be fields and suppose that $\theta : F_1 \to F_2$ is a ring homomorphism. If $a \in F_1$ with $a \neq 0$, then

$$\theta(a^{-1}) = \theta(a)^{-1}.$$

Definition 1.2 *Isomorphism of fields*

Let F_1 and F_2 be fields. We say that F_1 and F_2 are **isomorphic** if, and only if, there is a ring isomorphism $\theta : F_1 \to F_2$. We then also say that θ is a **field isomorphism**.

Proposition 1.4

Let F and K be fields and suppose that $f : F \to K$ is a non-zero ring homomorphism. Then f is an embedding, that is $\operatorname{Im} f \cong F$.

Proposition 1.5

Let F be a field and let $p(x)$ be an irreducible polynomial in $F[x]$. Then $E = F[x]/\langle p(x) \rangle$ is a field.

Proposition 1.7

Let $p(x) = p_0 + p_1x + \cdots + p_nx^n$ be an irreducible polynomial over a field F, so that $E = F[x]/\langle p(x)\rangle$ is a field. Let $\theta : F[x] \to E$ be the natural homomorphism and let $\alpha = \theta(x) = x + \langle p(x)\rangle$. Then:

(a) Any element of E can be represented, uniquely, in the form

$$a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1},$$

where $a_0, a_1, \ldots, a_{n-1} \in F$.

With this representation of elements of E, addition is given by

$$\begin{aligned}(a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1}) + (b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}) \\ = a_0 + b_0 + (a_1 + b_1)\alpha + \cdots + (a_{n-1} + b_{n-1})\alpha^{n-1}\,,\end{aligned}$$

and multiplication is given by

$$\begin{aligned}(a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1})(b_0 + b_1\alpha + \cdots + b_{n-1}\alpha^{n-1}) \\ = (r_0 + r_1\alpha + \cdots + r_{n-1}\alpha^{n-1})\,,\end{aligned}$$

where $r_0 + r_1x + \cdots + r_{n-1}x^{n-1} \in F[x]$ is the remainder in the division of $(a_0 + a_1x + \cdots + a_{n-1}x^{n-1})(b_0 + b_1x + \cdots + b_{n-1}x^{n-1})$ by $p(x)$.

(b) In E, we have

$$p(\alpha) = p_0 + p_1\alpha + \cdots + p_n\alpha^n = 0,$$

which means that α can be regarded as a root of the polynomial $p(x)$ with this representation of the elements of E.

Definition 1.9 *Characteristic of a field*

If the order of 1 in the additive group $(F, +)$ of a field F is infinite, then F is said to have **characteristic zero**. If the order of 1 in the group $(F, +)$ is finite and equal to m, we say that the field F has **characteristic** m.

Theorem 1.10

The characteristic of any field is either prime or zero.

Theorem 1.11

If a field F has prime characteristic p, then F contains a subfield isomorphic to $\mathbb{Z}_p$, and this is the smallest subfield of F with respect to inclusion, that is, it is contained in every other subfield of F.

Similarly, if F has characteristic zero, then F contains a subfield isomorphic to $\mathbb{Q}$, and this is the smallest subfield of F with respect to inclusion.

Theorem 1.12 *Idiot's Binomial Theorem*

Let F be a field of characteristic p. Then

$$(u + v)^p = u^p + v^p$$

for all $u, v \in F$.

Proposition 1.13

Let F_1 and F_2 be fields and suppose that there is a non-zero field homomorphism $\phi : F_1 \to F_2$. Then F_1 and F_2 have the same characteristic.

Theorem 1.14

Let F_1 and F_2 be subfields of $\mathbb{C}$ and let $\phi : F_1 \to F_2$ be an embedding. Then $\phi(x) = x$ for every $x \in \mathbb{Q}$.

Field extensions

Definition 2.1 *Extension of a field*

A field E is an **extension of a field** F if F is a subfield of E.

Definition 2.3 *Vector spaces over general fields*

A **vector space over a field** F consists of a set V of elements and two operations – vector addition and scalar multiplication – such that the following axioms hold.

Axioms for vector addition:

A1 Closure For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 \in V$.

A2 Additive identity There is a vector $\mathbf{0} \in V$ such that for each $\mathbf{v} \in V$, $\mathbf{v} + \mathbf{0} = \mathbf{0} + \mathbf{v} = \mathbf{v}$.

A3 Additive inverses For each $\mathbf{v} \in V$, there is an element $-\mathbf{v}$ (its additive inverse) such that $\mathbf{v} + (-\mathbf{v}) = (-\mathbf{v}) + \mathbf{v} = \mathbf{0}$.

A4 Associativity For all $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3 \in V$, $(\mathbf{v}_1 + \mathbf{v}_2) + \mathbf{v}_3 = \mathbf{v}_1 + (\mathbf{v}_2 + \mathbf{v}_3)$.

A5 Commutativity For all $\mathbf{v}_1, \mathbf{v}_2 \in V$, $\mathbf{v}_1 + \mathbf{v}_2 = \mathbf{v}_2 + \mathbf{v}_1$.

Axioms for scalar multiplication:

S1 Closure For all $\mathbf{v} \in V$ and $\alpha \in F$, $\alpha\mathbf{v} \in V$.

S2 Associativity For all $\mathbf{v} \in V$ and $\alpha, \beta \in F$, $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$.

S3 Scalar multiplicative identity For all $\mathbf{v} \in V$, $1\mathbf{v} = \mathbf{v}$.

Axioms combining addition and multiplication:

D1 Distributive law For all $\mathbf{v}_1, \mathbf{v}_2 \in V$ and $\alpha \in F$, $\alpha(\mathbf{v}_1 + \mathbf{v}_2) = \alpha\mathbf{v}_1 + \alpha\mathbf{v}_2$.

D2 Distributive law For all $\mathbf{v} \in V$ and $\alpha, \beta \in F$, $(\alpha + \beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$.

Proposition 2.5 *Field extensions as vector spaces*

If E is an extension of a field F, then E is a vector space over F.

Definition 2.6 *Linear independence and basis*

Let V be a vector space over a field F.

(a) Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k \in V$. A **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a vector of the form

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$.

(b) Let $S = \{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ be a finite set of vectors in a vector space V. Then the **span** $\langle S\rangle$ of S is the set of all possible linear combinations

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k,$$

where $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$; that is,

$$\langle S\rangle = \{\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k : \alpha_1, \alpha_2, \ldots, \alpha_k \in F\}.$$

We say that the set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ **spans** $\langle S\rangle$ or is a **spanning set** for $\langle S\rangle$, and that $\langle S\rangle$ is the set **spanned** by S.

(c) A finite set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ in a vector space V is a **linearly dependent** set if there exist $\alpha_1, \alpha_2, \ldots, \alpha_k \in F$, *not all zero*, such that

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}.$$

(d) A finite set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is **linearly independent** if it is not linearly dependent; that is, if

$$\alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \cdots + \alpha_k\mathbf{v}_k = \mathbf{0}$$

only when $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$.

(e) A *basis* for a vector space V is a linearly independent set of vectors that is a spanning set for V.

(f) Let V be a vector space. Then V has **finite dimension** if it contains a finite set of vectors S that forms a basis for V. If no such set exists, then V has **infinite dimension**.

Theorem 2.7 *Basis Theorem*

Let V be a finite-dimensional vector space. Then every basis for V contains the same number of vectors.

Definition 2.8 *Degree of a field extension*

Let E be an extension of a field F. If E has a finite basis as a vector space over F, we say that the **degree** of the extension is its dimension as a vector space over F, that is, the number of vectors in any basis for the space. We denote the degree by $[E : F]$. If E does not have a finite basis as a vector space over F, then we say that the extension is infinite and we write $[E : F] = \infty$.

Theorem 2.9

The order of a finite field is a power of a prime.

Definition 2.10 $F(\alpha)$

Let E be an extension of a field F and let $\alpha \in E$. We denote by $F(\alpha)$ the smallest subfield of E (with respect to inclusion) that contains both F and α.

Definition 2.11 *Algebraic and transcendental element*

Let the field E be an extension of a field F. An element $\alpha \in E$ is said to be **algebraic over** F if there is a non-zero polynomial $h(x)$ in $F[x]$ with $h(\alpha) = 0$. An element $\alpha \in E$ is called **transcendental over** F if there is no polynomial $h(x)$ in $F[x]$ with $h(\alpha) = 0$.

Definition 2.12 $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$

Let E be an extension of a field F and let $\{\alpha_1, \alpha_2, \ldots, \alpha_n\} \subseteq E$. We denote by $F(\alpha_1, \alpha_2, \ldots, \alpha_n)$ the smallest subfield of E (with respect to inclusion) that contains both F and $\{\alpha_1, \alpha_2, \ldots, \alpha_n\}$.

Proposition 2.13

Let E be an extension of a field F and let $\alpha_1, \alpha_2 \in E$. Then $F(\alpha_1, \alpha_2) = (F(\alpha_1))(\alpha_2)$.

Definition 2.14 *Algebraic and transcendental extension*

An extension $F(\alpha)$ of a field F is called an **algebraic extension** if α is an algebraic element over F. Analogously, $F(\alpha)$ is called a **transcendental extension** if α is transcendental over F.

Proposition 2.15

Let E be an extension of a field F and let $\alpha \in E$ be algebraic over F. Let $\beta \in F(\alpha)$. Then β is algebraic over F.

Definition 2.16 *Algebraic and transcendental extension (general)*

An extension E of a field F is called an **algebraic extension** if every element in E is algebraic over F. Otherwise (that is, if E contains a transcendental element), the extension is called **transcendental**.

Theorem 2.19

Let F be a field and let $p(x)$ be an irreducible polynomial of degree n in $F[x]$. Let $E = F[x]/\langle p(x)\rangle$ and let $\alpha = x + \langle p(x)\rangle$. Let us represent E in the form $E = \{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in F\}$. Then:

(a) E is an extension of F

(b) $\{1, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$ is a basis for E as a vector space over F

(c) the degree $[E : F]$ of the field extension E over F is n

(d) $p(\alpha) = 0$

(e) $E = F(\alpha)$ and it is an algebraic extension over F.

Corollary 2.20

Let F be a field and let $f(x)$ be a non-constant polynomial in $F[x]$. Then there is an extension E of F that contains an element α such that $f(\alpha) = 0$.

Theorem 2.21

Let E be an extension of a field F and suppose that $\alpha \in E \backslash F$ is algebraic over F. Then:

(a) there is an irreducible polynomial $h(x) \in F[x]$ such that $h(\alpha) = 0$ and $F[x]/\langle h(x)\rangle \cong F(\alpha)$

(b) there is a unique irreducible monic polynomial $m(x) \in F[x]$ such that $m(\alpha) = 0$ and $F[x]/\langle m(x)\rangle \cong F(\alpha)$

(c) $[F(\alpha) : F] = \deg m(x) = \deg h(x)$.

Definition 2.22 *Minimal polynomial*

Let E be an extension of a field F and suppose that $\alpha \in E \backslash F$ is algebraic over F. A non-zero polynomial $h(x) \in F[x]$ is called the **minimal polynomial** for α if it is monic and a polynomial of lowest degree in $F[x]$ such that $h(\alpha) = 0$.

Corollary 2.23 *to Theorem 2.21*

Let E be an extension of a field F and suppose that $h(x)$ is an irreducible polynomial in $F[x]$ and that $\alpha \in E$ with $h(\alpha) = 0$. Then $F(\alpha) \cong F[x]/\langle h(x)\rangle$.

Theorem 2.27

Let t_1 and t_2 be positive rational numbers such that both $\sqrt{t_1}$ and $\sqrt{t_2}$ are irrational. Then the fields $\mathbb{Q}(\sqrt{t_1})$ and $\mathbb{Q}(\sqrt{t_2})$ are isomorphic if, and only if, there exists a rational number s such that $t_1 = s^2 t_2$.

Theorem 2.28

Let F_1, F_2 be fields. Let $\phi : F_1 \to F_2$ be a field isomorphism and let $\hat{\phi}$ be the homomorphism $\hat{\phi} : F_1[x] \to F_2[x]$ that extends ϕ to $F_1[x]$, so that

$$\hat{\phi}(a_0 + a_1 x + \cdots + a_n x^n) = \phi(a_0) + \phi(a_1)x + \cdots + \phi(a_n)x^n.$$

Let $h_1(x)$ be an irreducible polynomial over F_1 and let $h_2(x) = \hat{\phi}(h_1(x))$ be its (irreducible) image under the mapping $\hat{\phi}$. Let α_1 be a root of $h_1(x)$ and let α_2 be a root of $h_2(x)$. Then the fields $F_1(\alpha_1)$ and $F_2(\alpha_2)$ are isomorphic.

Theorem 2.29 *The KLM Theorem for field extensions*

Let K, L and M be fields such that $K \subset L \subset M$ and let the degrees $[M : L]$ and $[L : K]$ be both finite. Then

$$[M : L][L : K] = [M : K].$$

Theorem 2.30

Let E be an extension of a field F and let $\alpha \in E$ be a transcendental element over F. Then the smallest subfield $F(\alpha)$ of E containing F and α has the form

$$F(\alpha) = \{f(\alpha)(g(\alpha))^{-1} : \ f(x), g(x) \in F[x], \ g(x) \neq 0\}.$$

That is, $F(\alpha)$ is isomorphic to the field of fractions of the integral domain $F[x]$.

Theorem 2.31

For every field L, the field of fractions of the integral domain $L[x]$ is an extension of L that contains a transcendental element over L.

Corollary 2.32 *to Theorems 2.30 and 2.31*

Let E be a field extension of a field F. Suppose that $\alpha, \beta \in E$ and both α and β are transcendental elements over F. Then $F(\alpha) \cong F(\beta)$.

Finite fields

Lemma 3.1

Let F be a finite field of order p^n for some prime p and let d be a positive divisor of $p^n - 1$. Let a_d be the number of elements of $F^* = F\backslash\{0\}$ that have multiplicative order equal to d. If $a_d > 0$ then $a_d = \phi(d)$ where $\phi(d)$ is the value of the Euler ϕ-function at d.

Lemma 3.2

Let $q = p^n$ for some prime p and let

$$A = \{d : 0 < d < q \text{ and } d \text{ divides } q - 1\}.$$

Then $q - 1 = \sum_{d \in A} \phi(d)$, where $\phi(d)$ is the Euler ϕ-function.

Theorem 3.3 *Multiplicative group of a finite field*

The multiplicative group of every finite field is cyclic.

Definition 3.4 *Primitive element*

Any generator of the multiplicative group of a finite field F is called a **primitive element** of F.

Corollary 3.5 *to Theorem 3.3*

For any prime p, let F be a field of order $q = p^n$.

(a) Let ξ be a primitive element of F. Then

$$\{\xi^j : 1 \leq j < q - 1,\ j \text{ and } q - 1 \text{ are relatively prime}\}$$

is the set of all the primitive elements of F.

(b) The number of primitive elements of a finite field of order p^n is equal to $\phi(p^n - 1)$.

Definition 3.6 *Roots of unity*

Let F be a field. Then any root of the polynomial $x^k - 1$ in F is known as a kth **root of unity**. Those elements $u \in F$ that have multiplicative order k in the multiplicative group F^* are called **primitive** kth roots of unity.

Theorem 3.8

A finite field F of order a prime power $q = p^n$ contains a primitive kth root of unity if, and only if, k is a divisor of $p^n - 1$. In this case, F contains precisely $\phi(k)$ primitive kth roots of unity.

Proposition 3.9

In a finite field F of order p^n, the element -1 has no square root in F if, and only if, $p \equiv 3 \pmod 4$ and n is odd.

Splitting fields

Definition 4.1 *Splitting field*

Let $f(x)$ be a polynomial of positive degree over a field F. A **splitting field** of $f(x)$ over F is a field E that has the following properties:

(a) E is an extension of F that contains all the roots of $f(x)$

(b) if L is an extension of F that contains all the roots of $f(x)$, then E is isomorphic to a subfield of L.

Theorem 4.2

Let $f(x)$ be a polynomial of positive degree over a field F. Then there exists an extension of F that contains all the roots of $f(x)$.

Definition 4.3 *Splitting polynomial*

Let F be a field. We say that a polynomial $f(x) \in F[x]$ **splits** over F if all roots of $f(x)$ lie in F.

Proposition 4.4

Suppose that $f(x)$ is a polynomial in $K[x]$ for some field K and

1. $f(x)$ splits over an extension F of K
2. $\{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ is the set of all roots of $f(x)$.

Then $K(\alpha_1, \ldots, \alpha_d)$ is a splitting field for $f(x)$ over K and $K(\alpha_1, \ldots, \alpha_d) \subseteq F$.

Corollary 4.5 *Existence of splitting fields*

Let $f(x)$ be a polynomial of positive degree over a field F. Then there exists a splitting field of $f(x)$ over F.

Theorem 4.6 *Uniqueness of splitting fields*

Let $\theta : K \to L$ be an isomorphism of fields and let $\hat{\theta} : K[x] \to L[x]$ be the extension of θ to the polynomial ring $K[x]$ defined by

$$\hat{\theta}(a_0 + a_1 x + \cdots + a_n x^n) = \theta(a_0) + \theta(a_1)x + \cdots + \theta(a_n)x^n.$$

Let $f(x) \in K[x]$ have degree $d \geq 1$ and let $g(x) = \hat{\theta}(f(x))$ be its image over L. Let K_S, L_S be fields over which $f(x)$ and $g(x)$, respectively, split. Further, let $A = \{\alpha_1, \alpha_2, \ldots, \alpha_d\}$ be the set of all the roots of $f(x)$ in K_S and let $B = \{\beta_1, \beta_2, \ldots, \beta_d\}$ be the set of all the roots of $g(x)$ in L_S.

Then there is an isomorphism $\phi : K(\alpha_1, \ldots, \alpha_d) \to L(\beta_1, \ldots, \beta_d)$ such that

(a) the restriction $\phi_A : A \to B$ of ϕ to A is a bijection

(b) $\phi(k) = \theta(k)$ for every $k \in K$.

Corollary 4.7

Let $f(x)$ be a polynomial over a field F. Then all splitting fields for $f(x)$ are mutually isomorphic.

Theorem 4.12

Let F be a finite field of order $q = p^n$ for some prime p. Then every element of F is a root of the polynomial $x^q - x$.

Definition 4.13 *Formal derivative*

Let K be a field. Let the mapping $d : K[x] \to K[x]$ be defined by

$$d(a_0 + a_1x + a_2x^2 + \cdots + a_nx^n) = a_1 + 2a_2x + \cdots + na_nx^{n-1}.$$

We call the map d the **formal derivative**.

Lemma 4.14

Let K be a field and let $d : K[x] \to K[x]$ be the formal derivative. Then, for any two polynomials $f(x), g(x) \in K[x]$, we have

(a) $d(f(x) + g(x)) = d(f(x)) + d(g(x))$

(b) $d(f(x)g(x)) = d(f(x))g(x) + f(x)d(g(x))$.

Lemma 4.15

Let K be a field and let $f(x)$ be a non-zero polynomial in $K[x]$. Let F be the splitting field for $f(x)$ over K. Then $f(x)$ has a root $\alpha \in F$ of multiplicity $m \geq 2$ if, and only if, the highest common factor of $f(x)$ and its formal derivative $d(f(x))$ is a polynomial of positive degree.

Lemma 4.16

Let $q = p^n$ be a prime power, let $f(x) = x^q - x \in \mathbb{Z}_p[x]$, and let F be a splitting field for $f(x)$ over $\mathbb{Z}_p$. Then the polynomial $f(x)$ has no repeated roots in F.

Theorem 4.17 *Classification of finite fields*

For every prime p and every $n \geq 1$ there exists a field of order p^n, and any two fields of order p^n are isomorphic.

Corollary 4.18

For any prime p and any positive integer n there exists an irreducible polynomial of degree n over $\mathbb{Z}_p$.

Chapter 19: Fields and geometry

Field extensions revisited

Definition 1.1 *Subfield generated by a set*

Let F be a subfield of a field E and let S be a subset of elements of E. We define $F(S)$ to be the smallest subfield of E containing F and all the elements of S. If no field F is specified, then the smallest subfield of E that contains S is called the subfield of E **generated** by S.

Theorem 1.2

Let E be an extension of a field F. An element α of E is algebraic over F if, and only if, the degree $[F(\alpha) : F]$ is finite.

Theorem 1.3 *Field extensions of finite degree*

Let $F \subset E$ be fields. Then the degree $[E : F]$ is finite if, and only if, E is an algebraic extension of F and there exists a finite subset S of E such that $E = F(S)$.

Theorem 1.5

Let $F \subset E$ be fields and let $\alpha, \beta \in E$ be algebraic over F. Then both $\alpha + \beta$ and $\alpha\beta$ are elements of E that are algebraic over F.

Strategy: to find a polynomial over F that has α as a root

Let E be an extension of a field F such that $[E : F] = n$ and let α be an element of E.

1. Write down a basis $B = \{u_1, \ldots, u_n\}$ for E as a vector space over F.
2. For $j = 1, \ldots, n$, express $u_j\alpha$ as a linear combination of elements of B, that is
$$u_j\alpha = c_{j,1}u_1 + \cdots + c_{j,j}u_j + \cdots + c_{j,n}u_n.$$
3. Rearrange the equations obtained in step 2 as
$$c_{j,1}u_1 + \cdots + (c_{j,j} - \alpha)u_j + \cdots + c_{j,n}u_n = 0.$$
4. Write down the coefficient matrix A of the resulting system of n homogeneous linear equations.
5. Since $u_1, u_2, \ldots, u_n$ is a non-zero solution for this system of equations, we know that $\det A = 0$.
6. The determinant $\det A$ will be a polynomial of degree at most n in α. This is the required polynomial.

Theorem 1.7

Let E and F be two fields such that $F \subset E$, and let A be the set of all the elements of E that are algebraic over F. Then A is a subfield of E.

Definition 1.8 *Field of algebraic numbers*

Let E be an extension of a field F. The subfield of E that consists of all the elements of E that are algebraic over F is called the **field of algebraic numbers in E over F** and it is denoted by $A_E(F)$.

When $E = \mathbb{C}$ and $F = \mathbb{Q}$, this subfield is denoted by $\overline{\mathbb{Q}}$ and it is called the **field of algebraic numbers over $\mathbb{Q}$**.

Definition 1.9 *Transcendental extension*

An extension E of a field F is said to be **transcendental** if E contains at least one element that is transcendental over F.

Theorem 1.10

Let $L(x)$ be the field of fractions of a polynomial ring $L[x]$ over a field L. Suppose that $z \in L(x)$ and $z \notin L$. Then z is transcendental over L.

Theorem 1.11

Let α be a transcendental element over a field F. Then, for any non-zero relatively prime integers m and n, the element $\sqrt[n]{\alpha^m}$ is transcendental over F.

Summary – Field extensions

Let E be a field with a subfield F.

- An element $\alpha \in E$ is said to be algebraic over F if α is a root of some polynomial in $F[x]$. An element $\alpha \in E$ is called transcendental over F if α is not a root of any non-zero polynomial in $F[x]$. (Definition 2.11 of Chapter 18)
- If $\alpha \in E \backslash F$ is algebraic over F then
  - α has a minimal polynomial $m(x)$ with $m(\alpha) = 0$
  - $F(\alpha) \cong F[x]/\langle m(x) \rangle$
  - $[F(\alpha) : F] = \deg(m)$
  - $F(\alpha) = \{a_0 + a_1\alpha + \cdots + a_{n-1}\alpha^{n-1} : a_0, a_1, \ldots, a_{n-1} \in F\}$ where $n = \deg(m)$ and addition and multiplication of elements are carried out modulo $m(x)$.

  (Theorem 2.21, Theorem 2.19 and Definition 2.22 of Chapter 18)
- Let K, L and M be fields such that $K \subset L \subset M$ and let the degrees $[M : L]$ and $[L : K]$ be both finite. Then $[M : L][L : K] = [M : K]$. (Theorem 2.29 of Chapter 18)
- An element α of E is algebraic over F if, and only if, the degree $[F(\alpha) : F]$ is finite. (Theorem 1.2 of Chapter 19)
- The degree $[E : F]$ is finite if, and only if, E is an algebraic extension of F and there exists a finite subset S of E such that $E = F(S)$. (Theorem 1.3 of Chapter 19)
- If $h(x)$ is an irreducible polynomial over a field F and α and β are roots of $h(x)$, then $F(\alpha) \cong F(\beta)$. (Exercise 2.5 of Chapter 18)
- If $f(x)$ is a polynomial over F, then up to isomorphism there is a unique extension of F – the splitting field of f – that contains all the roots of f. Moreover, every extension of F over which f splits contains a subfield that is isomorphic to the splitting field. (Definition 4.1, Proposition 4.4 and Corollary 4.7 of Chapter 18)
- Given a field F, the field of rational functions over F is a transcendental extension of F. (Theorem 2.31 of Chapter 18, and Definition 1.9 of Chapter 19)
- If α, β are transcendental elements over F, then $F(\alpha) \cong F(\beta)$. (Corollary 2.32 of Chapter 18)
- If α is transcendental over F and $\beta \in F(\alpha)$ but $\beta \notin F$ then β is transcendental over F. (Theorem 1.10 of Chapter 19)
- If α is transcendental over F and m and n are relatively prime integers then $\sqrt[n]{\alpha^m}$ is transcendental over F. (Theorem 1.11 of Chapter 19)

Ruler and compasses constructions

Definition 2.1 *Constructible sets of points*

Let S be a set of points in the Euclidean plane that contains at least two points. We introduce two operations corresponding to performing an elementary 'move' using ruler and compasses:

O1 Ruler operation Draw a straight line through any two points of S.

O2 Compasses operation Draw a circle whose centre is a point of S and whose radius is equal to the distance between some pair of points in S.

The points of intersection of any two distinct lines or circles drawn using operations O1 or O2 are said to be **constructible from a set** S in one move.

We say that a point B is **constructible** from an initial set S_0 of points in the Euclidean plane if there is a finite sequence $A_1, A_2, \ldots, A_k = B$ of points in the plane such that, for $i = 1, 2, \ldots, k$, the point A_i is constructible in one move from the set $S_0 \cup \{A_1, \ldots, A_{i-1}\}$.

Definition 2.3 *Constructible number*

(a) We say that a real number a is **constructible** if the point $(a, 0)$ is constructible from the set of points $\{(0, 0), (1, 0)\}$.

(b) Let S be a set of points in the plane that contains at least two points. We first determine a coordinate system. Choose two points in S and let one point be the origin, $(0, 0)$ and the other point be the point $(1, 0)$. We say that a real number a is **constructible from a set** S, with the given coordinate system, if the point $(a, 0)$ is constructible from the set of points S.

Proposition 2.4

The set of constructible numbers from a set S containing at least two points forms a subfield of $\mathbb{R}$.

Lemma 2.5

Let S be a set of at least two points in the plane and let K be the subfield of $\mathbb{R}$ generated by the x- and y-coordinates of the points in S. Let $A = (x, y)$ be a point constructed in one move from the set S and let $L = K(x, y)$ be the subfield of $\mathbb{R}$ obtained by adjoining x and y to K. Then the degree $[L : K]$ is either 1 or 2.

Theorem 2.6

Suppose that the point $A = (x, y)$ is constructible from a set S of at least two points in $\mathbb{R}^2$. Let E be the subfield of $\mathbb{R}$ generated by the coordinates of the points in S. Then the degree $[E(x, y) : E]$ is a power of 2.

Theorem 2.7

Let S be a set of points in the real plane $\mathbb{R}^2$ and let E be the subfield of $\mathbb{R}$ generated by the coordinates of all the points in S. Suppose that $A = (x, y)$ is a point in $\mathbb{R}^2$ such that the degree of the extension $[E(x, y) : E]$ is not a power of 2. Then the point A is not constructible from S.

Corollary 2.8

Let S be a set of points in the real plane $\mathbb{R}^2$ and let E be the subfield of $\mathbb{R}$ generated by the coordinates of all the points in S. Let $A = (x, 0)$ be a point in $\mathbb{R}^2$ such that the minimal polynomial for x over the field E has a degree that is not a power of 2. Then x is not a constructible number.

Theorem 2.9

The cube cannot be duplicated using ruler and compasses only.

Theorem 2.10

The angle $\pi/3$ cannot be trisected by a construction using ruler and compasses only.

Theorem 2.11

The quadrature of a circle cannot be constructed by using ruler and compasses only.

Theorem 2.12

Let S be a set of complex numbers (including 1 and 0) and let E be the subfield of $\mathbb{C}$ generated by S. Suppose that $a + ib$ is a complex number such that the degree of the extension $[E(a + ib) : E]$ is not a power of 2. Then $a + ib$ is not constructible from S.

Corollary 2.13

Let K be the field of constructible complex numbers. If the degree of the minimal polynomial for the complex number $a + ib$ over $\mathbb{Q}$ is not a power of 2, then $a + ib \notin K$.

Proposition 2.14

Let p be a prime number and let x be a root of $x^p - 1$ with $x \neq 1$. If x is constructible, then p is of the form $2^n + 1$.

Chapter 20: Public-key cryptography

Cryptosystems based on modular arithmetic

Lemma 1.3

Let $n = pq$, where p and q are distinct prime numbers, and let e be an integer such that $1 \leq e < \phi(n)$ and e is relatively prime to $\phi(n)$. If d is the multiplicative inverse of e modulo $\phi(n)$, then

$$x^{de} \equiv x \,(\text{mod } n)$$

for all integers x.

Lemma 1.4

Suppose that p is an odd prime and that $p - 1 = 2^r s$, where s is odd and $r \geq 1$. Then, for every b such that $1 < b < p$, either $b^s \equiv 1 \,(\text{mod } p)$, or there exists a t, $0 \leq t \leq r - 1$, such that $b^{2^t s} \equiv -1 \,(\text{mod } p)$.

Corollary 1.5

Suppose that k is an odd positive integer, $k \geq 3$, and that $k - 1 = 2^r s$, where s is odd and $r \geq 1$. If there exists an integer b, $1 < b < k$, such that $b^s \not\equiv 1 \pmod{k}$ and $b^{2^t s} \not\equiv -1 \pmod{k}$ for all t, $0 \leq t \leq r - 1$, then k is not a prime number.

Such an integer b is said to be a **witness of non-primality** for k.

Lemma 1.7

Let p be a prime. A non-zero element $\xi \in GF_p$ is a primitive element of GF_p if, and only if, $\xi^{(p-1)/q} \not\equiv 1 \pmod{p}$ for all prime divisors q of $p - 1$.

Lemma 1.9 *Euler's Criterion*

Let p be an odd prime and let b be a non-zero element of GF_p. Then b is a square in GF_p if, and only if, $b^{(p-1)/2} \equiv 1 \pmod{p}$, and b is a non-square in GF_p if, and only if, $b^{(p-1)/2} \equiv -1 \pmod{p}$.

Moreover, if $p \equiv -1 \pmod{4}$ and if b is a square in GF_p, then the two solutions of the equation $x^2 = b$ are $x = \pm b^{(p+1)/4}$.

Definition 1.11 *RSA triple, public key, private key*

Let $n = pq$ be a product of two distinct primes p, q and let e be a positive integer such that $1 < e < \phi(n)$ and e is relatively prime to $\phi(n)$. Let d be the unique positive integer such that $1 < d < \phi(n)$ and $de \equiv 1 \pmod{\phi(n)}$. Then:

(a) the triple (p, q, e) is called an **RSA triple**

(b) the pair (n, e) is the **public key**

(c) the integer d is the **private key**.

RSA encryption

Given the public key (n, e) and plaintext message m $(1 < m < n - 1)$, the ciphertext c is given by $c \equiv m^e \pmod{n}$.

RSA decryption

Given the public key (n, e), the ciphertext $c \equiv m^e \pmod{n}$ and the private key d, the plaintext message m is given by $c^d \pmod{n} \equiv m \pmod{n}$.

RSA eavesdropping?

Given the public key (n, e) and cyphertext $c \equiv m^e \pmod{n}$, the primes p and q (such that $n = pq$) are also needed to determine $\phi(n)$ and hence to find the inverse d of e and discover the message m.

However, in general, factorising n into prime factors is not feasible, provided the primes p and q are sufficiently large.

Lemma 1.13 *Witness of non-primality*

Let k be an odd positive integer, $k \geq 3$, and let $k - 1 = 2^r s$ where r and s are positive integers and s is odd. If there exists an integer b, $1 < b < k$, such that $b^s \not\equiv 1 \pmod{k}$ and $b^{2^t s} \not\equiv -1 \pmod{k}$ for all t, $0 \leq t \leq r - 1$, then k is not a prime number.

The Rabin–Miller test

To test a given positive integer k for primality, we choose a positive integer j. We then randomly generate j distinct integers b such that $1 < b < k$. If $S(b)$ is false for all the j values of b, we accept k as a strong probable prime, and then the probability that k is a prime is at least $1 - 4^{-j}$.

The integers k such that $S(b)$ is false for j values of b are called **strong probable primes to base** b.

Diffie–Hellman cryptosystem setup

Given the **public key** (p, ξ), where p is a prime and $\xi \in GF_p$ a primitive element, the two communicating parties each choose a **private key**, positive integers a and b, respectively, such that $1 < a < p - 1$ and $1 < b < p - 1$.

They then compute $\alpha \equiv \xi^a \pmod{p}$ and $\beta \equiv \xi^b \pmod{p}$, respectively, and exchange these numbers.

Finally, they each compute the **shared secret** σ and determine σ^{-1}, where $\sigma = \beta^a = \alpha^b = \xi^{ab}$, and hence $\sigma^{-1} = \alpha^{p-1-b} = \beta^{p-1-a}$.

Diffie–Hellman encryption and decryption

The information exchange can go both ways:

- Given the public key (p, ξ), message m $(1 < m < p - 1)$ and the shared secret σ, the cyphertext c is given by $c \equiv m\sigma \pmod{p}$.
- Given the public key (p, ξ), cyphertext c and shared secret σ, the message m is given by $c\sigma^{-1} \equiv m \pmod{p}$.

Diffie–Hellman security

Given the public key (p, ξ), numbers $\alpha \equiv \xi^a \pmod{p}$ and $\beta \equiv \xi^b \pmod{p}$, and the cyphertext $c \equiv m\sigma \pmod{p}$, the secret keys a and b are also needed to determine σ and hence σ^{-1} to find the message m.

However, in general, determining discrete logarithms is not feasible, provided the prime p is sufficiently large; that is, determining the exponent a from $\alpha \equiv \xi^a \pmod{p}$, and likewise b from $\beta \equiv \xi^b \pmod{p}$, is not feasible.

Definition 1.15 *Discrete logarithm in finite fields*

Let $\alpha \neq 0$ and $\beta \neq 0, 1$ be elements of a finite field GF_p. If there exists an integer a such that $0 \leq a < p - 1$ and $\beta^a = \alpha$, then a is called the **discrete logarithm of** α **to the base** β. In such a case we write $a = \log_\beta \alpha$.

Cryptosystems based on elliptic curves

Lemma 2.1

Let $f(x) = x^3 + ax + b$ be a polynomial over an arbitrary field F. If $4a^3 + 27b^2 \neq 0$, then $f(x)$ has no multiple roots.

Lemma 2.2

Let $f(x) = x^3 + ax + b$ be a cubic polynomial over $\mathbb{R}$ such that $4a^3 + 27b^2 \neq 0$. Then $f(x)$ has three real roots if $a < 0$ and $|b| < \frac{2|a|\sqrt{|a|}}{3\sqrt{3}}$, and exactly one real root otherwise.

Definition 2.3 *Elliptic curve*

Let F be a field of characteristic distinct from 2 and 3, and let $f(x) = x^3 + ax + b$ be a cubic polynomial over F such that $4a^3 + 27b^2 \neq 0$. Let $\mathcal{O}$ be a symbol, distinct from all elements of $F \times F$. The set

$$E_F(a, b) = \{\mathcal{O}\} \cup \{(x, y) \in F \times F : \ y^2 = f(x)\}$$

is called an **elliptic curve** over the field F.

Elements of an elliptic curve $E_F(a, b)$ are usually referred to as **points** on the curve and the special symbol $\mathcal{O}$ a **point at infinity**.

Definition 2.8 *Line with slope*

Let F be a field and let $\mathcal{O} \notin F \times F$ be the point at infinity.

(a) For any chosen $c, d \in F$, a **line with slope** c is the set

$$L(c, d) = \{(x, y) \in F \times F : y = cx + d\}.$$

(b) For any $x_0 \in F$, a **slopeless line** is the set

$$L(x_0) = \{\mathcal{O}\} \cup \{(x, y) \in F \times F : x = x_0\}.$$

A **line over** F refers to both lines with slope and slopeless lines.

Definition 2.9 *Tangent line with slope*

Let $E_F(a, b)$ be an elliptic curve over a field F of characteristic distinct from 2 and 3, and let $P = (x_0, y_0)$ be a point on this curve.

(a) If $y_0 \neq 0$, the line $L(c_0, d_0)$, where

$$c_0 = \frac{3x_0^2 + a}{2y_0} \quad \text{and} \quad d_0 = \frac{-x_0^3 + ax_0 + 2b}{2y_0},$$

is said to be the **tangent line with slope** c_0 to the curve at P.

(b) If $y_0 = 0$, the line $L(x_0)$ is said to be the **slopeless tangent line** to $E_F(a, b)$ at P.

A **tangent line** refers to both tangent lines with slope and slopeless tangent lines.

Lemma 2.10

Let $E_F(a, b)$ be an elliptic curve over a field F of characteristic distinct from 2 and 3.

(a) Let $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ be points on $E_F(a, b)$ distinct from $\mathcal{O}$ and such that $P_1 \neq \pm P_2$. Then the (unique) line through P_1 and P_2 is $L(c, d)$, where

$$c = \frac{y_2 - y_1}{x_2 - x_1} \quad \text{and} \quad d = \frac{y_1 x_2 - y_2 x_1}{x_2 - x_1}.$$

This line intersects the elliptic curve at the point $\pi(P_1, P_2) = (x_3, y_3)$, where $x_3 = c^2 - x_1 - x_2$ and $y_3 = cx_3 + d$.

(b) Let $P_0 = (x_0, y_0)$ be a point on $E_F(a, b)$ such that $y_0 \neq 0$. Then the tangent line $L(c_0, d_0)$, where c_0 and d_0 are as in Definition 2.9, intersects the curve $E_F(a, b)$ in the point $\pi'(P_0) = (x', y')$, where $x' = c_0^2 - 2x_0$ and $y' = c_0 x' + d_0$.

Definition 2.12 *Addition of points on an elliptic curve*

Let F be a field of characteristic distinct from 2 and 3, and let $E_F(a, b)$ be an elliptic curve over F. Let P, P_0, P_1 and P_2 be arbitrary points on $E_F(a, b)$ distinct from $\mathcal{O}$, with $P_1 \neq P_2$. We define a binary operation on $E_F(a, b)$ as follows:

(a) $\mathcal{O} + \mathcal{O} = \mathcal{O}$

(b) $P + \mathcal{O} = \mathcal{O} + P = P$

(c) $P + (-P) = \mathcal{O}$

(d) If P_0 has non-zero y-coordinate, then $2P_0 = P_0 + P_0 = -\pi'(P_0)$

(e) If P_1 and P_2 are such that $P_1 \neq \pm P_2$, then $P_1 + P_2 = -\pi(P_1, P_2)$.

Proposition 2.14

Let F be a field of characteristic distinct from 2 and 3, let $E = E_F(a, b)$ be an elliptic curve over F, and let $+$ be the binary operation on E introduced in Definition 2.12. Then $(E, +)$ is an abelian group with identity element $\mathcal{O}$. The inverse of a point P is the point $-P$.

Notation

For any non-negative integer r the symbol $\mathbb{Z}^r$ denotes the trivial group if $r = 0$, and the direct product

$$\underbrace{\mathbb{Z} \times \cdots \times \mathbb{Z}}_{r \text{ times}} \quad \text{if} \quad r \geq 1.$$

Theorem 2.19

Let E be an elliptic curve over the field of rational numbers. Then the group $(E, +)$ is isomorphic to a group of the form $A \times \mathbb{Z}^r$ for some $r \geq 0$, where A is a finite abelian group isomorphic to $\mathbb{Z}_n$ for some $n \in \{1, 2, \ldots, 9, 10, 12\}$, or to $\mathbb{Z}_2 \times \mathbb{Z}_{2m}$ for some $m \in \{1, 2, 3, 4\}$.

Theorem 2.20

Let E be an elliptic curve over a finite field of order q. Then the group $(E, +)$ is either cyclic, or isomorphic to a product $\mathbb{Z}_k \times \mathbb{Z}_m$ for some $k, m \geq 2$ such that k is a divisor of m. In either case, the order e_q of the group $(E, +)$ satisfies the inequality

$$|e_q - (q + 1)| \leq 2\sqrt{q}.$$

Diffie–Hellman–ElGamal cryptosystem setup

Given an elliptic curve $E = E_{GF_p}(a, b)$ and initial point $P \in E$, so p is a prime and $a, b \in GF_p$ such that $4a^3 + 27b^2 \neq 0$, the two communicating parties each choose a **private key** s_A and s_B, respectively, such that $1 < s_A < |E| - 1$ and $1 < s_B < |E| - 1$.

They then calculate the points $P_A = s_A P \in E$ and $P_B = s_B P \in E$, respectively, and exchange these points.

Finally, they each compute the **shared secret** the point Q where $Q = s_A P_B = s_B P_A$.

The **public key** in this system is the quadruple (E, P, P_A, P_B).

Diffie–Hellman–ElGamal encryption

First, the message m is converted to a point M on the elliptic curve.

Given the public key (E, P, P_A, P_B), message point M and a randomly generated number k such that $1 < k < |E|$, the two points $M_1 = kP$ and $M_2 = M + kP_B$ in the group $(E, +)$ are the ciphertext.

Diffie–Hellman–ElGamal decryption

Given the public key (E, P, P_A, P_B), the two points M_1 and M_2 (the cyphertext) and the private key s_B, the plaintext message M is given by $M = M_2 - s_B M_1$ in the group $(E, +)$.

Note that information exchange can go both ways, using s_A and s_B respectively.

Diffie–Hellman–ElGamal security

Given the public key (E, P, P_A, P_B) and the points M_1 and M_2, the secret keys s_A and s_B are also needed to determine the point M and hence to discover the message m.

However, in general, attempts to determine s_A and s_B from the knowledge of the points P, $P_A = s_A P$ and $P_B = s_B P$ are hopeless, provided the prime p is sufficiently large.

Menezes–Vanstone cryptosystem setup

As for the Diffie–Hellman–ElGamal cryptosystem, the **public key** is the quadruple (E, P, P_A, P_B) and the **private keys** are s_A and s_B, respectively.

Menezes–Vanstone encryption

Here, the message m takes the form of an ordered pair (m_1, m_2), where $m_1, m_2 \in \mathbb{Z}_p$.

Given the public key (E, P, P_A, P_B), the message (m_1, m_2) and a randomly generated number k, $1 < k < |E|$, compute the two points $kP \in E$ and $kP_B \in E$. Let the coordinates of the second point be $kP_B = (d_1, d_2)$ and compute $c_1 = d_1 m_1 \in \mathbb{Z}_p$ and $c_2 = d_2 m_2 \in \mathbb{Z}_p$. The ciphertext is the triple (kP, c_1, c_2).

Menezes–Vanstone decryption

Given the public key (E, P, P_A, P_B), the cyphertext (kP, c_1, c_2) and the private key s_B, compute the $(d_1, d_2) = s_B(kP) = k(s_B P) = kP_B$. Then determine d_1^{-1} and d_2^{-1} modulo p. The plaintext message (m_1, m_2) is given by $m_1 = c_1 d_1^{-1}$ and $m_2 = c_2 d_2^{-1}$, all modulo p.

Note that information exchange can go both ways, using s_A and s_B respectively.

Menezes–Vanstone comment

The number of messages that can be sent is far greater than for the Diffie–Hellman–ElGamal cryptosystem, as messages are ordered pairs of elements of $\mathbb{Z}_p$, rather than points on an elliptic curve over GF_p.

Definition 2.23 *Discrete logarithm for elliptic curves*

Let $E = E_{GF_p}(a, b)$ be an elliptic curve over a Galois field GF_p. Let $P, P' \neq \mathcal{O}$ be two points on E, regarded as elements of the abelian group $(E, +)$. If there exists an integer s, $1 < s < |E|$, such that $P' = sP$, then s is called the **discrete logarithm of P' to the base P**.

Book F: Metric spaces 2

Chapter 21: Connectedness

Homeomorphisms

Definition 1.1 *Homeomorphism*

Let (X, d) and (Y, e) be two metric spaces. Then (X, d) and (Y, e) are **homeomorphic** metric spaces if there exists a bijection $f\colon X \to Y$ such that $f\colon X \to Y$ is (d, e)-continuous and $f^{-1}\colon Y \to X$ is (e, d)-continuous. We say that f is a **homeomorphism**.

Proposition 1.2 *Restriction Rule for homeomorphisms*

Let $h\colon X \to Y$ be a homeomorphism between two metric spaces (X, d) and (Y, e), and let $X' \subseteq X$. Then $h|_{X'}$, the restriction of h to X', gives a homeomorphism from X' to the image $h(X') \subseteq Y$ so that $(X', d_{X'})$ and $(h(X'), e_{h(X')})$ are homeomorphic metric spaces.

Theorem 1.8

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$. Then f is (d, e)**-continuous** on X if, and only if, for each e-closed set $E \subseteq Y$ $f^{-1}(E)$ is d-closed.

Proposition 1.10

Let (X, d_0) be any space with the discrete metric. Let (Y, d) be any metric space and $f\colon X \to Y$ be any function. Then f is (d_0, d)-continuous.

Definition 1.12 *Topological invariant*

We say that a property of metric spaces is a **topological invariant** if it is preserved by homeomorphisms. That is, if two spaces are homeomorphic and one has the property, then so must the other.

Closed and open sets revisited

Theorem 2.1

Let (X, d) be a metric space and let $A \subseteq X$.

A set $U \subseteq A$ is d_A-open if, and only if, there is a set E in X that is d-open and for which $U = A \cap E$.

Characterisations of continuity

Let (X, d) and (Y, e) be metric spaces and let $f\colon X \to Y$ be a function. The following are equivalent ways of defining continuity of f:

- *The sequential characterisation*

  The function f is (d, e)-**continuous** at $a \in X$ if, and only if, whenever (a_k) is a sequence in X for which $a_k \xrightarrow{d} a$ as $k \to \infty$, then the sequence $f(a_k) \xrightarrow{e} f(a)$ as $k \to \infty$.
- *The 'pre-image of closed is closed' characterisation*

  The function f is (d, e)-**continuous** if, and only if, $f^{-1}(U)$ is a d-closed set in (X, d) whenever U is an e-closed set in (Y, e).

Characterisations of closed sets

Let (X, d) be a metric space and let A be a subset of X.

- The set A is a d-closed subset of X if, and only if, every d-convergent sequence with all its terms in A has its limit in A.
- The set A is d-closed if, and only if, the complement, $A^c = X - A$, is d-open (that is, if A^c contains a d-open ball around each of its points).

Characterisation of closed sets for the induced metric

Let (X, d) be a metric space. Let E be a subset of A where A is a subset of X, and let d_A be the metric on A induced by the metric d on X.

- The subset E of A is d_A-closed considered as a subset of (A, d_A) if, and only if, there is a d-closed subset F of X with $E = A \cap F$.

Connectedness

Definition 3.1 *Disconnection*

Let (X, d) be a metric space. A **disconnection** $\{U, V\}$ of X is a pair of disjoint non-empty subsets, U and V, each of which is simultaneously d-closed and d-open, with $X = U \cup V$. We say that the space (X, d) is **disconnected** if X has a disconnection. It is **connected** if X has no disconnection.

When we wish to emphasise the metric, we say *d-disconnection*, *d-connected* and *d-disconnected*.

Both U and V must be proper non-empty subsets of X.

Theorem 3.2

The metric space (X, d) is connected if, and only if, the only subsets of X which are simultaneously d-closed and d-open are $\emptyset$ and X.

Proposition 3.4

The following statements are equivalent for a metric space (X, d).

(a) (X, d) is disconnected.

(b) X can be expressed as the disjoint union of two non-empty d-closed sets.

(c) X can be expressed as the disjoint union of two non-empty d-open sets.

Lemma 3.6

Let (X, d) and (Y, d') be metric spaces, with $f : X \to Y$ a (d, d')-continuous map. If (X, d) is d-connected then $\big(f(X), d'_{f(X)}\big)$ is a $d'_{f(X)}$-connected metric space. In other words, the continuous image of a connected metric space is connected.

Corollary 3.7

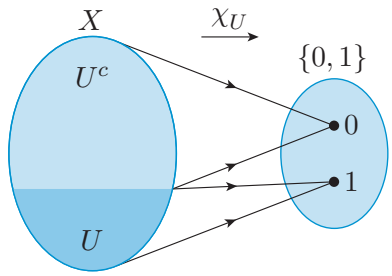
Connnectedness is preserved by homeomorphism.

Using a continuous function to characterise disconnectedness

Definition 3.8 *Characteristic function*

Let (X, d) be a metric space. If $U \subseteq X$, then the **characteristic function** of U is the function $\chi_U : X \to \{0, 1\}$ defined by

$$\chi_U(x) = \begin{cases} 1, & \text{if } x \in U, \\ 0, & \text{if } x \in U^c. \end{cases}$$



Characteristic function

Theorem 3.9

Let (X, d) be a metric space. Then X is d-connected if, and only if, every (d, d_0)-continuous function $f : X \to (Y, d_0)$, where $Y = \{0, 1\}$ and d_0 is the discrete metric, is constant.

A consequence is that if X is connected, then no function from X to $\{0, 1\}$ can be both continuous and *onto*.

Connectedness and closure

Lemma 3.10

Let (X, d) be a metric space and let A be a non-empty subset of X. Suppose that V is a non-empty d-open subset of $\mathrm{Cl}_{(X,d)}(A)$. Then $A \cap V$ is non-empty.

Definition 3.11 *Connected subspace*

Let (X, d) be a metric space with $A \subseteq X$. Then A is a **connected subspace** of X if (A, d_A) is connected when viewed as a metric space in its own right.

Theorem 3.12

Let (X, d) be a metric space and let A be a d_A-connected subset of X. Write $\mathrm{Cl}(A)$ for $\mathrm{Cl}_{(X,d)}(A)$. Suppose that

$$A \subseteq B \subseteq \mathrm{Cl}(A).$$

Then B is d_B-connected.

In particular, $\mathrm{Cl}(A)$ is $d_{\mathrm{Cl}(A)}$-connected.

Components

Definition 3.13 *Connected component*

Let (X, d) be a metric space and let $x \in X$.

The **d-component** $C_d(x)$ of x in X is the largest d-connected subset of X that contains x.

Definition 3.14 *Totally disconnected*

X is **totally d-disconnected** if, for each point $x \in X$, the d-component of x is the set $\{x\}$.

When we do not wish to emphasise the metric, we say that X is totally disconnected.

Lemma 3.15

Let (X, d) be a metric space, and let $\{A_i : i \in I\}$ be a family of subsets of X and write d_{A_i} for the metric on A_i induced by d. Assume that the intersection of the family, $\bigcap_{i\in I} A_i$, is non-empty and that for each $i \in I$, A_i is d_{A_i}-connected. Then $A = \bigcup_{i\in I} A_i$ is d_A-connected (where d_A is the metric on A induced by the metric d on X).

Theorem 3.16

Let (X, d) be a metric space and let $x \in X$. Let

$$\mathcal{F}_x = \{A \subseteq X : x \in A, \text{ and } A \text{ is } d_A\text{-connected}\}.$$

Then $C_d(x)$, the d-component of x, exists and is given by

$$C_d(x) = \bigcup_{A\in\mathcal{F}_x} A.$$

Theorem 3.20

Let (X, d) be a metric space. Let us write $x \sim y$ to mean that $y \in C_d(x)$. Then $\sim$ is an equivalence relation on X. Consequently, X is the union of its mutually disjoint d-components.

Corollary 3.21

Let (X, d) be a metric space. For each $y \in C_d(x)$, $C_d(y) = C_d(x)$. In other words, the component of each point in $C_d(x)$ is $C_d(x)$ itself.

The components are closed

Lemma 3.22

Let (X, d) be a metric space.

(a) If $\{U, V\}$ is a disconnection of a metric space (X, d), then

$$\mathrm{Cl}_{(X,d)}(U) \cap V = \emptyset \text{ and } U \cap \mathrm{Cl}_{(X,d)}(V) = \emptyset.$$

(b) If there is a pair of non-empty sets U and V for which

$$U \cup V = X,\ \mathrm{Cl}_{(X,d)}(U) \cap V = \emptyset \text{ and } U \cap \mathrm{Cl}_{(X,d)}(V) = \emptyset,$$

then U and V are d-closed and form a disconnection of X.

Lemma 3.23

Let (X, d) be a metric space and let $x \in X$. Then $C_d(x)$, the component of x in (X, d), is d-closed.

Connectedness in Euclidean space

Lemma 4.1

Let (X, d) be a metric space. A d-closed subset of X contains all of its d-closure points.

Theorem 4.2

$\mathbb{R}$ with the Euclidean metric is connected.

Proposition 4.3

Let $a < b$ be real numbers. The interval (a, b) is a connected subset of $\mathbb{R}$ (with the Euclidean metric).

Theorem 4.4

A subset A of $\mathbb{R}$ is connected for the Euclidean metric if, and only if, it is an interval.

Corollary 4.5

$\mathbb{R} - \{a\}$ is not homeomorphic to $\mathbb{R}$ for the Euclidean metrics, for each $a \in \mathbb{R}$.

Theorem 4.6 *Intermediate Value Theorem*

Let (X, d) be a connected metric space, and let $f\colon X \to \mathbb{R}$ be a $(d, d^{(1)})$-continuous function. Let a and b be points of X. Then f takes each value between $f(a)$ and $f(b)$.

Homeomorphic subsets of $\mathbb{R}$

There are five distinct classes of homeomorphic connected subsets of $\mathbb{R}$:

- the empty set $\emptyset$
- the singleton sets $\{a\} = [a, a]$
- the open intervals (a, b) for $a < b$, (a, ∞) and $(-\infty, a)$ for $a \in \mathbb{R}$, and $\mathbb{R}$
- the closed intervals $[a, b]$ for $a < b$
- the intervals of the form $[a, b)$ and $(a, b]$ for $a < b$ and $[a, \infty)$ and $(-\infty, a]$.

Intervals in the same class are homeomorphic; intervals in different classes are not homeomorphic.

Products of connected spaces

Theorem 4.8

Let (X, d_1) and (Y, d_2) be non-empty metric spaces. The product space $(X \times Y, e)$, where e is a product metric, is e-connected if, and only if, (X, d_1) and (Y, d_2) are both connected (for their respective metrics).

Corollary 4.9

Euclidean space, $(\mathbb{R}^n, d^{(n)})$, is connected.

Corollary 4.10

The n-dimensional 'rectangles' $I_1 \times \cdots \times I_n$ are connected for the Euclidean metric, where I_i is an interval in $\mathbb{R}$.

Path-connected spaces

Definition 5.1 *Path*

Let (X, d) be a metric space and let $[0, 1]$ have the induced Euclidean metric $d^{(1)}_{[0,1]}$.

A function $p\colon [0, 1] \to X$ is a **path** in X if p is a $(d^{(1)}_{[0,1]}, d)$-continuous function.

A path p **goes from** a **to** b in X if p is a path in X and if there are t_x, $t_y \in [0, 1]$ with $t_x \leq t_y$, $p(t_x) = a$ and $p(t_y) = b$. If the order in which the points are reached is unimportant, then we say that p **joins** a and b.

The **initial point** of a path p is the point $p(0)$ in X and the **final point** of p is the point $p(1)$.

The path is a **closed path** if the initial and final points of a path are the same (that is, $p(0) = p(1)$).

The **trail** of a path is its image set $p([0, 1])$.

For each point x in a metric space (X, d), there is always a path joining a point x to itself – namely, the constant function $p\colon [0, 1] \to X$ given by $p(t) = x$.

Lemma 5.3

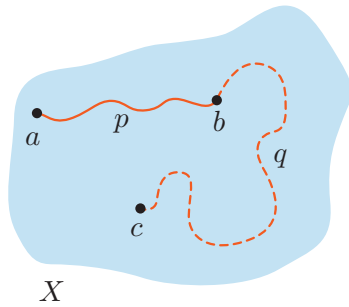
Suppose we have maps $f\colon [a,b] \to \mathbb{R}$ and $g\colon [b,c] \to \mathbb{R}$, with $f(b) = g(b)$ and where f and g are continuous with respect to the metrics induced by the Euclidean metric on $\mathbb{R}$.
Then $h\colon [a,c] \to \mathbb{R}$ defined by:

$$h(t) = \begin{cases} f(t), & \text{for } a \leq t \leq b, \\ g(t), & \text{for } b \leq t \leq c, \end{cases}$$

is also continuous with respect to the metrics induced by the Euclidean metric on $\mathbb{R}$.

Lemma 5.4

Let (X,d) be a metric space and let $a,b,c \in X$. Suppose that there is a path p in X with initial point a and final point b, and a path q in X with initial point b and final point c. Then there is a path that joins a and c in X.



Definition 5.5 *Path-connected*

Let (X,d) be a metric space.

The space (X,d) is **path-connected** if, for each a and b in X, there is a path in X that joins a and b.

A set $A \subseteq X$ is **path-connected** if (A, d_A) is path-connected. (Here d_A denotes the induced metric for A.)

Proposition 5.7

The metric space $\mathbb{R}^n$ with the Euclidean metric is path-connected.

Theorem 5.8

Let (X,d_X) and (Y,d_Y) be metric spaces, let X be path-connected, and let $f\colon X \to Y$ be (d_X,d_Y)-continuous. Then $f(X)$ is path-connected.

Corollary 5.9

Path-connectedness is a topological invariant.

Theorem 5.11

Every $d^{(n)}$-open ball in $\mathbb{R}^n$ is path-connected. Every $d^{(n)}$-closed ball in $\mathbb{R}^n$ is path-connected.

Theorem 5.13

Let (X,d) be a metric space. If X is d-path-connected, then X is d-connected.

Corollary 5.14

The space $\big(C[0,1], d_{\max}\big)$ is connected.

Theorem 5.16

Let (X,d_X) and (Y,d_Y) be non-empty metric spaces. The product space $(X \times Y, e)$, where e is a product metric, is path-connected if, and only if, (X,d_X) and (Y,d_Y) are both path-connected.

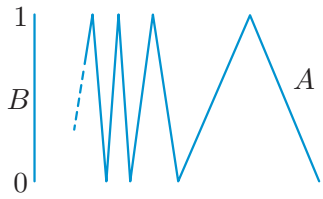
Theorem 5.17

Let (X, d) be a metric space, and let $\{A_i : i \in I\}$ be a family of path-connected subsets of X whose intersection is non-empty. Then $A = \bigcup_{i \in I} A_i$ is path-connected.

The topologist's cosine

Let $l : (0, 1] \to \mathbb{R}$ be given by

$$l(x) = \begin{cases} 2n(n+1)x - 2n, & \text{if } \frac{1}{n+1} < x \leq \frac{1}{2}\left(\frac{1}{n} + \frac{1}{n+1}\right),\ n \in \mathbb{N}, \\ 2(n+1) - 2n(n+1)x, & \text{if } \frac{1}{2}\left(\frac{1}{n} + \frac{1}{n+1}\right) < x \leq \frac{1}{n},\ n \in \mathbb{N}. \end{cases}$$



The topologist's cosine, $C = A \cup B$

Let A be the subset of the plane that is the graph of the function l, so

$$A = \{(x, l(x)) : 0 < x \leq 1\},$$

let $B = \{(0, y) : 0 \leq y \leq 1\}$ and define the topologist's cosine to be the set C given by $C = A \cup B$.

Corollary 6.2

A is path-connected for d_A, the metric induced by the Euclidean metric on the plane.

Theorem 6.3

C is d_C-connected, where d_C is the metric induced by the Euclidean metric on the plane.

Theorem 6.4

C is not path-connected for the Euclidean metric.

Connectedness of standard sets and spaces

Set or metric space	Connected	Path-connected
$\emptyset$ (any metric)	yes	yes
singleton set $\{a\}$ (any metric)	yes	yes
intervals with Euclidean metric	yes	yes
$(\mathbb{R}^n, d^{(n)})$	yes	yes
$B_{d^{(n)}}(\mathbf{a}, r)$, $B_{d^{(n)}}[\mathbf{a}, r]$, $\mathbf{a} \in \mathbb{R}^n$	yes	yes
$(\mathbb{R}^2 - A, d^{(2)}_{\mathbb{R}^2 - A})$, A countable	yes	yes
$(C[0,1], d_{\max})$	yes	yes
topologist's cosine	yes	no
$(\mathbb{R} - \{a\}, d^{(1)}_{\mathbb{R}-\{a\}})$, $a \in \mathbb{R}$	no	no
(X, d_0), X any set with at least two elements	no, totally disconnected	no
$(\mathbb{Q}, d^{(1)}_{\mathbb{Q}})$	no, totally disconnected	no
$(\mathbb{R} - \mathbb{Q}, d^{(1)}_{\mathbb{R}-\mathbb{Q}})$	no, totally disconnected	no
$(K_C, d^{(1)}_{K_C})$	no, totally disconnected	no
$(\mathbf{C}, d_{\mathbf{C}})$	no, totally disconnected	no

Chapter 22: Compactness

Two important theorems

Lemma 1.1

Let (a_n) be a sequence of real numbers. Then (a_n) contains a monotonic subsequence – that is, there is a subsequence (a_{n_k}) such that either:

$$a_{n_1} \leq a_{n_2} \leq a_{n_3} \leq \cdots \leq a_{n_k} \leq \cdots$$

or

$$a_{n_1} \geq a_{n_2} \geq a_{n_3} \geq \cdots \geq a_{n_k} \geq \cdots .$$

Theorem 1.2

Let X be any $d^{(1)}$-closed and bounded subset of $\mathbb{R}$. Then every sequence in X contains a $d^{(1)}_X$-convergent subsequence.

Theorem 1.3 *Boundedness Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a $(d^{(1)}_{[a,b]}, d^{(1)})$-continuous function on $[a,b]$. Then f is bounded on $[a,b]$. That is, there is a real number M such that $|f(x)| \leq M$ for each $x \in [a,b]$.

Theorem 1.4 *Extreme Value Theorem*

Let $f\colon [a,b] \to \mathbb{R}$ be a $(d^{(1)}_{[a,b]}, d^{(1)})$-continuous function on $[a,b]$. Then there are numbers $c, d \in [a,b]$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for each } x \in [a,b].$$

Sequential compactness

Definition 2.1 *d-sequentially compact*

A metric space (X,d) is **d-sequentially compact** if each sequence in X has a d-convergent subsequence (which is d-convergent to a point in X).

A subset $A \subseteq X$ is a **sequentially compact subset** if (A, d_A) is sequentially compact as a metric space.

If the metric is clear from the context, then we suppress mention of the metric and say that X is *sequentially compact*.

If $A \subseteq X$ is a sequentially compact subset, then we sometimes refer to A as d_A-sequentially compact.

Fact

Any d-sequentially compact subset of a metric space (X,d) is d-closed.

Corollary 2.2

Any $d^{(1)}$-closed and bounded subset of $\mathbb{R}$ is sequentially compact.

Theorem 2.3 *The Heine–Borel Theorem for* $\mathbb{R}$

A subset of $\mathbb{R}$ is $d^{(1)}$-sequentially compact if, and only if, it is $d^{(1)}$-closed and bounded.

Examples of sequentially compact spaces

1. The middle-third Cantor set, K_C, is sequentially compact for the Euclidean metric. (Exercise 2.3)
2. Finite subsets of any metric space are sequentially compact. (Exercise 2.4)

Subsets, images and products of sequentially compact spaces

Theorem 2.5

Let $A \subseteq X$ where (X,d) is a d-sequentially compact metric space. Then A is d-closed if, and only if, A is d_A-sequentially compact.

Lemma 2.6

Let (X,d) and (Y,e) be metric spaces with X being sequentially compact. Let $f\colon X \to Y$ be a (d,e)-continuous function. Then the image, $f(X)$, is sequentially compact.

Corollary 2.7

Sequential compactness is preserved by homeomorphism.

Theorem 2.8 *Tikhonov's Theorem*

Let (X, d) and (Y, d') be two sequentially compact metric spaces. Then $(X \times Y, e)$ is sequentially compact (where e is a product metric).

Sequentially compact subsets of $\mathbb{R}^n$

Definition 2.9 *Bounded set*

Let (X, d) be a metric space and let $A \subseteq X$. Then A is **bounded** if there is $M > 0$ such that $d(x, y) \leq M$ for all $x, y \in A$. If we want to emphasise the metric, we say that A is **d-bounded**.

Proposition 2.11

Any $d^{(n)}$-closed and bounded subset of $\mathbb{R}^n$ is $d^{(n)}$-sequentially compact.

Theorem 2.12 *The Heine–Borel Theorem*

A subset of $\mathbb{R}^n$ is $d^{(n)}$-sequentially compact if, and only if, it is $d^{(n)}$-closed and bounded.

The Generalised Extreme Value Theorem

Theorem 2.13 *General Extreme Value Theorem*

Let (X, d) be a non-empty sequentially compact metric space, and let $f\colon X \to \mathbb{R}$ be $(d, d^{(1)})$-continuous. Then there are $c, d \in X$ such that

$$f(c) \leq f(x) \leq f(d) \quad \text{for all } x \in X.$$

Functions and sequential compactness

Proposition 3.1

Let $A \subseteq \mathbb{R}$ and $f\colon A \to \mathbb{R}$. Then f is (d_A, d)-continuous at $a \in A$ if, and only if, for each $\varepsilon > 0$, there is $\delta > 0$ such that

$$|f(x) - f(a)| < \varepsilon, \text{ whenever } x \in A \text{ and } |x - a| < \delta.$$

(Here d denotes the Euclidean metric for $\mathbb{R}$ and d_A denotes the corresponding induced metric for A.)

This equivalent formulation of continuity is known as the **ε-δ definition of continuity**.

Theorem 3.2

Let (X, d) and (Y, e) be metric spaces and let $f\colon X \to Y$. Then f is (d, e)-continuous at $a \in X$ if, and only if, for each $\varepsilon > 0$, there is $\delta > 0$ so that $e(f(x), f(a)) < \varepsilon$ whenever $d(x, a) < \delta$.

Theorem 3.3

Let $f\colon [0, 1] \to \mathbb{R}$ be a $(d^{(1)}_{[0,1]}, d^{(1)})$-continuous function. Then for each $\varepsilon > 0$, there is $\delta > 0$ such that for each $x, y \in [0, 1]$, $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$.

(A similar result holds if the interval $[0, 1]$ is replaced by a closed bounded interval $[a, b]$.)

Theorem 3.4

Let (X, d) and (Y, e) be metric spaces and suppose $f\colon X \to Y$ is (d, e)-continuous on X.

If X is d-sequentially compact, then for each $\varepsilon > 0$, there is $\delta > 0$ such that for each $x, y \in X$, $e(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

Definition 3.5 *Uniform continuity*

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$ satisfies the following condition: for each $\varepsilon > 0$, there is $\delta > 0$ such that for each $x, y \in X$, if $d(x, y) < \delta$, then $e(f(x), f(y)) < \varepsilon$.

Then f is (d, e)-**uniformly continuous**.

Functions in $C[0, 1]$ are uniformly continuous

Every function in $C[0, 1]$ is uniformly continuous on $[0, 1]$ for the Euclidean metrics.

Lemma 3.6

Let (X, d) and (Y, e) be metric spaces and suppose that $f\colon X \to Y$. If f is (d, e)-uniformly continuous on X, then f is (d, e)-continuous on X.

Sequentially compact subsets of $C[0, 1]$

Definition 3.7 *Pointwise bounded*

Let A be a set of functions from X to $\mathbb{R}$. We say that the set A is **pointwise bounded** if for each $x \in X$, there is $M_x \geq 0$ such that for each $f \in A$, $|f(x)| \leq M_x$.

Examples of pointwise bounded sets

1. Let $g\colon \mathbb{R} \to \mathbb{R}$ be a function and let A_g be the set consisting of all functions $f\colon \mathbb{R} \to \mathbb{R}$ such that $|f(x)| \leq |g(x)|$ for each $x \in \mathbb{R}$. Then A is pointwise bounded. (Exercise 3.5)
2. Let (X, d) be a metric space, let $L \geq 0$, and fix $a \in X$ and $b \in \mathbb{R}$. Define

$$A = \{f\colon X \to \mathbb{R} : f(a) = b \text{ and } f \text{ is } (d, d^{(1)})\text{-Lipschitz with Lipschitz constant } L\}.$$

   Then A is a pointwise bounded set. (Exercise 3.7)
3. If A is a sequentially compact subset of $(C[0, 1], d_{\max})$, then it is pointwise bounded. (Exercise 3.9)

Definition 3.8 *Equicontinuity*

Let (X, d) and (Y, e) be metric spaces and let A be a set of functions from X to Y.

The set A is a (d, e)-**equicontinuous set of functions** if for each $\varepsilon > 0$, there is a $\delta > 0$ such that for each $f \in A$ and every $x, y \in X$, $e(f(x), f(y)) < \varepsilon$ whenever $d(x, y) < \delta$.

We usually suppress mention of the metrics when it is clear from the context and just say that A is an equicontinuous set (or family) of functions.

Examples of equicontinuous sets of functions

1. Let $L \geq 0$ be given and let

$$A = \{f : [0,1] \to \mathbb{R} : f \text{ is } (d^{(1)}_{[0,1]}, d^{(1)})\text{-Lipschitz with Lipschitz constant } L\}.$$

   Then A is an equicontinuous set of functions (for the Euclidean metrics on $[0,1]$ and $\mathbb{R}$). (Worked Exercise 3.9)

2. Let (X,d) and (Y,e) be metric spaces and let A be a (d,e)-equicontinuous set of functions from X to Y. If $B \subseteq A$, then B is also a (d,e)-equicontinuous set of functions. (Exercise 3.10)

Lemma 3.10

Let (X,d) and (Y,e) be metric spaces and let A be a (d,e)-equicontinuous set of functions from X to Y. Then each $f \in A$ is (d,e)-uniformly continuous and hence (d,e)-continuous on X.

Proposition 3.11

If A is a $d_{\max}$-sequentially compact subset of $C[0,1]$, then A is equicontinous.

Theorem 3.12 *The Arzelà–Ascoli Theorem*

Let A be a subset of $C[0,1]$. Suppose that:

1. A is $d_{\max}$-closed
2. A is a pointwise bounded set of functions
3. A is an equicontinuous set of functions.

Then A is $d_{\max}$-sequentially compact.

Compact metric spaces

Definition 4.2 *Open cover*

Let (X,d) be a metric space and let $A \subseteq X$.

A collection $\mathcal{S}$ of d-open subsets of X is an **open cover** of A if

$$A \subseteq \bigcup_{U \in \mathcal{S}} U;$$

that is, for each $a \in A$, there exists $U \in \mathcal{S}$ such that $a \in U$.

If we wish to emphasise the metric, we say d-open cover.

Definition 4.4 *Finite subcover*

Let (X, d) be a metric space and let $A \subseteq X$. Let $\mathcal{S}$ be a d-open cover of A. A collection $\mathcal{R}$ of d-open subsets of X is a **finite (d-open) subcover** of A from $\mathcal{S}$ if:

1. $\mathcal{R} \subseteq \mathcal{S}$
2. $A \subseteq \bigcup_{U \in \mathcal{R}} U$
3. $\mathcal{R}$ is a finite collection of sets.

Theorem 4.5

Every $d^{(1)}$-open cover of the interval $[0, 1]$ has a finite subcover.

The definition of a compact metric space

Definition 4.7

A metric space (X, d) is **compact** if each d-open cover of X contains a finite subcover of X.

A subset $A \subseteq X$ is **compact** if (A, d_A) is compact as a metric space in its own right.

If we want to emphasise the metric, then we refer to a compact space (X, d) as a d-compact metric space.

Corollary 4.8 *to Theorem 4.5*

The interval $[0, 1]$, with the metric induced by the usual Euclidean metric on $\mathbb{R}$, is compact.

Examples of compact spaces

1. The empty set, $\emptyset$, is a compact subset of every metric space. (Exercise 4.6)
2. Each finite subset of a metric space is compact. (Exercise 4.7)
3. The only compact subsets of (X, d_0), where X is a set and d_0 is the discrete metric, are the finite subsets. (Example 4.9)

Sequentially compact and compact are equivalent for metric spaces

Definition 5.1 *ε-net*

Let (X, d) be a metric space. For $\varepsilon > 0$, an **ε-net** for X is a set $N(\varepsilon) \subseteq X$ such that

$$X \subseteq \bigcup_{p \in N(\varepsilon)} B_d(p, \varepsilon).$$

When we want to emphasise the metric, we use the notation $N_d(\varepsilon)$.

Elementary facts about ε-nets

- For each $\varepsilon > 0$, X is an ε-net for itself.
- If S is an ε-net for X, then it is also a δ-net for X for each $\delta > \varepsilon$.

Lemma 5.2

Let (X, d) be a sequentially compact metric space. Then X has a *finite* ε-net for each $\varepsilon > 0$.

Corollary 5.3

A sequentially compact metric space is bounded.

Totally bounded sets

Definition 5.4 *The min metric*

Let (X, d) be a metric space and define a distance function $d_{\min}\colon X \times X \to \mathbb{R}$ as follows:

$$d_{\min}(a, b) = \begin{cases} d(a, b), & \text{if } d(a, b) \leq 1, \\ 1, & \text{if } d(a, b) > 1. \end{cases}$$

Properties of $d_{\min}$

1. $d_{\min}$ is a metric. (Exercise 5.2)
2. Let (X, d) be a metric space and let (a_n) be a sequence in X. Then (a_n) is d-convergent if, and only if, (a_n) is $d_{\min}$-convergent. (Exercise 5.3)
3. Let (X, d) be a metric space and let $A \subseteq X$. Then A is $d_{\min}$-bounded. (Exercise 5.4)

Definition 5.5 *Totally bounded*

A metric space (X, d) is **totally bounded** if there is a finite ε-net for each $\varepsilon > 0$.

Sequentially compact metric spaces are compact

Lemma 5.6

Each compact metric space (X, d) is totally bounded.

Corollary 5.7 *to Lemma 5.2*

Each sequentially compact metric space is totally bounded.

Lemma 5.8

Let (X, d) be a sequentially compact metric space and let $\mathcal{S}$ be an open cover of X. Then there is an $\varepsilon > 0$ such that, for each $x \in X$, there is a $U \in \mathcal{S}$ with $B_d(x, \varepsilon) \subseteq U$.

Definition 5.9 *Lebesgue number*

The number ε given in Lemma 5.8 is called a **Lebesgue number** of the open cover $\mathcal{S}$.

Theorem 5.10

Every sequentially compact metric space is compact.

Compact metric spaces are sequentially compact

Theorem 5.11

Let (X, d) be a compact metric space. Then each sequence in X has a convergent subsequence and so (X, d) is sequentially compact.

Theorem 5.12

A metric space is compact if, and only if, it is sequentially compact.

Corollary 5.13

1. Let $A \subseteq X$ where (X, d) is a d-compact metric space. Then A is d-closed if, and only if, A is d_A-compact.
2. Let (X, d) and (Y, d') be metric spaces with X being d-compact. Let $f\colon X \to Y$ be a (d, d')-continuous function. Then the image, $f(X)$, is d'-compact.
3. Let (X, d) and (Y, d') be two compact metric spaces. Then $(X \times Y, e)$ is compact (where e is a product metric).
4. Compactness is a topological invariant.

Theorem 5.14 *The Heine–Borel Theorem*

A subset of $\mathbb{R}^n$ is $d^{(n)}$-compact if, and only if, it is $d^{(n)}$-closed and bounded.

Unions and intersections of compact spaces

Theorem 6.1

Let (X, d) be a metric space. Then the intersection of any collection of compact subsets of X is compact.

Corollary 6.2

Let (X, d) be a metric space. Then the intersection of any collection of sequentially compact subsets of X is sequentially compact.

Theorem 6.3

Let (X, d) be a metric space, and let K_n $(n = 1, 2, 3, \ldots)$ be non-empty compact sets in X for which

$$K_1 \supseteq K_2 \supseteq K_3 \supseteq \cdots \supseteq K_n \supseteq K_{n+1} \supseteq \cdots .$$

Then $\bigcap_{n=1}^{\infty} K_n$ is a non-empty compact set.

Proposition 6.4

The finite union of a collection of compact subsets is compact.

Corollary 6.5

The finite union of a collection of sequentially compact subsets is sequentially compact.

Epilogue

Compact and sequentially compact sets need not be the same in topological spaces that are not determined by a metric.

Chapter 23: Completeness

Completeness

Lemma 1.2

Let (X, d) be a metric space and let (a_n) be a convergent sequence in X. Then, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$d(a_n, a_m) < \varepsilon \text{ for each } n, m > N.$$

Definition 1.3 *Cauchy sequence*

Let (X, d) be a metric space and let (a_n) be a sequence in X.

Then (a_n) is a **d-Cauchy sequence** if, for each $\varepsilon > 0$, there is an $N \in \mathbb{N}$ such that

$$d(a_n, a_m) < \varepsilon \text{ for each } n, m > N.$$

When it is clear from the context, we suppress the d and just refer to a sequence (a_n) as being a Cauchy sequence. If we wish to emphasise the metric, we say that (a_n) is a Cauchy sequence *for* d or *with respect to* d, or that (a_n) is a d-Cauchy sequence.

Showing a sequence is not Cauchy

To show that a given sequence is *not* a Cauchy sequence, it is enough to find just *one* $\varepsilon > 0$ such that, for each $N \in \mathbb{N}$, there are n and m both larger than N with $d(a_n, a_m) \geq \varepsilon$.

Lemma 1.4

Let (X, d) be a metric space and let (a_n) be a d-Cauchy sequence in X. Then (a_n) is a bounded sequence in X – that is, there are an $a \in X$ and an $M > 0$ such that $\{a_n : n \in \mathbb{N}\} \subseteq B_d(a, M)$.

Corollary 1.5

Each $d^{(1)}$-Cauchy sequence in $\mathbb{R}$ is bounded.

Lemma 1.6

Let (X, d) be a metric space and let (a_n) be a Cauchy sequence in X. If (a_n) has a convergent subsequence with limit a in X, then (a_n) is convergent with limit a.

Theorem 1.8

Each $d^{(1)}$-Cauchy sequence in $\mathbb{R}$ is $d^{(1)}$-convergent.

Corollary 1.9

A sequence in $\mathbb{R}$ is $d^{(1)}$-convergent if, and only if, it is a $d^{(1)}$-Cauchy sequence.

Complete metric spaces

Definition 2.1 *Complete metric space*

A metric space (X, d) is a **complete metric space** if each Cauchy sequence in (X, d) is convergent; otherwise it is an **incomplete metric space**.

Proposition 2.2

Let (X, d) be a metric space and let $A \subseteq X$. Let d_A denote the induced metric on A from d.

1. If (A, d_A) is a complete metric space, then A is d-closed.
2. If (X, d) is a complete metric space and A is d-closed, then (A, d_A) is a complete metric space.

Theorem 2.3

Each compact metric space is complete.

Lemma 2.4

A sequence $(\mathbf{a}_k)_{k\in\mathbb{N}}$ is a $d^{(n)}$-Cauchy sequence in $\mathbb{R}^n$ if, and only if, for each $j = 1, 2, \ldots, n$ the component sequence $(a_{j,k})_{k\in\mathbb{N}}$ is a $d^{(1)}$-Cauchy sequence in $\mathbb{R}$.

Corollary 2.5

The metric space $(\mathbb{R}^n, d^{(n)})$ is complete.

Lemma 2.6

Let $(\mathbf{x}_n)$ be a $d_{\mathbf{C}}$-Cauchy sequence in $\mathbf{C}$. Then for each $k \in \mathbb{N}$, there is $N_k \in \mathbb{N}$ such that for $n, m > N_k$,

$$x_{i,n} = x_{i,m} \text{ for } i = 1, \ldots, k.$$

Theorem 2.7

$(\mathbf{C}, d_{\mathbf{C}})$ is a complete metric space.

Theorem 2.8

The metric space $(C[0, 1], d_{\max})$ is complete.

Completeness is not a homeomorphism invariant. (Example 2.10.)

Definition 2.11 *Isometric metric spaces*

Let (X, d_X) and (Y, d_Y) be metric spaces.

A function $f\colon X \to Y$ is an **isometry** if

1. f is one–one and onto
2. for each $x_1, x_2 \in X$, $d_X(x_1, x_2) = d_Y(f(x_1), f(x_2))$.

The metric spaces (X, d_X) and (Y, d_Y) are **isometric** if there is an isometry between X and Y.

Facts about isometries

1. If $f\colon X \to Y$ is an isometry, then it is invertible (since it is one–one and onto) and its inverse, f^{-1}, is also an isometry.
2. Since an isometry preserves distances and is a bijection, a sequence in X converges if, and only if, its image under the isometry also converges in Y. Equivalently, a sequence in X is a d_X-Cauchy sequence if, and only if, its image under the isometry is a d_Y-Cauchy sequence in Y.
3. Since closed sets and sequentially compact sets can be characterised in terms of sequences, an isometry preserves closed and (sequentially) compact sets, and thus also preserves open sets. This implies that connectedness is also preserved under isometry.

Theorem 2.12

Let (X, d_X) and (Y, d_Y) be isometric metric spaces. If (X, d_X) is complete, then (Y, d_Y) is also complete.

The Contraction Mapping Theorem

Definition 3.1 *Fixed point*

Let X be a set and let $T\colon X \to X$ be a function.

A **fixed point** of T is a point $x \in X$ such that $T(x) = x$.

Converting zeros to fixed points (Exercise 3.2)

Let $f\colon \mathbb{R} \to \mathbb{R}$ be a function and for $r \in \mathbb{R} - \{0\}$ define $g_r\colon \mathbb{R} \to \mathbb{R}$ by $g_r(x) = x - rf(x)$. Then X is a fixed point of g_r if, and only if, X is a zero of f.

Definition 3.3 *Contraction mapping*

Let (X, d) be a metric space. A function $T\colon X \to X$ is a **contraction mapping** for d if there is a real number $0 \leq \lambda < 1$ such that

$$d(T(x), T(y)) \leq \lambda d(x, y) \text{ for each } x, y \in X.$$

Any such real number λ is known as a **contraction ratio** for T.

Examples of contraction mappings

1. The map $T\colon C[0,1] \to C[0,1]$ defined by

   $$T(f)(t) = 1 + \int_0^t sf(s)ds, \text{ for } 0 \leq t \leq 1$$

   is a contraction mapping for $d_{\max}$ with contraction ratio $\frac{1}{2}$. (Worked Exercise 3.4)
2. Constant functions are contraction mappings with contraction ratio 0. (Exercise 3.4)

3. The map $T\colon C[0,1] \to C[0,1]$ defined by

$$T(f)(t) = \int_0^t \tfrac{1}{2}\cos f(s)\,ds, \text{ for } 0 \le t \le 1$$

is a contraction mapping for $d_{\max}$ with contraction ratio $\frac{1}{2}$. (Exercise 3.5)

Lemma 3.5

Let (X, d) be a metric space. Suppose that $T\colon X \to X$ is a contraction mapping for d. Then T has at most one fixed point.

Theorem 3.6 *Contraction Mapping Theorem*

Let (X, d) be a complete metric space, and let $T\colon X \to X$ be a contraction mapping for d. Then there is a unique element $x_T \in X$ such that

$$T(x_T) = x_T.$$

Moreover, for each $x \in X$, the sequence of iterates

$$x,\ T(x),\ T(T(x)),\ T(T(T(x))),\ \ldots$$

converges to this unique fixed point x_T; that is, if

$$T^n(x) = (T \circ T \circ \cdots \circ T)(x) \ (T \text{ composed with itself } n \text{ times}),$$

then for each $x \in X$

$$d(T^n(x), x_T) \to 0 \quad \text{as } n \to \infty.$$

Corollary 3.8

Let (X, d) be a complete metric space and let $T\colon X \to X$ be a contraction mapping for d with contraction ratio $\lambda \in [0, 1)$. Then for each $x \in X$ and each $n \in \mathbb{N}$,

$$d(x_T, T^n(x)) \le \frac{\lambda^n}{1-\lambda} d(x, T(x)),$$

where x_T is the fixed point of the map T.

Theorem 3.9

Let $a < b \in \mathbb{R}$, and suppose that $f\colon [a, b] \to \mathbb{R}$ is continuous and differentiable on $[a, b]$. If

1. $f(a) < 0 < f(b)$
2. there are $m, M \in \mathbb{R}$ with $0 < m \le M$ such that

$$m \le f'(x) \le M \text{ for each } x \in [a, b],$$

then for $0 < r \le \frac{1}{M}$, the function $g_r\colon [a, b] \to \mathbb{R}$ given by $g_r(x) = x - rf(x)$ maps $[a, b]$ into itself and is a contraction mapping of $[a, b]$ with contraction ratio $1 - mr$.

Theorem 3.10

Suppose that $F\colon \mathbb{R}^2 \to \mathbb{R}$ is a $(d^{(2)}, d^{(1)})$-continuous function and $a \in \mathbb{R}$ is given. Then $f \in C[0,1]$ is a solution of the differential equation

$$\frac{df}{ds} = F(s, f(s)), \quad f(0) = a \quad (0 \le s \le 1)$$

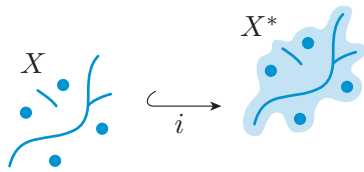
if, and only if,

$$f(t) = a + \int_0^t F(s, f(s))\, ds \quad \text{for each } 0 \le t \le 1.$$

Completion

Definition 4.1 *Completion of a metric space*

Let (X, d) be a metric space. A **completion** of X is a complete metric space (X^*, d^*) that contains a dense subspace isometric to (X, d).



$X^* = \mathrm{Cl}_{(X^*,d^*)}(i(X))$

A note about completions

If (X^*, d^*) and (Y^*, e^*) are both completions of (X, d), then they are isometric – thus, up to isometry, completions are unique. Consequently, we usually talk about *the* completion of a space.

Constructing the completion of a space

Given a metric space (X, d), we let X_C denote the collection of all d-Cauchy sequences in X.

Lemma 4.2

Let (X, d) be a metric space. If (a_n) and (b_n) are sequences in X that converge to the same limit, then $d(a_n, b_n) \to 0$ as $n \to \infty$.

An equivalence relation on X_C

We say that $\mathbf{a} = (a_n) \in X_C$ and $\mathbf{b} = (b_n) \in X_C$ are equivalent and write $\mathbf{a} \sim \mathbf{b}$ if, and only if, $(d(a_n, b_n))$ is a real null sequence. This is an equivalence relation on X_C. (Exercise 4.3)

We let X^* be the collection of equivalence classes of X_C for the relation $\sim$ and for a Cauchy sequence (a_n) of points in X, we define $[(a_n)]$ to be the equivalence class in X^* that contains (a_n).

Lemma 4.3

Let $\mathbf{a} = (a_n) \in X_C$ and $\mathbf{b} = (b_n) \in X_C$. Then

1. $\lim_{n\to\infty} d(a_n, b_n)$ exists
2. if $\mathbf{a}' = (a'_n) \in [\mathbf{a}]$, $\mathbf{b}' = (b'_n) \in [\mathbf{b}]$, then
$$\lim_{n\to\infty} d(a'_n, b'_n) = \lim_{n\to\infty} d(a_n, b_n).$$

Defining a metric on X^*

For $[\mathbf{a}] = [(a_n)]$, $[\mathbf{b}] = [(b_n)] \in X^*$, we define $d^*([\mathbf{a}], [\mathbf{b}]) = \lim_{n\to\infty} d(a_n, b_n)$.

Theorem 4.4

(X^*, d^*) is a metric space.

Embedding X into X^*

We define an embedding of X into X^*, $i\colon X \to X^*$, by

$$i(x) = [(x, x, x, \ldots)] = [\mathbf{x}].$$

Lemma 4.5

The set $i(X)$ is a d^*-dense subset of (X^*, d^*).

Theorem 4.6

The metric space (X^*, d^*) is complete.

Theorem 4.7

Any metric space (X, d) has a completion (X^*, d^*).

Examples of complete and incomplete spaces

Space	Status	Chapter reference
finite spaces	complete	Exercise 2.5
$(\mathbb{R}, d^{(1)})$	complete	Theorem 1.8
$(\mathbb{R}^n, d^{(n)})$	complete	Corollary 2.5
(X, d_0), X any set	complete	Exercise 2.2
closed subsets of a complete space	complete	Proposition 2.2(1)
continuous image of a compact space	complete	Exercise 2.4
$(\mathbf{C}, d_{\mathbf{C}})$, Cantor space	complete	Theorem 2.7
$(C[0,1], d_{\max})$	complete	Theorem 2.8
$((0,1], d^{(1)}_{(0,1]})$	incomplete	Exercise 2.1
$C[0,1]$ with the integral metric	incomplete	Example 2.9

Chapter 24: Fractals

Examples of fractals

Theorem 1.1

The middle-third Cantor set K_C is uncountable.

Definition 1.2 *Similarity*

A mapping $S\colon \mathbb{R}^m \to \mathbb{R}^m$ is a **similarity** if there is a $c > 0$ such that

$$d^{(m)}(S(\mathbf{x}), S(\mathbf{y})) = c\, d^{(m)}(\mathbf{x}, \mathbf{y}) \quad \text{for each } \mathbf{x}, \mathbf{y} \in \mathbb{R}^m.$$

The constant c is the **similarity ratio**.

Definition 1.3 *Self-similar sets*

A set $A \subseteq \mathbb{R}^m$ is **self-similar** if there are similarities $S_i : \mathbb{R}^m \to \mathbb{R}^m$ with similarity ratios $0 < c_i < 1$, for $1 \leq i \leq k$, such that

$$A = \bigcup_{i=1}^{k} S_i(A).$$

(The set A consists of k copies of itself, each of which is the image of A under a similarity.)

Using the Contraction Mapping Theorem

The Hausdorff metric

Definition 2.3

For each $m \in \mathbb{N}$, the set $\mathcal{K}(\mathbb{R}^m)$ is defined as follows:

$$\mathcal{K}(\mathbb{R}^m) = \{K \subseteq \mathbb{R}^m : K \text{ is non-empty and compact}\}.$$

Each member of $\mathcal{K}(\mathbb{R}^m)$ is *itself* a set. For example, the *set* $[0, 1] \subseteq \mathbb{R}$ is non-empty and compact, so it is a member of $\mathcal{K}(\mathbb{R})$. Thus we write $[0, 1] \in \mathcal{K}(\mathbb{R})$.

Definition 2.4 *Euclidean distance between a point and a set*

Let $\mathbf{x} \in \mathbb{R}^m$ and let $K \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$.

The **Euclidean distance** between the point $\mathbf{x}$ and the set K is

$$d^{(m)}(\mathbf{x}, K) = \min\{d^{(m)}(\mathbf{x}, \mathbf{k}) : \mathbf{k} \in K\}.$$

Lemma 2.5

Let $\mathbf{x} \in \mathbb{R}^m$ and let $K \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$. Then

$$d^{(m)}(\mathbf{x}, K) = 0 \text{ if, and only if, } \mathbf{x} \in K.$$

Lemma 2.6

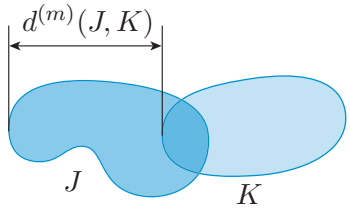
Let $\mathbf{x} \in \mathbb{R}^m$ and let $J, K \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$. Let $J \subseteq K$; then

$$d^{(m)}(\mathbf{x}, K) \leq d^{(m)}(\mathbf{x}, J).$$

Definition 2.7 *Euclidean distance between compact sets*

Let $J, K \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$. The **Euclidean distance** between the set J and the set K is

$$d^{(m)}(J, K) = \max\{d^{(m)}(\mathbf{j}, K) : \mathbf{j} \in J\}.$$



Lemma 2.9

Let $J, K, L \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$. Then

$$d^{(m)}(J, L) \leq d^{(m)}(J, K) + d^{(m)}(K, L).$$

Definition 2.10 *Hausdorff distance*

Let $J, K \in \mathcal{K}(\mathbb{R}^m)$, for some $m \in \mathbb{N}$. The **Hausdorff distance** d_{H} between the sets J and K is

$$d_{\mathrm{H}}(J, K) = \max\{d^{(m)}(J, K), d^{(m)}(K, J)\}.$$

Theorem 2.11

The Hausdorff distance, d_{H}, is a metric on $\mathcal{K}(\mathbb{R}^m)$.

Theorem 2.12

The metric space $(\mathcal{K}(\mathbb{R}^m), d_{\mathrm{H}})$ is complete.

Proposition 2.13

Let $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$ be $(d^{(m)}, d^{(m)})$-continuous, and let S be defined by

$$S(K) = \bigcup_{i=1}^{k} S_i(K) \quad \text{for each } K \in \mathcal{K}(\mathbb{R}^m).$$

Then S maps $\mathcal{K}(\mathbb{R}^m)$ to $\mathcal{K}(\mathbb{R}^m)$.

Proposition 2.14

Let $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$ be contraction mappings. Then the mapping $S \colon \mathcal{K}(\mathbb{R}^m) \to \mathcal{K}(\mathbb{R}^m)$ defined by

$$S(K) = \bigcup_{i=1}^{k} S_i(K) \quad \text{for each } K \in \mathcal{K}(\mathbb{R}^m)$$

is a contraction mapping with respect to the Hausdorff distance d_{H} on $\mathcal{K}(\mathbb{R}^m)$; that is, there is a number λ such that $0 \leq \lambda < 1$ and

$$d_{\mathrm{H}}(S(J), S(K)) \leq \lambda d_{\mathrm{H}}(J, K),$$

for all $J, K \in \mathcal{K}(\mathbb{R}^m)$.

Theorem 2.15 *Iterated function schemes*

Let $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$ be contraction mappings, and let the mapping $S \colon \mathcal{K}(\mathbb{R}^m) \to \mathcal{K}(\mathbb{R}^m)$ be defined by

$$S(K) = \bigcup_{i=1}^{k} S_i(K) \quad \text{for each } K \in \mathcal{K}(\mathbb{R}^m).$$

Then there is a unique set $K_S \in \mathcal{K}(\mathbb{R}^m)$ such that

$$S(K_S) = K_S.$$

Moreover, for each $K \in \mathcal{K}(\mathbb{R}^m)$, the sequence of iterates

$$K,\ S(K),\ S^2(K),\ S^3(K),\ \ldots$$

converges to K_S; that is,

$$d_{\mathrm{H}}(S^n(K), K_S) \to 0 \text{ as } n \to \infty.$$

This method of generating a fractal is called an *iterated function scheme*.

Definition 2.16 *Invariant*

A set A is **invariant** for the mappings $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$ if

$$\bigcup_{i=1}^{k} S_i(A) = A.$$

A self-similar set is just a set that is invariant for *similarities* $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$.

Lemma 2.17

Let $S_1, S_2, \ldots, S_k \colon \mathbb{R}^m \to \mathbb{R}^m$ be contraction mappings, and let $S \colon \mathcal{K}(\mathbb{R}^m) \to \mathcal{K}(\mathbb{R}^m)$ be the mapping defined by

$$S(K) = \bigcup_{i=1}^{k} S_i(K) \quad \text{for each } K \in \mathcal{K}(\mathbb{R}^m).$$

Let $S(K_0) \subseteq K_0$, for some $K_0 \in \mathcal{K}(\mathbb{R}^m)$; then

$$\bigcap_{n \in \mathbb{N}} S^n(K_0) = K_S,$$

where K_S is the unique set in $\mathcal{K}(\mathbb{R}^m)$ such that $S(K_S) = K_S$.

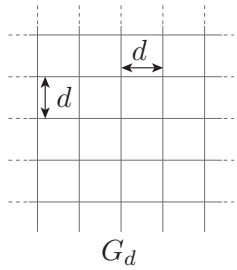
Dimensions of fractals

Definition 3.1 *Grid of side d*

Let $d \in (0, 1)$. The **grid** G_d is a particular set of closed boxes of side d:

$$G_d = \{[nd, (n+1)d] \times [md, (m+1)d] : n, m \in \mathbb{Z}\}.$$

The number of (closed) boxes in G_d that intersect a set A is denoted by $N_d(A)$.



Definition 3.2 *Box dimension*

Let A be a bounded subset of $\mathbb{R}^2$. The **box dimension** of A is

$$\dim A = \lim_{d \to 0} \frac{\log N_d(A)}{-\log d},$$

provided that this limit exists.

Definition 3.4 *Open set condition*

The similarities $S_1, S_2, \ldots, S_k$ satisfy the **open set condition** if there exists a non-empty bounded open set U such that

$$S_i(U) \cap S_j(U) = \emptyset, \quad \text{for } i \neq j, \text{ and } \bigcup_{i=1}^{k} S_i(U) \subseteq U.$$

Theorem 3.5

Let K_S be the non-empty compact invariant set of the similarities $S_1, S_2, \ldots, S_k$, and let $c_i \in (0, 1)$ be the similarity ratio of S_i, for $1 \leq i \leq k$. If $S_1, S_2, \ldots, S_k$ satisfy the open set condition, then

$$\dim K_S = s,$$

where s is the solution of the equation $\sum_{i=1}^{k} c_i^s = 1$.

Theorem 3.7

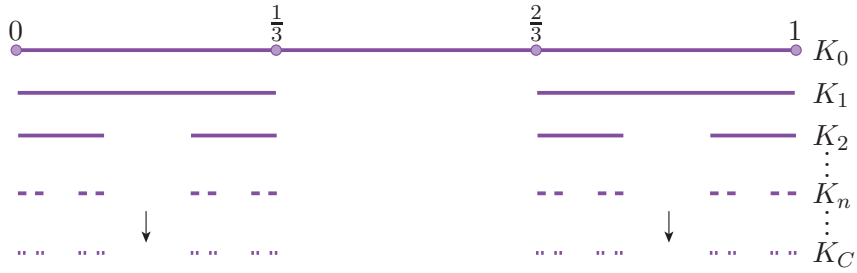
Let K_S be the non-empty compact invariant set of the similarities $S_1, S_2, \ldots, S_k$, each with similarity ratio $c \in (0,1)$. If $S_1, S_2, \ldots, S_k$ satisfy the open set condition, then

$$\dim K_S = \frac{\log k}{-\log c}.$$

Appendix 6: Three important fractals

The middle-third Cantor set

This uncountable non-empty compact subset of the real line is an invariant set for the iterated function scheme $\{S_1, S_2\}$ where $S_i \colon \mathbb{R} \to \mathbb{R}$ is given by $S_i(x) = \frac{x}{3} + \frac{2}{3}(i-1)$ for $i = 1, 2$. Each similarity has similarity ratio $\frac{1}{2}$. It has box dimension $\frac{\log 2}{\log 3}$.



The von Koch curve

This non-empty compact subset of the plane is an invariant set for the iterated function scheme $\{S_1, S_2, S_3, S_4\}$ where $S_1, S_2, S_3, S_4 \colon \mathbb{R}^2 \to \mathbb{R}^2$ are given by

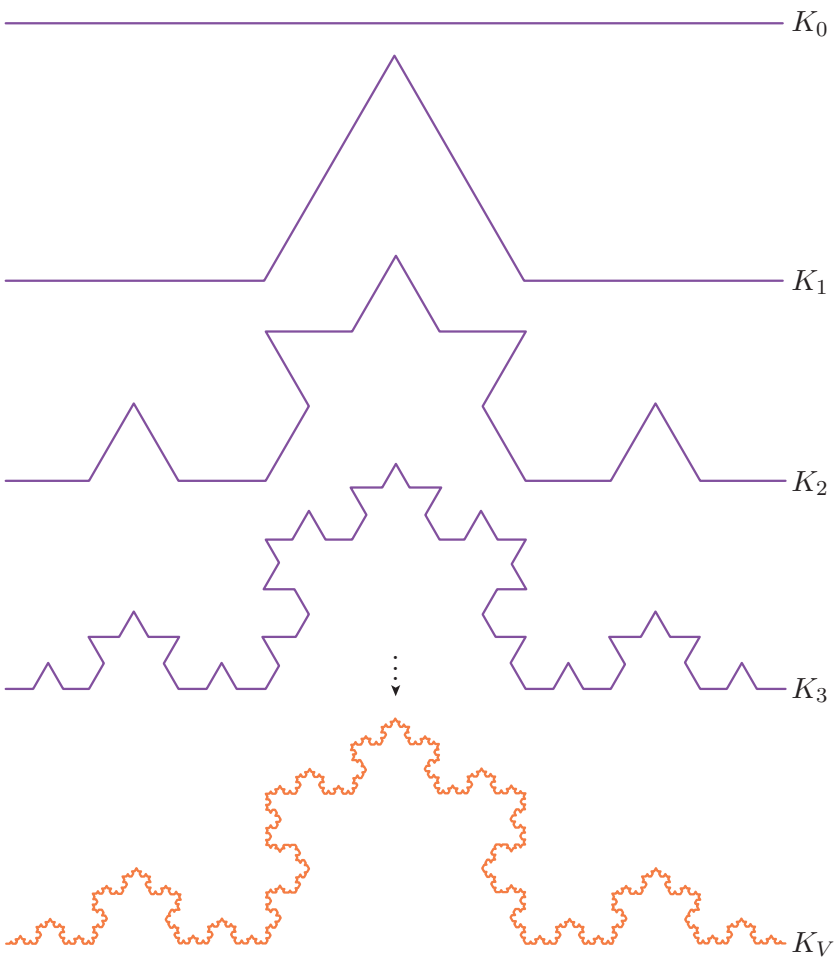
$$S_1(x,y) = \left(\frac{1}{3}x,\ \frac{1}{3}y\right),$$

$$S_2(x,y) = \left(\frac{1}{3}\left(\frac{1}{2}x - \frac{\sqrt{3}}{2}y\right) + \frac{1}{2},\ \frac{1}{3}\left(\frac{\sqrt{3}}{2}x + \frac{1}{2}y\right)\right),$$

$$S_3(x,y) = \left(\frac{1}{3}\left(\frac{1}{2}x + \frac{\sqrt{3}}{2}y\right) + \frac{1}{2},\ \frac{1}{3}\left(-\frac{\sqrt{3}}{2}x + \frac{1}{2}y\right) + \frac{1}{2}\right),$$

$$S_4(x,y) = \left(\frac{1}{3}x + \frac{2}{3},\ \frac{1}{3}y\right).$$

Each similarity has similarity ratio $\frac{1}{3}$ and the curve has box dimension $\frac{\log 4}{\log 3}$.



The Sierpiński gasket

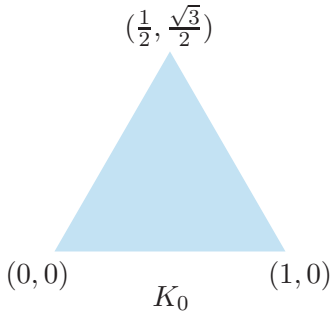
This non-empty compact subset of the plane is an invariant set for the iterated function scheme $\{S_1, S_2, S_3\}$ where $S_1, S_2, S_3 \colon \mathbb{R}^2 \to \mathbb{R}^2$ are given by

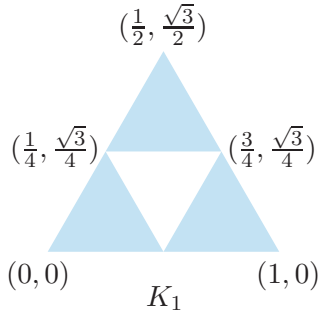
$$S_1(x, y) = \left(\tfrac{1}{2}x,\ \tfrac{1}{2}y\right),$$

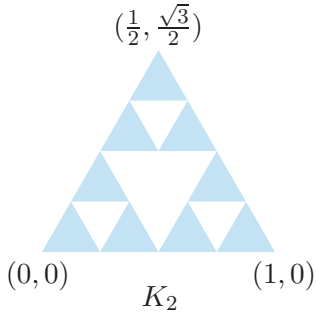
$$S_2(x, y) = \left(\tfrac{1}{2}x + \tfrac{1}{2},\ \tfrac{1}{2}y\right),$$

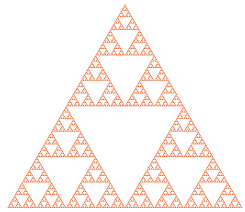
$$S_3(x, y) = \left(\tfrac{1}{2}x + \tfrac{1}{4},\ \tfrac{1}{2}y + \tfrac{\sqrt{3}}{4}\right).$$

Each similarity has similarity ratio $\frac{1}{2}$ and the gasket has box dimension $\frac{\log 3}{\log 2}$.


$(\frac{1}{2}, \frac{\sqrt{3}}{2})$
$(0,0)$
$(1,0)$
K_0


$(\frac{1}{2}, \frac{\sqrt{3}}{2})$
$(\frac{1}{4}, \frac{\sqrt{3}}{4})$
$(\frac{3}{4}, \frac{\sqrt{3}}{4})$
$(0,0)$
$(1,0)$
K_1


$(\frac{1}{2}, \frac{\sqrt{3}}{2})$
$(0,0)$
$(1,0)$
K_2



Index