Instructor: **Yaroslav Korobka** Updated: October 27, 2025

Course: **Mathematics**

1 Basic topology

1.1 Metric spaces

Definition 1.1. A set X, whose elements we shall call *points*, is said to be *a metric space* if with any two points p and q of X there is associated a real number $d(p,q)$, called the *distance* from p to q, such that

(a) $d(p,q) > 0$ if $p \neq q$ and $d(p,p) = 0$,

(b) $d(p,q) = d(q,p)$,

(c) $d(p,q) \leq d(p,r) + d(r,q)$, for $\forall r \in X$.

Any function with these three properties is called a *distance function*, or *a metric*.

Example 1.2 (Metric spaces)**.** The following are examples of the metric spaces:

1. the set of real numbers $\mathbb{R}$ with a metric $d(p,q) = |p-q|$,
2. a real plane $\mathbb{R}^2$ with a metric $d(\mathbf{p},\mathbf{q}) = \sqrt{(p_1-q_1)^2+(p_2-q_2)^2} := \|\mathbf{p}-\mathbf{q}\|$ (Eucledian distance),
3. a real plane $\mathbb{R}^2$ with a metric $d(\mathbf{p},\mathbf{q}) = |p_1-q_1| + |p_2-q_2|$ (Manhattan distance),
4. the set of probability distributions defined on the same measurable space with a metric $d(P,Q) = \frac{1}{\sqrt{2}}\left(\int\left(\sqrt{p(x)}-\sqrt{q(x)}\right)^2 dx\right)^{1/2}$ (Hellinger distance).

It is important to observe that every subset Y of a metric space X is a metric space in its own right, with the same distance function. Thus, every subset of a Euclidean space is a metric space.

Definition 1.3. By the *segment* (a,b) we mean the set of all real numbers x such that $a < x < b$. By the *interval* $[a,b]$ we mean the set of all real numbers x such that $a \leq x \leq b$.

If $a_i < b_i$ for $i = 1,\ldots,k$, the set of all points $\mathbf{x} = (x_1,\ldots,x_k)$ in $\mathbb{R}^k$ whose coordinates satisfy the inequalities $a_i \leq x_i \leq b_i$ $(1 \leq i \leq k)$ is called a *k-cell*. Thus, a 1-cell is an interval, a 2-cell is a rectangle, etc.

If $\mathbf{x} \in \mathbb{R}^k$ and $r > 0$, the *open* (or *closed*) *ball* B with center at $\mathbf{x}$ and radius r is defined to be the set of all $\mathbf{y} \in \mathbb{R}^k$ such that $\|\mathbf{y}-\mathbf{x}\| < r$ (or $\|\mathbf{y}-\mathbf{x}\| \leq r$).

We call a set $E \subset \mathbb{R}^k$ *convex* if

$$\lambda\mathbf{x} + (1-\lambda)\mathbf{y} \in E$$

whenever $\mathbf{x} \in E$, $\mathbf{y} \in E$, and $0 < \lambda < 1$. For example, balls are convex. It is also easy to see that k-cells are convex.

Definition 1.4. Let X be a metric space. All points and sets mentioned below are understood to be elements and subsets of X.

(a) *A neighborhood* of a point p is a set $N_r(p)$ consisting of all points q such that $d(p,q) < r$. The number r is called the *radius* of $N_r(p)$.

(b) A point p is a *limit point* of the set E if *every* neighborhood of p contains a point $q \neq p$ such that $q \in E$. Example: take a set $A := (0,1)$. Point 0 is a limit point, because any open interval, say $(-\varepsilon, \varepsilon)$, intersects A.

(c) If $p \in E$ and p is not a limit point of E, then p is called an *isolated point* of E. Example: take a set $A = \{n^{-1} : n \in \mathbb{N}\}$. Each element is an isolated point because you can take a small interval around n^{-1} that avoids the other fractions in the set.

(d) E is *closed* if every limit point of E is a point of E. Example: take $A = [0,1]$. Both 0 and 1 are limit points and both belong to the set A. A set $B = (0,1]$ is not closed because a limit point 0 does not belong to the set.

(e) A point p is an *interior point* of E if there is a neighborhood $N_r(p)$ of p such that $N \subset E$. Example: take a set $A = (0,1)$. A point 0.5 is an interior point because there is a neighborhood around it, say, $N_{0.1}(0.5)$ that belongs to the set A; if $N_{0.1}(0.5) = (0.4, 0.6) := B$, we have $B \subset A$. On the other hand, if $C = [0.5, 1]$, 0.5 is not an interior point of C, because there is no neighborhood around it that is a subset of C; some points of that neighborhood are outside of C.

(f) E is *open* if every point of E is an interior point of E.

(g) The *complement* of E (denoted by E^c) is the set of all points $p \in X$ such that $p \notin E$.

(h) E is *perfect* if E is closed and if every point of E is a limit point of E. Example: take $A = [0,1]$, which is closed with all points being limit points, so it is perfect. On the other hand, $B = [0,1] \cup \{3\}$ is not perfect because it contains a point 3, which is not a limit point (it is an isolated point).

(i) E is *bounded* if there is a real number M and a point $q \in X$ such that $d(p,q) < M$ for $\forall p \in E$.

(j) E is *dense in* X if every point of X is a limit point of E, or a point of E (or both).

Let us note that in $\mathbb{R}^1$ neighborhoods are segments, whereas in $\mathbb{R}^2$ neighborhoods are interiors of circles.

Theorem 1.5. *Every neighborhood is an open set.*

Proof. Consider neighborhood $E = N_r(p)$, and let q be any point of E. Then there is a positive real number h such that

$$d(p,q) = r - h.$$

For all points s such that $d(q,s) < h$, we have then

$$d(p,s) \leq d(p,q) + d(q,s) < r - h + h = r,$$

so that $s \in E$. Thus, q is an interior point of E. □

Theorem 1.6. *If p is a limit point of a set E, then every neighborhood of p contains infinitely many points of E.*

Proof. Suppose there is a neighborhood N of p which contains only a finite number of points of E. Let $q_1, \ldots, q_n$ be those points of $N \cap E$, which are distinct from p, and put

$$r = \min_{1 \leq m \leq n} d(p, q_m)$$

The minimum of a finite set of positive numbers is clearly positive, so that $r > 0$.

The neighborhood $N_r(p)$ contains no point q of E such that $q \neq p$, so that p is not a limit point of E. This contradiction established the theorem. □

Corollary 1.7. *A finite point set has no limit points.*

Theorem 1.8. *A set E is open if and only if its complement is closed.*

1.2 Compact sets

Definition 1.9. By an *open cover* of a set E in a metric space X we mean a collection $\{G_\alpha\}$ of open subsets of X such that $E \subset \cup_\alpha G_\alpha$.

Definition 1.10. A subset K of a metric space X is said to be *compact* if every open cover of K contains a *finite subcover*. More explicitly, the requirement is that if $\{G_\alpha\}$ is an open cover of K, then there are finitely many indices $\alpha_1, \ldots, \alpha_n$ such that

$$K \subset G_{\alpha_1} \cup \ldots \cup G_{\alpha_n}.$$

Corollary 1.11. *A set E is compact if it is both closed and bounded.*

1.3 Functions

Definition 1.12. Consider two sets A and B, whose elements may be any objects whatsoever, and suppose that with each element x of A there is associated, in some manner, an element of B, which we denote by $f(x)$. Then f is said to be a *function* from A to B (or a *mapping* from A into B). The set A is called the *domain* of f (we also say f is defined on A), and the elements $f(x)$ are called the *values* of f. The set of *all* values of f is called the *range* of f.

Definition 1.13. If for every $y \in B$ there is at most one $x \in A : f(x) = y$, the function f is said to be a 1-1 (*one-to-one*) mapping of A into B. This may also be expressed as follows: f is a 1-1 mapping of A into B provided that $f(x_1) \neq f(x_2)$ whenever $x_1 \neq x_2$, $x_1 \in A$, $x_2 \in A$.

Definition 1.14. Let A and B be two sets and let f be a mapping of A into B. If $f(A) = B$, we say that f maps A *onto* B. If, additionally, f is 1-1, then f is *one-to-one and onto (bijection)*.

Definition 1.15. If there exists a 1-1 mapping of A *onto* B, we say that A and B can be put in 1-1 *correspondence*, or that A and B have the same *cardinal number*, or, briefly, that A and B are *equivalent*, and we write $A \sim B$.

Definition 1.16. For any positive integer n, let J_n be the set whose elements are the integers $1, 2, \ldots, n$; let J be the set consisting of all positive integers. For any set A, we say:

(a) A is finite if $A \sim J_n$ for some n.

(b) A is *infinite* if A is not finite.

(c) A is *countable* if $A \sim J$.

(d) A is *uncountable* if A is neither finite nor countable.

(e) A is *at most countable* if A is finite or countable.

For two finite sets A and B, we evidently have $A \sim B$ if and only if A and B contain the same number of elements (same *cardinality*). For infinite sets, however, the idea of cardinality becomes quite vague, whereas the notion of 1-1 correspondence retains its clarity.

Example 1.17. Let A be the set of all integers. Then A is countable. Consider, the following arrangement of the sets A and J:

$$\begin{aligned} A: &\quad 0, 1, -1, 2, -2, \ldots \\ J: &\quad 1, 2, 3, 4, 5, \ldots \end{aligned}$$

We can, in this example, even give an explicit formula for a function f from J to A which sets up a 1-1 correspondence:

$$f(n) = \begin{cases} \frac{n}{2} & \text{if } n \text{ is even,} \\ -\frac{n-1}{2} & \text{if } n \text{ is odd.} \end{cases}$$

Remark 1.18. A finite set cannot be equivalent to one of its proper subsets. That this is, however, possible for infinite sets, is shown by Example 1.17, in which J is a proper subset of A.

Definition 1.19. In the following, assume that the set A is a subset of $\mathbb{R}$.

(a) If there exists $x \in \mathbb{R}$ such that for every $y \in A$ we have $x \geq y$, then the set A is *bounded from above*.

(b) If there exists $x \in \mathbb{R}$ such that for every $y \in A$ we have $x \leq y$, then the set A is *bounded from below*.

(c) The *supremum* of A, denoted as $\sup A$, is the smallest upper bound of the set A.

(d) The *infimum* of A, denoted as $\inf A$, is the largest lower bound of the set A.

We note that the set A is bounded, if it is bounded both from below and from above, which is equivalent to the Definition 1.4(i). If the set A is not bounded from above, then $\sup A = \infty$, and if it is not bounded from below, then $\inf A = -\infty$.

2 Sequences and limits

Definition 2.1. By a *sequence*, we mean a function f defined on the set J of all positive integers. If $f(n) = x_n$ for $n \in J$, it is customary to denote the sequence f by the symbol $\{x_n\}$, or sometimes by $x_1, x_2, x_3, \ldots$. The values of f, that is, the elements x_n, are called the *terms* of the sequence. If A is a set and if $x_n \in A$ for all $n \in J$, then $\{x_n\}$ is said to be a *sequence in* A, or a *sequence of elements of* A.

Note that the terms $x_1, x_2, x_3, \ldots$ of a sequence need not be distinct.

Since every countable set is the range of a 1-1 function defined on J, we may regard every countable set as the range of a sequence of distinct terms. Speaking more loosely, we may say that the elements of any countable set can be "arranged in a sequence".

Definition 2.2. For a given sequence $\{x_n\}$, if $x_{n+1} > x_n$ for $\forall n \in J$, then the sequence is *increasing*. If $x_{n+1} < x_n$ for $\forall n \in J$, then the sequence is *decreasing*. If $x_{n+1} \geq x_n$ for $\forall n \in J$, then the sequence is *non-decreasing*. If $x_{n+1} \leq x_n$ for $\forall n \in J$, then the sequence is *non-increasing*.

If at least one of these four conditions is satisfied, the sequence is called *monotonic*.

Example 2.3. We give examples of different sequences below.

(a) A sequence that is defined via a formula for the nth term: $x_n = \left(\frac{2}{3}\right)^n$.

(b) A sequence that is defined recursively (Fibonacci sequence): $x_n = x_{n-1} + x_{n-2}$ for $n \geq 3$, and $x_1 = x_2 = 1$.

(c) A sequence $x_n = (-1)^n$.

(d) A sequence $x_n = 2^n$.

Note that the sequence (a) is decreasing with n, while the sequence (b) is non-decreasing with n. The sequence (c) is non-monotonic.

Definition 2.4. A sequence $\{x_n\}$ in a metric space X is said to *converge* if there is a point $x \in X$ with the following property: for every $\varepsilon > 0$ there is an integer N such that $n \geq N$ implies that $d(x_n, x) < \varepsilon$.

In this case, we also say that $\{x_n\}$ converges to x, or that x is the limit of $\{x_n\}$, and we write $x_n \to x$, or

$$\lim_{n\to\infty} x_n = x.$$

If $\{x_n\}$ does not converge, it is said to *diverge*.

We recall that the set of all points x_n $(n = 1, 2, 3, \ldots)$ is the *range* of $\{x_n\}$. The range of a sequence may be a finite set, or it may be infinite. The sequence $\{x_n\}$ is said to be *bounded* if its range is bounded. In the Example 2.3, (a) and (c) are bounded sequences, while (b) and (d) are not.

Example 2.5. Show that $\lim_{n\to\infty}(\frac{2}{3})^n = 0$.

We need to show that for a given $\varepsilon > 0$, after some $n \in J$, the distance between the elements of the sequence and the limit 0 is smaller than ε. In other words, that there exists some N such that for all n larger than N we have $d(x_n, 0) < \varepsilon$. Taking the absolute value, we have $|(\frac{2}{3})^n| < \varepsilon$ for $\forall n \geq N$, and rewriting

$$\begin{aligned}
\left(\frac{2}{3}\right)^n &< \varepsilon, \\
\log\left(\frac{2}{3}\right)^n &< \log\varepsilon, \\
n\log\left(\frac{2}{3}\right) &< \log\varepsilon, \\
n &> \frac{\log\varepsilon}{\log 2/3}.
\end{aligned}$$

Denote the smallest integer larger than a as $\lceil a \rceil$. Then, one can take $N = \lceil n \rceil$, and for all $n \geq N$, the inequality $n > \frac{\log\varepsilon}{\log 2/3}$ is satisfied. Then, 0 is a limit of $\left(\frac{2}{3}\right)^n$.

Theorem 2.6. *Every bounded, monotonic sequence converges.*

Example 2.7. Show that the sequence

$$x_n = \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!} = \sum_{k=1}^{n} \frac{1}{k!}$$

converges.

To show that the sequence converges, we use the Theorem 2.6, hence, it is sufficient to show that the sequence is monotonic and bounded. To show monotonicity, note that

$$x_{n+1} = x_n + \frac{1}{(n+1)!} > x_n,$$

so $\{x_n\}$ is increasing and hence monotonic. To show that it is bounded, note that

$$\frac{1}{n!} = \frac{1}{1 \cdot 2 \cdot 3 \cdot \ldots \cdot n} = \frac{1}{2 \cdot 3 \cdot \ldots \cdot n} \leq \frac{1}{2 \cdot 2 \cdot \ldots \cdot 2} = \frac{1}{2^{n-1}},$$

with strict inequality for $n > 1$. $x_1 = 1$ is finite, hence does not contradict boundedness. For $n > 1$, we have

$$x_n < 1 + \frac{1}{2^1} + \frac{1}{2^2} + \ldots + \frac{1}{2^{n-1}} = \frac{1 - (1/2)^n}{1 - 1/2} = 2 - \left(\frac{1}{2}\right)^{n-1} < 2.$$

Because each element of the sequence x_n for $\forall n > 1$ is bounded by 2, the sequence is bounded.

2.1 Limit laws (i)

Corollary 2.8. *Let $\{x_n\}$ and $\{y_n\}$ are convergent sequences, and let c be a constant. Then,*

(a) $\lim_{n\to\infty} (x_n + y_n) = \lim_{n\to\infty} x_n + \lim_{n\to\infty} y_n$.

(b) $\lim_{n\to\infty} (x_n - y_n) = \lim_{n\to\infty} x_n - \lim_{n\to\infty} y_n$.

(c) $\lim_{n\to\infty} cx_n = c \lim_{n\to\infty} x_n$.

(d) $\lim_{n\to\infty} c = c$.

(e) $\lim_{n\to\infty} (x_n y_n) = \lim_{n\to\infty} x_n \lim_{n\to\infty} y_n$.

(f) $\lim_{n\to\infty} \frac{x_n}{y_n} = \frac{\lim_{n\to\infty} x_n}{\lim_{n\to\infty} y_n}$ *if* $\lim_{n\to\infty} y_n \neq 0$.

(g) $\lim_{n\to\infty} x_n^p = \left(\lim_{n\to\infty} x_n\right)^p$ *if* $p > 0$ *and* $x_n > 0$.

Example 2.9. Find the limit of $\{x_n\}$, where

$$x_n = \frac{2n^3 + n^2 - 7n}{n^3 + 2n + 2}.$$

Rewrite the nth term of the sequence as

$$\frac{2 + n^{-1} - 7n^{-2}}{1 + 2n^{-2} + 2n^{-3}}.$$

The limit of the numerator and the denominator respectively is

$$\lim_{n\to\infty} \left(2 + \frac{1}{n} - \frac{7}{n^2}\right) = 2, \quad \lim_{n\to\infty} \left(1 + \frac{2}{n^2} + \frac{2}{n^3}\right) = 1,$$

so that $\lim_{n\to\infty} x_n = 2$.

Definition 2.10. Given a sequence $\{x_n\}$, consider a sequence $\{n_k\}$ of natural numbers, such that $n_1 < n_2 < n_3 < \ldots$. Then the sequence $\{x_{n_i}\}$ is called a *subsequence* of $\{x_n\}$. If $\{x_{n_i}\}$ converges, its limit is called a *subsequential limit* of $\{x_n\}$.

The sequence $\{x_n\}$ converges to x if and only if every subsequence of $\{x_n\}$ converges to x.

Example 2.11. Consider a sequence $x_n = (-1)^n$ that we know to be divergent. Now, consider two sequences of natural numbers, $\{n_k\} = \{1, 3, 5, \ldots\}$ and $\{m_k\} = \{2, 4, 6, \ldots\}$. The subsequence corresponding to $\{n_k\}$ is $\{-1, -1, -1, \ldots\}$ with the limit -1, and the subsequence corresponding to $\{m_k\}$ is $\{1, 1, 1, \ldots\}$ with the limit 1. Hence, it is possible for subsequences to converge even though the whole sequence does not.

2.2 Upper and lower limits

Definition 2.12. Let $\{x_n\}$ be a sequence of real numbers with the following property: for every real M there is an integer N such that $n \geq N$ implies $x_n \geq M$. We then write

$$x_n \to +\infty.$$

Similarly, if for every real M there is an integer N such that $n \geq N$ implies $x_n \leq M$, we write

$$x_n \to -\infty.$$

Definition 2.13. Let $\{x_n\}$ be a sequence or real numbers. Let E be the set of numbers x such that $x_{n_k} \to x$ for some subsequence $\{x_{n_k}\}$. This set E contains all subsequential limits as defined in the Definition 2.10, plus possibly the numbers $+\infty$, $-\infty$.

Put

$$x^\star = \sup E, \quad x_\star = \inf E.$$

The numbers $x^\star$ and $x_\star$ are called the *upper* and *lower limits* of $\{x_n\}$. We use the notation

$$\limsup_{n\to\infty} x_n = x^\star, \quad \liminf_{n\to\infty} x_n = x_\star.$$

Theorem 2.14. *If $s_n \leq t_n$ for $n \geq N$, where N is fixed, then*

$$\liminf_{n\to\infty} s_n \leq \liminf_{n\to\infty} t_n,$$
$$\limsup_{n\to\infty} s_n \leq \limsup_{n\to\infty} t_n.$$

3 Continuity

3.1 Limits of functions

Definition 3.1. Let X and Y be metric spaces; suppose $E \subset X$, f maps E into Y, and p is a limit point of E. We write $f(x) \to q$ as $x \to p$, or

$$\lim_{x\to p} f(x) = q$$

if there is a point $q \in Y$ with the following property: for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$d_Y(f(x), q) < \varepsilon$$

for all points $x \in E$ for which

$$0 < d_X(x, p) < \delta.$$

The symbols d_X and d_Y refer to the distances in X and Y, respectively.

If X and/or Y are replaced by the real line, the complex plane, or by some Euclidean space $\mathbb{R}^k$, the distances d_X, d_Y are of course replaced by absolute values, or by appropriate norms.

Corollary 3.2. *If f has a limit at p, this limit is unique.*

Definition 3.3. One can also define *one-sided (left-sided and right-sided limits)* by manipulating the definition such that it considers not all x in the δ-neighborhood of p but those x that are smaller (or larger) than p:

$$\lim_{x\to p^-} f(x) = q,$$
$$\lim_{x\to p^+} f(x) = q.$$

Theorem 3.4. *It holds that $\lim_{x\to p} f(x) = q$ if and only if $\lim_{x\to p^-} f(x) = \lim_{x\to p^+} f(x) = q$.*

3.2 Limit laws (ii)

Corollary 3.5. *If* $\lim_{x\to p} f(x)$ *and* $\lim_{x\to p} g(x)$ *exist and* c *is a constant, then*

(a) $\lim\limits_{x\to p}(f(x)+g(x)) = \lim\limits_{x\to p} f(x) + \lim\limits_{x\to p} g(x)$.

(b) $\lim\limits_{x\to p}(f(x)-g(x)) = \lim\limits_{x\to p} f(x) - \lim\limits_{x\to p} g(x)$.

(c) $\lim\limits_{x\to p}(cf(x)) = c\lim\limits_{x\to p} f(x)$.

(d) $\lim\limits_{x\to p} c = c$.

(e) $\lim\limits_{x\to p} x = p$.

(f) $\lim\limits_{x\to p}(f(x)g(x)) = \lim\limits_{x\to p} f(x)\lim\limits_{x\to p} g(x)$.

(g) $\lim\limits_{x\to p}\dfrac{f(x)}{g(x)} = \dfrac{\lim\limits_{x\to p} f(x)}{\lim\limits_{x\to p} g(x)}$ *if* $\lim\limits_{x\to p} g(x) \neq 0$.

(h) $\lim\limits_{x\to p}(f(x))^n = \left(\lim\limits_{x\to p} f(x)\right)^n$, $n \in \mathbb{N}$.

Definition 3.6. We write $f(x) \to +\infty$ as $x \to p$, or

$$\lim_{x\to p} f(x) = +\infty,$$

if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $f(x) > \varepsilon$ for every x for which $0 < |x - p| < \delta$. An example of such a function is $f(x) = x^{-1}$ with a limit $\lim_{x\to 0} f(x)$.

3.3 Continuous functions

Definition 3.7. Suppose X and Y are metric spaces, $E \subset X$, $p \in E$, and f maps E into Y. Then f is said to be *continuous at* p if for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$$d_Y(f(x), f(p)) < \varepsilon$$

for all points $x \in E$ for which $d_X(x, p) < \delta$.

If f is continuous at every point of E, then f is said to be *continuous on* E. It should be noted that f has to be defined at the point p in order to be continuous at p.

We now turn to compositions of functions. A brief statement of the following theorem is that a continuous function of a continuous function is continuous.

Theorem 3.8. *Suppose* X, Y, Z *are metric spaces,* $E \subset X$, f *maps* E *into* Y, g *maps the range of* f, $f(E)$, *into* Z, *and* h *is the mapping of* E *into* Z *defined by*

$$h(x) = g(f(x)) \quad (x \in E).$$

If f *is continuous at point* $p \in E$ *and if* g *is continuous at the point* $f(p)$, *then* h *is continuous at* p.

This function h is called the *composition* or the *composite* of f and g. The notation

$$h = g \circ f$$

is frequently used in this context.

Example 3.9. Consider two functions $f(x) = \frac{x}{2}$ and $g(x) = x^2$. We have

(a) $f \circ g = f(g(x)) = \dfrac{g(x)}{2} = \dfrac{x^2}{2}$.

(b) $g \circ f = g(f(x)) = \left(\dfrac{x}{2}\right)^2 = \dfrac{x^2}{4}$.

(c) $g \circ g = g(g(x)) = (x^2)^2 = x^4$.

Theorem 3.10. *Let* f *and* g *be functions defined on the same interval. If* $f(x)$ *and* $g(x)$ *are continuous at* p, *so are* $f(x) + g(x)$ *and* $f(x)g(x)$. *If* $g(p) \neq 0$, $f(x)/g(x)$ *is also continuous at* p.

4 Differentiation

In this section we shall confine our attention to *real* functions defined on intervals or segments.

Definition 4.1. Let f be defined (and real-valued) on $[a, b]$. For any $x \in [a, b]$ form the quotient

$$\phi(t) = \frac{f(t) - f(x)}{t - x} \quad (a < t < b, t \neq x),$$

and define

$$f'(x) = \lim_{t \to x} \phi(t), \tag{1}$$

provided that this limit exists.

We thus associate with the function f a function f' whose domain is the set of points x at which the limit (1) exists; f' is called the *derivative of* f.

If f' is defined at a point x, we say that f is *differentiable* at x. If f' is defined at every point of a set $E \subset [a, b]$, we say that f is differentiable on E.

It is possible to consider right-hand and left-hand limits in (1); this leads to the definition of right-hand and left-hand derivatives. In particular, at the endpoints a and b, the derivative, if it exists, is a right-hand or left-hand derivative respectively.

If f is defined on a segment (a, b) and if $a < x < b$, then $f'(x)$ is defined by (4.1) and (1), as above. But $f'(a)$ and $f'(b)$ are not defined in this case.

Theorem 4.2. *Let f be defined on $[a, b]$. If if is differentiable at a point $x \in [a, b]$, then f is continuous at x.*

Proof. As $t \to x$, we have

$$f(t) - f(x) = \frac{f(t) - f(x)}{t - x} \cdot (t - x) \to f'(x) \cdot 0 = 0.$$

□

The converse of this theorem is not true.

Example 4.3. Consider two functions,

$$f(x) = \begin{cases} x, & x < 0, \\ x^2, & x \geq 0, \end{cases} \quad g(x) = \begin{cases} 0, & x \leq 0, \\ 1 & x > 0. \end{cases}$$

The function $g(x)$ is discontinuous at 0, hence it is not differentiable. The function $f(x)$ is continuous at 0, but not differentiable. To show this, note

$$\lim_{x \to 0^-} \frac{f(x) - f(0)}{x} = \lim_{x \to 0^-} \frac{x - 0}{x} = 1 \neq \lim_{x \to 0^+} \frac{f(x) - f(0)}{x} = \lim_{x \to 0^+} \frac{x^2 - 0}{x} = 0.$$

Because one-sided derivatives are not equal, the derivative at 0, $f'(0)$, does not exist.

Theorem 4.4. *Suppose f and g are defined on $[a, b]$ and are differentiable at a point $x \in [a, b]$. Then $f + g$, $f \cdot g$, and f/g are differentiable at x, and*

(a) $(f + g)'(x) = f'(x) + g'(x)$.

(b) $(f \cdot g)(x) = f'(x)g(x) + f(x)g'(x)$.

(c) $(f/g)'(x) = \dfrac{f'(x)g(x) - f(x)g'(x)}{g^2(x)}$, $g(x) \neq 0$.

Example 4.5. The derivative of any constant is clearly zero. If f is defined by $f(x) = x$, then $f'(x) = 1$. Repeated application of *(b)* and *(c)* then shows that $f(x) = x^n$ is differentiable, and that its derivative is $f'(x) = nx^{n-1}$, for any integer n. Thus, every polynomial is differentiable and so is every rational function, except at the points where the denominator is zero.

Example 4.6. Consider $f(x) = x^2$, $g(x) = 1 + x$. Then we have

$$\begin{aligned} f'(x) &= 2x, \\ g'(x) &= 1, \\ (f+g)'(x) &= (x^2 + 1 + x)' = 2x + 1, \\ (f \cdot g)'(x) &= (x^2 \cdot (1+x))' = 2x \cdot (1+x) + x^2 = 2x + 3x^2, \\ \left(\frac{f(x)}{g(x)}\right)' &= \frac{2x(1+x) - x^2}{(1+x)^2} = \frac{2x + x^2}{(1+x)^2}. \end{aligned}$$

The following theorem is known as the "chain rule" for differentiation. It deals with differentiation of composite functions and is probably the most important theorem about derivatives.

Theorem 4.7. *Suppose f is continuous on $[a, b]$, $f'(x)$ exists at some point $x \in [a, b]$, g is defined on an interval I which contains the range of f, and g is differentiable at the point $f(x)$. If*

$$h(t) = g(f(t)) \quad (a \leq t \leq b),$$

then h is differentiable at x, and

$$h'(x) = g'(f(x))f'(x).$$

Example 4.8. Consider two functions, $f(x) = \frac{x}{2}$ and $g(x) = x^2$, and their composite function $h(x) = \left(\frac{x}{2}\right)^2$. Then,

$$\begin{aligned} f'(x) &= \frac{1}{2}, \\ g'(x) &= 2x, \\ h'(x) &= g'(f(x))f'(x) = \frac{x}{2}. \end{aligned}$$

4.1 Mean value theorems

Definition 4.9. Let f be a real function defined on a metric space X. We say that f has a *local maximum* at a point $p \in X$ if there exists $\delta > 0$ such that $f(q) \leq f(p)$ for all $q \in X$ with $d(p, q) < \varepsilon$.

Local minima are defined likewise. Our next theorem is the basis of many applications of differentiation.

Theorem 4.10. *Let f be defined on $[a, b]$; if f has a local maximum at a point $x \in (a, b)$, and if $f'(x)$ exists, then $f'(x) = 0$. The analogous statement for local minima is also true.*

Proof. Choose δ in accordance with Definition 4.9, so that

$$a < x - \delta < x < x + \delta < b.$$

If $x - \delta < t < x$, then

$$\frac{f(t) - f(x)}{t - x} \geq 0.$$

Letting $t \to x$, we see that $f'(x) \geq 0$.

If $x < t < x + \delta$, then

$$\frac{f(t) - f(x)}{t - x} \leq 0,$$

which shows that $f'(x) \leq 0$. Hence, $f'(x) = 0$. □

The following result is usually referred to as the mean value theorem:

Theorem 4.11. *If f is a real continuous function on $[a, b]$ which is differentiable in (a, b), then there is a point $x \in (a, b)$ at which*

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

Theorem 4.12. *Suppose f is differentiable in (a, b).*

(a) If $f'(x) \geq 0$ for all $x \in (a, b)$, then f is monotonically increasing.

(b) If $f'(x) = 0$ for all $x \in (a, b)$, then f is constant.

(c) If $f'(x) \leq 0$ for all $x \in (a, b)$, then f is monotonically decreasing.

4.2 o and $\mathcal{O}$ notation

Suppose we have a function $f(x)$ with $f(a) = 0$ and we want to consider how quickly the function goes to zero around a. Then ideally, we would want to find a simple function g (for example, $g(x) = (x-a)^n$) which also vanishes at a such that g and f are almost equal around a. The "small-o" and "big-o" notation expresses this notion, but only states that f goes to zero faster than g.

Definition 4.13. We say

$$f(x) = \mathcal{O}(g(x))$$

as $x \to a$ if there exists a constant M such that $|f(x)| \leq M\,|g(x)|$ in some punctured neighborhood of a, that is for $x \in (a-\delta, a+\delta)\backslash\{a\}$ for some value of δ.

We say

$$f(x) = o(g(x))$$

as $x \to a$ if $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$. This implies that there exists a punctured neighborhood of a on which g does not vanish.

Example 4.14. The first two examples are derived from Taylor polynomials, the rest can be checked directly:

a) $e^x = 1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3 + \mathcal{O}(x^4)$ as $x \to 0$,

b) $\frac{1}{1-x} = 1 + x + x^2 + \mathcal{O}(x^3) = 1 + x + x^2 + o(x^2)$ as $x \to 0$,

c) $\left|x^3\right| = \mathcal{O}(x^3) = o(x^2)$ as $x \to 0$,

d) $\cosh(x) = \mathcal{O}\left(e^x\right) = o\left(e^{\frac{5}{4}x}\right)$ as $x \to 0$,

e) $\frac{1}{\sin(x)} = \mathcal{O}\left(\frac{1}{x}\right) = o\left(\frac{1}{x^{\frac{3}{2}}}\right)$ as $x \to 0$.

Theorem 4.15. *The following holds:*

(a) $f(x) = \mathcal{O}\left(f(x)\right)$.

(b) If $f(x) = o\left(g(x)\right)$ *then* $f(x) = \mathcal{O}\left(g(x)\right)$.

(c) If $f(x) = \mathcal{O}\left(g(x)\right)$ *then* $\mathcal{O}\left(f(x) + g(x)\right) = \mathcal{O}\left(g(x)\right)$.

(d) If $f(x) = \mathcal{O}\left(g(x)\right)$ *then* $o\left(f(x) + g(x)\right) = o\left(g(x)\right)$.

(e) Let $c \neq 0$, *then* $c \cdot \mathcal{O}\left(g(x)\right) = \mathcal{O}\left(g(x)\right)$ *and* $c \cdot o\left(g(x)\right) = o\left(g(x)\right)$.

(f) $\mathcal{O}\left(f(x)\right)\mathcal{O}\left(g(x)\right) = \mathcal{O}\left(f(x)g(x)\right)$.

(g) $o\left(f(x)\right)\mathcal{O}\left(g(x)\right) = o\left(f(x)g(x)\right)$.

(h) If $g(x) = o(1)$ *then* $\frac{1}{1+o(g(x))} = 1 + o\left(g(x)\right)$, *and* $\frac{1}{1+\mathcal{O}(g(x))} = 1 + \mathcal{O}\left(g(x)\right)$.

In the case when functions $f(\cdot)$ and $g(\cdot)$ are polynomials these rules simplify to the following.

Corollary 4.16. *Around 0 we have*

a) $x^a = \mathcal{O}(x^b)$ *for all* $b \leq a$, *and* $x^a = o(x^b)$ *for all* $b < a$.

b) $\mathcal{O}(x^a) + \mathcal{O}(x^b) = \mathcal{O}\left(x^{\min(a,b)}\right)$, $o(x^a) + o(x^b) = o\left(x^{\min(a,b)}\right)$, *and*

$$\mathcal{O}(x^a) + o(x^b) = \begin{cases} o(x^b), & b < a, \\ \mathcal{O}(x^a), & b \geq a. \end{cases}$$

c) For $c \neq 0$, $c \cdot \mathcal{O}(x^a) = \mathcal{O}(x^a)$, *and* $c \cdot o(x^a) = o(x^a)$.

d) $x^b\mathcal{O}(x^a) = \mathcal{O}(x^{a+b})$, *and* $x^b o(x^a) = o(x^{a+b})$.

e) $\mathcal{O}(x^a)\mathcal{O}(x^b) = \mathcal{O}(x^{a+b})$, $\mathcal{O}(x^a)o(x^b) = o(x^{a+b})$, *and* $o(x^a)o(x^b) = o(x^{a+b})$.

4.3 Differentiation of functions of several variables

So far, we have focused on functions of one variable; a straightforward extension of the differentiation ideas to functions of several variables involves *partial derivatives*.

Definition 4.17. Let $f : \mathbb{R}^n \to \mathbb{R}$. Then for each x_i at each point $x = (x_1, \ldots, x_n)$ in the domain of f, the *partial derivative* of f at x is

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h \to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)}{h}$$

provided that this limit exists.

5 Integration

Definition 5.1. Let f be a function defined on $[a, b]$. Divide the interval $[a, b]$ into n subintervals of equal width, $\Delta x = (b - a)/n$. Let $x_0, x_1, \ldots, x_n$ be the endpoints of these subintervals, and let $x_1^\star, \ldots, x_n^\star$ be any points in these subintervals, so that $x_i^\star$ lies in the ith subinterval $[x_{i-1}, x_i]$. The *definite integral* of f from a to b is

$$\int_a^b f(x)dx = \lim_{n \to \infty} \sum_{i=1}^{n} f(x_i^\star)\Delta x$$

provided that the limit exists. If it does, we say that f is *integrable* on $[a, b]$.

Sometimes instead of definite integrals, we work with indefinite integrals.

Definition 5.2. *Indefinite integral* (or *antiderivative*) of the function f is defined as

$$\int f(x)dx = F(x),$$

such that $F'(x) = f(x)$.

Note that if $F(x)$ is the antiderivative of $f(x)$, then $F(x) + C$ is also the antiderivative of $f(x)$ for any constant C. Thus, an indefinite integral represents the whole family of functions.

Theorem 5.3. *If f is continuous on $[a, b]$, or if f has only a finite number of jumpdiscontinuities, then f is integrable on $[a, b]$.*

Corollary 5.4. *Let f and g be integrable on $[a, b]$, and k be a constant. Then we have*

(a) $\int_a^b k dx = k(b - a).$

(b) $\int_a^b (f(x) + g(x))\, dx = \int_a^b f(x)dx + \int_a^b g(x)dx.$

(c) $\int_a^b kf(x)dx = k \int_a^b f(x)dx.$

(d) $\int_a^b (f(x) - g(x))\, dx = \int_a^b f(x)dx - \int_a^b g(x)dx.$

(e) $\int_a^c f(x)dx + \int_c^b f(x)dx = \int_a^b f(x)dx$ *for some* $c \in [a, b]$.

(f) $\int_a^b f(x)dx = -\int_b^a f(x)dx.$

(g) $\int_a^b f(x)dx \geq \int_a^b g(x)dx$ *if* $f(x) \geq g(x)$ *for all* $x \in [a, b]$.

Lemma 5.5 (Integration by parts). *Let f and g be integrable, and assume $f'(x)$ and $g'(x)$ exist for all x. Then,*

$$\int f(x)g'(x)dx + \int g(x)f'(x)dx = f(x)g(x).$$

For definite integrals defined on $[a, b]$, it holds that

$$\int_a^b f(x)'g(x)dx + \int_a^b f(x)g(x)'dx = (f(x)g(x))\Big|_a^b.$$

Example 5.6. Consider $\int x \sin x dx$. Pick $f(x) = x$, $g(x) = -\cos x$. Then

$$\int x \sin x dx = -x\cos x - \int -\cos x dx = -x\cos x + \sin x + C.$$

Example 5.7. Consider $\int_0^\pi e^x \sin x dx$. First, pick $f(x) = e^x$ and $g(x) = -\cos x$. Then

$$\int_0^\pi e^x \sin x dx = (e^x(-\cos x)\Big|_0^\pi + \int_0^\pi e^x \cos x dx.$$

Let us integrate by parts again. Now pick $f(x) = e^x$ and $g(x) = \sin x$. Then

$$\int_0^\pi e^x \sin x dx = (e^x(-\cos x)\Big|_0^\pi + (e^x \sin x)\Big|_0^\pi - \int_0^\pi e^x \sin x dx.$$

Regrouping, we have

$$\begin{aligned}\int_0^\pi e^x \sin x dx &= \frac{1}{2}\left((e^x(-\cos x)\Big|_0^\pi + (e^x \sin x)\Big|_0^\pi\right) \\ &= \frac{1}{2}\left[e^\pi(-\cos\pi) - e^0(-\cos 0)\right] + \frac{1}{2}\left[e^\pi \sin\pi - e^0 \sin 0\right] \\ &= \frac{e^\pi + 1}{2}.\end{aligned}$$

Lemma 5.8 (Integration by substitution). *If $u = g(x)$ is a differentiable function whose range is an interval I, and f is continuous on I, then*

$$\int f(g(x))g'(x)dx = \int f(u)du.$$

If g' is continuous on $[a, b]$, and f is continuous on the range of $u = g(x)$, then

$$\int_a^b f(g(x))g'(x)dx = \int_{g(a)}^{g(b)} f(u)du.$$

Example 5.9. Consider $\int x^3 \cos(x^4 + 2)dx$. Let $u = x^4 + 2$, then $du = 4x^3 dx$ and $dx = du/4x^3$. So we have

$$\int x^3 \cos(x^2 + 4)dx = \frac{1}{4}\int \cos u du = \frac{1}{4}\sin u + C = \frac{1}{4}\sin(x^2 + 4) + C.$$

Example 5.10. Consider $\int_1^2 \frac{1}{(3-5x)^2}dx$. Let $u = g(x) = 3 - 5x$, then $du = -5dx$, and $dx = -du/5$. The lower bound is $x = 1$, hence, $u = g(1) = -2$, and the upper bound is $x = 2$, hence, $u = g(2) = -7$. We have

$$\int_1^2 \frac{1}{(3-5x)^2}dx = -\frac{1}{5}\int_{-2}^{-7} \frac{1}{u^2}du = -\frac{1}{5}\left(-\frac{1}{u}\right)\Big|_{-2}^{-7} = \frac{1}{14}.$$

Definition 5.11. If $\int_a^t f(x)dx$ exists for every $t \geq a$, then

$$\int_a^\infty f(x)dx = \lim_{t\to\infty}\int_a^t f(x)dx$$

provided this limit exists.

If $\int_t^b f(x)dx$ exists for every $t \leq b$, then

$$\int_{-\infty}^b f(x)dx = \lim_{t\to-\infty} \int_t^b f(x)dx$$

provided this limit exists. Improper integrals $\int_a^\infty f(x)dx$ and $\int_{-\infty}^b f(x)dx$ are called *convergent* if the corresponding limit exists, and *divergent* if the limit does not exist.

Definition 5.12. If f is continuous on $[a, b)$ and is discontinuous at b, then

$$\int_a^b f(x)dx = \lim_{t\to b^-} \int_a^t f(x)dx$$

provided this limit exists.

If f is continuous on $(a, b]$ and is discontinuous at a, then

$$\int_a^b f(x)dx = \lim_{t\to a^+} \int_t^b f(x)dx$$

provided this limit exists.

The improper integral $\int_a^b f(x)dx$ is called *convergent* if the corresponding limit exists, and *divergent* if the limit does not exist.

If f has a discontinuity at c, where $a < c < b$, and both $\int_a^c f(x)dx$ and $\int_c^b f(x)dx$ are convergent, then we define

$$\int_a^b f(x)dx = \int_a^c f(x)dx + \int_c^b f(x)dx.$$

Example 5.13. Consider $\int_0^3 \frac{1}{x-1}dx$. First, note that $\frac{1}{x-1}$ is not defined at $x = 1$, so it is discontinuous at $x = 1$. Then

$$\int_0^3 \frac{1}{x-1}dx = \int_0^1 \frac{1}{x-1}dx + \int_1^3 \frac{1}{x-1}dx = \lim_{t\to1^-} \int_0^t \frac{1}{x-1}dx + \lim_{t\to1^+} \int_t^3 \frac{1}{x-1}dx.$$

Consider the first term:

$$\lim_{t\to1^-} \int_0^t \frac{1}{x-1}dx = \lim_{t\to1^-} \log|x-1|\Big|_0^t = \lim_{t\to1^-} (\log|t-1| - \log|-1|) = \infty.$$

The first term diverges, hence, the whole integral diverges. Note that if we fail to take discontinuity at $x = 1$ into account, we get

$$\int_0^3 \frac{1}{x-1}dx = \log|x-1|\Big|_0^3 = \log 2,$$

which is incorrect.

Theorem 5.14. *Suppose f is continuous on $[-a, a]$.*

(a) If f is even, $f(-x) = f(x)$, then $\int_{-a}^a f(x)dx = 2\int_0^a f(x)dx$.

(b) If f is odd, $f(-x) = -f(x)$, then $\int_{-a}^a f(x)dx = 0$.

Theorem 5.15. *Suppose f is continuous on $[a, b]$.*

(a) If $g(x) = \int_a^x f(t)dt$, then $g'(x) = f(x)$.

(b) $\int_a^b f(x)dx = F(b) - F(a)$, where F is any antiderivative of f, that is, $F' = f$.

5.1 Differential equations and their systems

Consider an equation that contains y', for example, $y' = x^3$. It is called a *differential equation*, and you might think of y as of some function of x: $y = f(x)$. So the objective is to find y such that its derivative is x^3. It is straightforward that $y = \frac{1}{4}x^4 + C$ (where C is some constant) solves the differential equation $y' = x^3$. In fact, $\frac{1}{4}x^4 + C$ is the *general solution* (for any C it solves $y = x^3$, thus, it represents a family of solutions).

Now consider that, in addition to the equation $y' = x^3$, we are given some *initial condition*, for example, $y(0) = 0.5$. Then we need only solutions that satisfy this initial condition: $y = \frac{1}{4}x^4 + C$ such that $y(0) = 0.5$. Note that $y(0) = C$, then the *particular solution* to this equation that satisfies the given initial condition is $y = \frac{1}{4}x^4 + 0.5$. Here, we focus on two types of the differential equations.

Definition 5.16. The differential equation $y' = f(x, y)$ is *separable* if we can factorize $f(x, y) = g_1(x)g_2(y)$.

In this case, it holds that

$$\frac{dy}{dx} = g_1(x)g_2(y), \tag{2}$$

$$\frac{dy}{g_2(y)} = g_1(x)dx, \tag{3}$$

$$\int \frac{dy}{g_2(y)} = \int g_1(x)dx. \tag{4}$$

Example 5.17. Consider $y' = \dfrac{x^2}{y^2}$. It is separable, with $g_1(x) = x^2$ and $g_2(y) = y^{-2}$. Now, using (4) we have

$$\begin{aligned}
\int y^2 dy &= \int x^2 dx, \\
\frac{1}{3}y^3 + C_1 &= \frac{1}{3}x^3 + C_2, \\
y^3 &= x^3 + C, \\
y &= \sqrt[3]{x^3 + C},
\end{aligned}$$

which is the general solution. Now, suppose there is an initial condition $y(0) = 2$. Then, $\sqrt[3]{0 + C} = 2 \Rightarrow C = 8$. Hence, the corresponding particular solution is $y = \sqrt[3]{x^3 + 8}$.

Definition 5.18. A differential equation $y' = f(x, y)$ is *first-order linear* if we can represent it by $y' + P(x)y = Q(x)$, where $P(x)$ and $Q(x)$ are some continuous functions.

After, it is useful to multiply both sides by $\exp\left(\int P(x)dx\right)$.

Example 5.19. Consider $y' + 3x^2y = 6x^2$. It is not separable but it is first-order linear. Multiply both sides by e^{x^2} (note that $\exp\left(\int P(x)dx\right) = e^{x^3}e^C$ but $e^C > 0$ is a non-zero constant, so we can omit it):

$$\begin{aligned}
\frac{dy}{dx}e^{x^3} + y3x^2e^{x^3} &= 6x^2e^{x^3}, \\
\frac{d(e^{x^3} \cdot y)}{dx} &= 6x^2e^{x^3}, \\
\int \frac{d(e^{x^3} \cdot y)}{dx}dx &= \int 6x^2e^{x^3}, \\
e^{x^3}y &= 2e^{x^3} + C, \\
y &= 2 + \frac{C}{e^{x^3}}.
\end{aligned}$$

6 Topics in optimization

6.1 Unconstrained optimization

Definition 6.1. Let $F : U \to \mathbb{R}$ be a real-valued function of n variables, whose domain U is a subset of $\mathbb{R}^n$.

(a) A point $\mathbf{x}^\star \in U$ is a *maximum* of F on U if $F(\mathbf{x}^\star) \geq F(\mathbf{x})$ for all $\mathbf{x} \in U$.

(b) $\mathbf{x}^\star \in U$ is a *strict maximum* if $\mathbf{x}^\star$ is a maximum and $F(\mathbf{x}^\star) > F(\mathbf{x})$ for all $\mathbf{x} \neq \mathbf{x}^\star$ in U.

(c) $\mathbf{x}^\star \in U$ is a *local maximum* of F if there is a ball $B_r(\mathbf{x}^\star)$ such that $F(\mathbf{x}^\star) \geq F(\mathbf{x})$ for all $\mathbf{x} \in B_r(\mathbf{x}^\star) \cup U$.

In other words, a point $\mathbf{x}^\star$ is a local maximum if there are no nearby points at which F takes on a larger values. Of course, a maximum is always a local maximum. If we want to emphasize that a point $\mathbf{x}^\star$ is a maximum of F on the whole domain U, not just a local maximum, we call $\mathbf{x}^\star$ a global maximum of F on U.

Reversing the inequalities in the above definitions leads to the definitions of a global minimum, a strict global minimum, and a local minimum, respectively.

Theorem 6.2. *Let $F : U \to \mathbb{R}$ be a C^1 function defined on a subset U of $\mathbb{R}^n$. If $\mathbf{x}^\star$ is a local maximum or minimum of F in U and if $\mathbf{x}^\star$ is an interior point of U, then*

$$\frac{\partial F}{\partial x_i}(\mathbf{x}^\star) = 0, \quad i = 1, \ldots, n.$$

The first-order condition for a point $x^\star$ to be a maximum or minimum of a function f of one variable is that $f'(x^\star) = 0$, that is, that $x^\star$ be a *critical point* of f. This condition requires that $x^\star$ not be an endpoint of the interval under consideration, in other words, that $x^\star$ lie in the interior of the domain of f. The same first order condition works for a function F of n variables. However, a function of n variables has n first derivatives: the partials $\partial F/\partial x_i = 0$ at $\mathbf{x}^\star$. The n-dimensional analogue of $f'(x^\star) = 0$ is that each $\partial F/\partial x_i = 0$ at $\mathbf{x}^\star$. In this case, $\mathbf{x}^\star$ is an *interior point* of the domain of F if there is a whole ball $B_r(\mathbf{x}^\star)$ about $\mathbf{x}^\star$ in the domain of F.

Example 6.3. To find the local maxima and minima of $F(x, y) = x^3 - y^3 + 9xy$, one computes the first order partial derivatives and sets them equal to zero:

$$\frac{\partial F}{\partial x} = 3x^2 + 9y = 0, \quad \frac{\partial F}{\partial y} = -3y^2 + 9x = 0. \tag{5}$$

The first equation yields $y = -\frac{1}{3}x^2$. Substitute into the second equation to get

$$0 = -3y^2 + 9x = -\frac{1}{3}x^4 + 9x.$$

This equation is equivalent to $27x - x^4 = x(27 - x^3) = 0$ whose solutions are $x = 0$ and $x = 3$. Substituting these solutions into $y = -\frac{1}{3}x^2$ implies that the solutions to (5) are the two points $(0, 0)$ and $(3, -3)$. At this stage, we can conclude that the only candidates for a local maximum or minimum of F are these two points. We are unable to say whether either of these two is a maximum or a minimum.

Definition 6.4. We say that the n-vector $\mathbf{x}^\star$ is a *critical point* of a function $F(x_1, \ldots, x_n)$ if $\mathbf{x}^\star$ satisfies

$$\frac{\partial F}{\partial x_i}(\mathbf{x}^\star) = 0, \quad i = 1, \ldots, n.$$

The critical points of $F(x, y) = x^3 - y^3 + 9xy$ in Example 6.3 are $(0, 0)$ and $(3, -3)$. To determine whether either of these of these critical points is a maximum or a minimum, we need to use a condition on the second derivatives of F.

Definition 6.5. A C^2 function of n variables has n^2 second order partial derivatives at each point in its domain and it is natural to combine then into an $n \times n$ matrix, called *the Hessian* of F:

$$D^2F(\mathbf{x}^\star) = \begin{pmatrix} \frac{\partial^2 F}{\partial x_1^2}(\mathbf{x}^\star) & \cdots & \frac{\partial^2 F}{\partial x_n \partial x_1}(\mathbf{x}^\star) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_1 \partial x_n}(\mathbf{x}^\star) & \cdots & \frac{\partial^2 F}{\partial x_n^2}(\mathbf{x}^\star) \end{pmatrix}.$$

Since cross-partial derivatives are equal for a C^2 function, $D^2F(\mathbf{x}^\star)$ is a symmetric matrix.

Theorem 6.6. *Let $F : U \to \mathbb{R}$ be a C^2 function whose domain is an open set U in $\mathbb{R}^n$. Suppose that $\mathbf{x}^\star$ is a critical point of F in that it satisfies Definition 6.4.*

(a) If the Hessian $D^2F(\mathbf{x}^\star)$ is a negative definite symmetric matrix, then $\mathbf{x}^\star$ is a strict local maximum of F.

(b) If the Hessian $D^2F(\mathbf{x}^\star)$ is a positive definite symmetric matrix, then $\mathbf{x}^\star$ is a strict local minimum of F.

(c) If $D^2F(\mathbf{x}^\star)$ is indefinite, then $\mathbf{x}^\star$ is neither a local maximum nor a local minimum of F.

The second order condition for a critical point $x^\star$ of a function f on $\mathbb{R}$ to be a maximum is that the second derivative $f''(x^\star)$ be negative. The corresponding condition for a function F of n variables is that the second derivative $D^2F(\mathbf{x}^\star)$ be *negative definite* as a symmetric matrix at the critical point $\mathbf{x}^\star$. Similarly, the second order sufficient condition for a critical point of a function f of one variable to be a local minimum is that $f''(x^\star)$ be positive; the analogous second order condition fo an n-dimensional critical point $\mathbf{x}^\star$ to be a local minimum is that the Hessian of F at $\mathbf{x}^\star$, $D^2F(\mathbf{x}^\star)$, be *positive definite*.

Definition 6.7. A critical point $\mathbf{x}^\star$ of F for which the Hessian $D^2F(\mathbf{x}^\star)$ is indefinite is called a *saddle point* of F.

A saddle point $\mathbf{x}^\star$ is a minimum of F in some directions and a maximum of F in other directions.

6.2 Constrained optimization

Theorem 6.8. *Let f and h be C^1 functions of two variables. Suppose that $\mathbf{x}^\star = (x_1^\star, x_2^\star)$ is a solution of the problem*

$$\max f(x_1, x_2) \quad \textit{subject to} \quad h(x_1, x_2) = c.$$

Suppose further that $(x_1^\star, x_2^\star)$ is not a critical point of h. Then, there is a real number $\mu^\star$ such that $(x_1^\star, x_2^\star, \mu^\star)$ is a critical point of the Lagrangian function

$$\mathcal{L}(x_1, x_2, \mu) := f(x_1, x_2) - \mu\left(h(x_1, x_2) - c\right).$$

In other words, at $(x_1^\star, x_2^\star, \mu^\star)$

$$\frac{\partial \mathcal{L}}{\partial x_1} = 0, \quad \frac{\partial \mathcal{L}}{\partial x_2} = 0, \quad \frac{\partial \mathcal{L}}{\partial \mu} = 0.$$

Remark 6.9. If we were minimizing f instead of maximizing f on the constrained set C_h, we would have used the same arguments that we used in the proof of Theorem 6.8. In other words, the conclusion of Theorem 6.8 holds whether we are maximizing f or minimizing f on C_h.

Example 6.10. To solve the maximization problem

$$\max f(x_1, x_2) = x_1 x_2, \quad \text{subject to} \quad h(x_1, x_2) := x_1 + 4x_2 = 16,$$

we use Theorem 6.8. Since the gradient $\nabla h(x_1, x_2) = \begin{pmatrix} \frac{\partial h}{\partial x_1} \\ \frac{\partial h}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$, h has no critical points and the constraint qualification is satisfied. Form the Lagrangian

$$\mathcal{L}(x_1, x_2, \mu) = x_1 x_2 - \mu(x_1 + 4x_2 - 16),$$

and set its partial derivatives equal to zero:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} &= x_2 - \mu = 0, \\ \frac{\partial \mathcal{L}}{\partial x_2} &= x_1 - 4\mu = 0, \\ \frac{\partial \mathcal{L}}{\partial \mu} &= -(x_1 + 4x_2 - 16) = 0. \end{aligned}$$

From the first two equations,

$$\mu = x_2 = \frac{1}{4} x_1,$$

and therefore, $x_1 = 4x_2$. Substituting into the third equation,

$$4x_2 + 4x_2 = 16, \quad \text{or} \quad x_2 = 2.$$

We conclude that the solution is

$$x_1 = 8, \quad x_2 = 2, \quad \mu = 2.$$

Theorem 6.8 states that the only candidate for a solution is $x_1 = 8$, $x_2 = 2$.

The statement of the general theorem for maximizing a function of n variables constrained by m equality constraints is a straightforward generalization of Theorem 6.8.

Theorem 6.11. *Let $f, h_1, \ldots, h_m$ be C^1 functions of n variables. Consider the problem of maximizing (or minimizing) $f(\mathbf{x})$ on the constrained set*

$$C_{\mathbf{h}} := \{\mathbf{x} = (x_1, \ldots, x_n) : h_1(\mathbf{x}) = a_1, \ldots, h_m(\mathbf{x}) = a_m\}.$$

Suppose that $x^\star \in C_{\mathbf{h}}$ and that $\mathbf{x}^\star$ is a local maximum or minimum of f on $C_{\mathbf{h}}$. Suppose further that $\mathbf{x}^\star$ satisfies the nondegenerate constraint qualification at $\mathbf{x}^\star$. Then, there exist $\mu_1^\star, \ldots, \mu_m^\star$ such that $(x_1^\star, \ldots, x_n^\star, \mu_1^\star, \ldots, \mu_m^\star) := (\mathbf{x}^\star, \mu^\star)$ is a critical point of the Lagrangian

$$\mathcal{L}(\mathbf{x}, \mu) := f(\mathbf{x}) - \mu_1(h_1(\mathbf{x}) - a_1) - \ldots - \mu_m(h_m(\mathbf{x}) - a_m).$$

In other words, at $(\mathbf{x}^\star, \mu^\star)$

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial x_1} = 0, \quad \ldots, \quad \frac{\partial \mathcal{L}}{\partial x_n} = 0, \\ \frac{\partial \mathcal{L}}{\partial \mu_1} = 0, \quad \ldots, \quad \frac{\partial \mathcal{L}}{\partial \mu_m} = 0. \end{aligned}$$

Theorem 6.12. *Suppose that f and g are C^1 functions on $\mathbb{R}^2$ and that $(x^\star, y^\star)$ maximizes f on the constraint set $g(x, y) \leq b$. If $g(x^\star, y^\star) = b$, suppose that at $(x^\star, y^\star)$*

$$\frac{\partial g}{\partial x} \neq 0 \quad \textit{or} \quad \frac{\partial g}{\partial y} \neq 0.$$

In any case, form the Lagrandian function

$$\mathcal{L}(x, y, \lambda) = f(x, y) - \lambda \cdot (g(x, y) - b).$$

Then, there is a multiplier $\lambda^\star$ such that

(a) $\dfrac{\partial \mathcal{L}}{\partial x}(x^\star, y^\star, \lambda^\star) = 0,$

(b) $\dfrac{\partial \mathcal{L}}{\partial y}(x^\star, y^\star, \lambda^\star) = 0,$

(c) $\lambda^\star \cdot (g(x^\star, y^\star) - b) = 0,$

(d) $\lambda^\star \geq 0,$

(e) $g(x^\star, y^\star) \leq b.$

Remark 6.13. In some texts, the constraint is written as $g(x, y) \geq b$ instead of as $g(x, y) \leq b$ and the Lagrangian is then written as $\mathcal{L}(x, y, \lambda) = f(x, y) + \lambda \cdot (g(x, y) - b)$. These two changes cancel each other so that the conclusion of Theorem 6.12 still holds at a constrained maximum.

Remark 6.14. Notice the similarities and differences between the statement of Theorem 6.8 which treats equality constraints and the statement of Theorem 6.12 which covers inequality constraints:

(a) Both use the same Lagrangian $\mathcal{L}$ and both require that the derivatives of $\mathcal{L}$ with respect to the x_i's be zero.

(b) The condition that $\partial L/\partial \mu = h(x,y) - c = 0$ for equality constraints may no longer hold for inequality constraints since the constraint need not be binding at the maximizer in the inequality constraint case. It is replaced by two conditions:

$$\lambda \cdot (g(x,y) - b) = 0 \quad \text{and} \quad \frac{\partial \mathcal{L}}{\partial \lambda} = g(x,y) - b \leq 0.$$

The second of these two conditions is simply a repetition of the inequality constraint itself.

(c) Both situations require that we check a constraint qualification. However, we need only check the constraint qualification for an inequality constraint if that constraint is binding at the solution candidate.

(d) There were no restrictions on the sign of the multiplier in the equality constraint situation; however, the multiplier for inequality constraints must be nonnegative.

(e) For equality constraints, the same first order conditions that work for maximization problems also hold for minimization problems. However, the argument that $\nabla f(\mathbf{p})$ and $\nabla g(\mathbf{p})$ point in the *same* direction for inequality constrains holds only for the maximization problem. The same line of reasoning concludes that $\nabla f(\mathbf{p})$ and $\nabla g(\mathbf{p})$ must point in *opposite* directions in a constrained minimization problem.

Example 6.15. Consider the problem of maximizing $f(x,y) = xy$ on the constraint set $g(x,y) = x^2 + y^2 \leq 1$. The only critical of g occurs at the origin – far away from the boundary of the constraint set $x^2 + y^2 = 1$. So, the constraint qualification will be satisfied at any candidate for a solution. Form the Lagrangian

$$\mathcal{L}(x,y,\lambda) = xy - \lambda(x^2 + y^2 - 1),$$

and write out the first order conditions described in Theorem 6.12:

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial x} = y - 2\lambda x = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = x - 2\lambda y = 0, \\ &\lambda(x^2 + y^2 - 1) = 0, \quad x^2 + y^2 \leq 1, \quad \lambda \geq 0. \end{aligned}$$

The first two equations yield

$$\lambda = \frac{y}{2x} = \frac{x}{2y}, \quad \text{or} \quad x^2 = y^2. \tag{6}$$

If $\lambda = 0$, then $x = y = 0$. This combination satisfies all the first order conditions, so it is a candidate for a solution. If $\lambda \neq 0$, then the third equation becomes $x^2 + y^2 - 1 = 0$. Combining this with (6), we find that $x^2 = y^2 = 1/2$, or $x = \pm 1/\sqrt{2}$, $y = \pm 1/\sqrt{2}$. Combining these with the equation for λ in (6), we find the following four candidates:

$$\begin{aligned} &x = \frac{1}{\sqrt{2}}, \quad y = \frac{1}{\sqrt{2}}, \quad \lambda = \frac{1}{2}, \\ &x = -\frac{1}{\sqrt{2}}, \quad y = -\frac{1}{\sqrt{2}}, \quad \lambda = \frac{1}{2}, \\ &x = \frac{1}{\sqrt{2}}, \quad y = -\frac{1}{\sqrt{2}}, \quad \lambda = -\frac{1}{2}, \\ &x = -\frac{1}{\sqrt{2}}, \quad y = \frac{1}{\sqrt{2}}, \quad \lambda = -\frac{1}{2}. \end{aligned}$$

We disregard the last two candidates since they involve a negative multiplier. So, including $(0,0,0)$, there are three candidates which satisfy all five first order conditions. Plugging these three into the objective function, we find that

$$x = \frac{1}{\sqrt{2}}, y = \frac{1}{\sqrt{2}} \quad \text{and} \quad x = -\frac{1}{\sqrt{2}}, y = -\frac{1}{\sqrt{2}}$$

are the solutions of our original problem.

6.3 Kuhn-Tucker formulation

The most common constrained maximization problems in economics involve only inequality constraints and a complete set of nonnegativity constraints:

$$\max f(x_1, \ldots, x_n) \quad \text{subject to} \quad g_1(x_1, \ldots, x_n) \leq b_1, \ldots, g_k(x_1, \ldots, x_n) \leq b_k, \tag{7}$$

$$x_1 \geq 0, \ldots, x_n \geq 0. \tag{8}$$

If we use the techniques as before to solve the problem (7), we would write the Lagrangian as

$$\begin{aligned} &\mathcal{L}(\mathbf{x}, \lambda_1, \lambda_k, \nu_1, \ldots, \nu_n) \\ &= f(\mathbf{x}) - \lambda_1 \cdot (g_1(\mathbf{x}) - b_1) - \ldots - \lambda_k \cdot (g_k(\mathbf{x}) - b_k) + \nu_1 x_1 + \ldots + \nu_n x_n. \end{aligned}$$

Kuhn and Tucker worked with a Lagrangian $\tilde{\mathcal{L}}$ which did *not* include the nonnegativity constraints:

$$\tilde{\mathcal{L}}(\mathbf{x}, \lambda_1, \ldots, \lambda_k) = f(\mathbf{x}) - \lambda_1(g_1(\mathbf{x}) - b_1) - \ldots - \lambda_k(g_k(\mathbf{x}) - b_k).$$

It can be shown that the first order conditions in terms of the Kuhn-Tucker Lagrangian are

$$\begin{gathered} \frac{\partial \tilde{\mathcal{L}}}{\partial x_1} \leq 0, \ldots, \frac{\partial \tilde{\mathcal{L}}}{\partial x_n} \leq 0, \quad \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda_1} \geq 0, \ldots, \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda_n} \geq 0, \\ x_1 \frac{\partial \tilde{\mathcal{L}}}{\partial x_1} = 0, \ldots, x_n \frac{\partial \tilde{\mathcal{L}}}{\partial x_n} = 0, \quad \lambda_1 \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda_1} = 0, \ldots, \lambda_n \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda_n} = 0. \end{gathered}$$

The Kuhn-Tucker formulation is sometimes more advantageous because it involves $n + k$ equations in $n + k$ unknowns, compared with the $2n + k$ equations in $2n + k$ unknowns as in the original formulation.

Example 6.16. In this framework, the Kuhn-Tucker Lagrangian for the usual utility maximization problem would be

$$\tilde{\mathcal{L}}(x_1, x_2, \lambda) = U(x_1, x_2) - \lambda(p_1 x_1 + p_2 x_2 - I),$$

and the first order conditions are

$$\begin{gathered} \frac{\partial U}{\partial x_1} - \lambda p_1 \leq 0, \quad \frac{\partial U}{\partial x_2} - \lambda p_2 \leq 0, \\ x_1 \cdot \left(\frac{\partial U}{\partial x_1} - \lambda p_1 \right) = 0, \quad x_2 \cdot \left(\frac{\partial U}{\partial x_2} - \lambda p_1 \right) = 0, \\ \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda} = I - p_1 x_1 - p_2 x_2 \geq 0, \quad \lambda \frac{\partial \tilde{\mathcal{L}}}{\partial \lambda} = \lambda \left(I - p_1 x_1 - p_2 x_2 \right) = 0. \end{gathered}$$

7 Topics in linear algebra

7.1 Vector spaces

Definition 7.1. Consider a *vector space* of a set V along with two operations "$+$" and "$\cdot$" such that

1. if $\mathbf{v}, \mathbf{w} \in V$, then their *vector sum* $\mathbf{v} + \mathbf{w} \in V$, and
   - $\mathbf{v} + \mathbf{w} = \mathbf{w} + \mathbf{v}$,
   - $(\mathbf{v} + \mathbf{w}) + \mathbf{u} = \mathbf{v} + (\mathbf{w} + \mathbf{u})$ for some $\mathbf{u} \in V$,
   - there is a zero vector $\mathbf{0} \in V$ such that $\mathbf{v} + \mathbf{0} = \mathbf{v}$ for any $\mathbf{v} \in V$,
   - each $\mathbf{v} \in V$ has an additive inverse $\mathbf{w} \in V$ such that $\mathbf{w} + \mathbf{v} = \mathbf{0}$;
2. if r, s are scalars and $\mathbf{v}, \mathbf{w} \in V$, then each scalar multiple $r \cdot \mathbf{v} \in V$, and
   - $(r + s) \cdot \mathbf{v} = r \cdot \mathbf{v} + s \cdot \mathbf{v}$,
   - $r \cdot (\mathbf{v} + \mathbf{w}) = r \cdot \mathbf{v} + r \cdot \mathbf{w}$,
   - $(r \cdot s) \cdot \mathbf{v} = r \cdot (s \cdot \mathbf{v})$,
   - $1 \cdot \mathbf{v} = \mathbf{v}$.

Example 7.2. $\mathbb{R}^2$ is a vector space; $\mathbb{R}^2_+$ is not a vector space; $\{(x, y) \mid y = cx\}$ for some $c \in \mathbb{R}$ is a vector space; is a vector space.

Definition 7.3. A *subspace* is a subset of a vector space that is a vector space itself, under the inherited operations.

Definition 7.4. A subset of a vector space is *linearly independent* if none of its elements is a linear combination of the others. Otherwise, it is *linearly dependent*.

Example 7.5. Consider the following elements of $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 2 \\ 2 \\ 0 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_4 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v}_5 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

Any subset that contains $\mathbf{v}_5$ is linearly dependent: zero vector is a linear combination of any other vectors with zero coefficients. The subset $\{\mathbf{v}_1, \mathbf{v}_3, \mathbf{v}_4\} \subset \mathbb{R}^3$ is linearly independent: for example, the third coordinate is non-zero in only one of the vectors, it cannot be a combination of the other two. The subset $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\} \subset \mathbb{R}^3$ is linearly dependent as $2\mathbf{v}_1 = \mathbf{v}_2$.

Definition 7.6. Let $\alpha_1, \ldots, \alpha_n$ be some scalars. A *linear combination* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is

$$\mathbf{w} = \alpha_1 \mathbf{v}_1 + \ldots + \alpha_n \mathbf{v}_n.$$

If all vectors $\mathbf{w} \in V$ are linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_n$, we say that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ *span the space* V.

Definition 7.7. A *basis* for V is a collection of vectors that

1. are linearly independent, and
2. span the space V.

Example 7.8. Consider $\mathbb{R}^3$ and the following vectors:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The "natural" basis is $(\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3)$. Some other possibilities are $(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3)$, $(\mathbf{v}_1, \mathbf{e}_2, \mathbf{e}_3)$, $(\mathbf{v}_2, \mathbf{e}_2, \mathbf{e}_3)$.

Definition 7.9. Any two bases for V contain the same number of vectors; this number is called the *dimension of* V.

Example 7.10. Consider $V = \{(x, y) \mid y = x\}$. One possible basis is a vector $(1, 1)$. Other possibilities are the multiples of this vector; V is a line, so any element of it can be obtained by multiples of $(1, 1)$. The dimension of V is one (one element in the basis).

7.2 Vector properties

Lemma 7.11. *Consider a scalar α, and vectors*

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

The following operations holds:

(a) vector addition,

$$\mathbf{v} + \mathbf{w} = \begin{pmatrix} v_1 + w_1 \\ \vdots \\ v_n + w_n \end{pmatrix},$$

(b) scalar multiplication,

$$\alpha\mathbf{v} = \begin{pmatrix} \alpha v_1 \\ \vdots \\ \alpha v_n \end{pmatrix},$$

(c) dot product (inner product),

$$\mathbf{v}'\mathbf{w} = \begin{pmatrix} v_1 & \dots & v_n \end{pmatrix} \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = v_1 w_1 + \ldots + v_n w_n.$$

(d) outer product,

$$\mathbf{v}\mathbf{w}' = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \begin{pmatrix} w_1 & \dots & w_n \end{pmatrix} = \begin{pmatrix} v_1 w_1 & \dots & v_1 w_n \\ \vdots & \ddots & \vdots \\ v_n w_1 & \dots & v_n w_n \end{pmatrix}.$$

7.3 Matrix properties

Lemma 7.12. *Consider a scalar α and matrices*

$$\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & \dots & b_{1k} \\ \vdots & \ddots & \vdots \\ b_{n1} & \dots & b_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k}.$$

The following operations holds:

(a) matrix addition,

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} a_{11} + b_{11} & \dots & a_{1k} + b_{1k} \\ \vdots & \ddots & \vdots \\ a_{n1} + b_{n1} & \dots & a_{nk} + b_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k},$$

(b) scalar multiplication,

$$\alpha\mathbf{A} = \begin{pmatrix} \alpha a_{11} & \dots & \alpha a_{1k} \\ \vdots & \ddots & \vdots \\ \alpha a_{n1} & \dots & \alpha a_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k},$$

(c) matrix multiplication,

$$\mathbf{A}\mathbf{B}' = \begin{pmatrix} \sum_{i=1}^{k} a_{1i} b_{1i} & \dots & \sum_{i=1}^{k} a_{1i} b_{ni} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^{k} a_{ni} b_{1i} & \dots & \sum_{i=1}^{k} a_{ni} b_{ni} \end{pmatrix} \in \mathbb{R}^{n \times n},$$

(d) Kronecker product,

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & \dots & a_{1k}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{1k}\mathbf{B} & \dots & a_{nk}\mathbf{B} \end{pmatrix} \in \mathbb{R}^{n^2 \times k^2},$$

(e) Hadamard product,

$$\mathbf{A} \odot \mathbf{B} = \begin{pmatrix} a_{11}b_{11} & \dots & a_{1k}b_{1k} \\ \vdots & \dots & \vdots \\ a_{1k}b_{1k} & \dots & a_{nk}b_{nk} \end{pmatrix} \in \mathbb{R}^{n \times k}.$$

One important case of the matrix manipulation is the following: consider a matrix $\mathbf{X} \in \mathbb{R}^{n \times k}$ and a vector $\boldsymbol{\beta} \in \mathbb{R}^k$. Then the product is a vector $\mathbf{X}\boldsymbol{\beta} \in \mathbb{R}^n$:

$$\mathbf{X}\boldsymbol{\beta} = \begin{pmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} x_{11}\beta_1 + \ldots + x_{1k}\beta_k \\ \vdots \\ x_{n1}\beta_1 + \ldots + x_{nk}\beta_k \end{pmatrix}$$
$$= \begin{pmatrix} x_{11} \\ \vdots \\ x_{n1} \end{pmatrix} \beta_1 + \ldots + \begin{pmatrix} x_{1k} \\ \vdots \\ x_{nk} \end{pmatrix} \beta_k.$$

Hence, some matrix $\mathbf{X}$ times some vector $\boldsymbol{\beta}$ is a linear combination of columns of $\mathbf{X}$ with coefficients equal to elements of $\boldsymbol{\beta}$; hence, columns of $\mathbf{X}$ span some vector space and $\mathbf{X}\boldsymbol{\beta}$ is an element of that space.

Example 7.13. Consider a vector space of all 2×2 matrices. The "natural" basis is

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Definition 7.14. A square matrix $\mathbf{X}$ with elements x_{ij}, such that $x_{ij} = x_{ji}$ for all i, j is called a *symmetric* matrix.

Note that the definition above is equivalent to $\mathbf{X}' = \mathbf{X}$.

Example 7.15. Consider a vector space of all symmetric 2×2 matrices. The "natural" basis is

$$\left(\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right).$$

Definition 7.16. A square matrix $\mathbf{X}$ with elements x_{ij}, such that $x_{ij} = 0$ for all $i \neq j$ is called a *diagonal* matrix.

An important example of a diagonal matrix is the *identity* matrix $\mathbf{I}_n \in \mathbb{R}^{n \times n}$,

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix},$$

for which holds $\mathbf{I}_n\mathbf{X} = \mathbf{X}$ for all matrices $\mathbf{X}$.

Example 7.17. Consider a vector of ones $\mathbf{r}_n \in \mathbf{R}^n$. Its outer product with itself is

$$\mathbf{r}_n\mathbf{r}_n' = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} = \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} \in \mathbf{R}^{n \times n}.$$

Also,

$$\mathbf{I}_k \otimes \mathbf{r}_n\mathbf{r}_n' = \begin{pmatrix} \mathbf{r}_n\mathbf{r}_n' & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{r}_n\mathbf{r}_n' & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{r}_n\mathbf{r}_n' \end{pmatrix} \in \mathbf{R}^{nk \times nk},$$

which is an example of a *block-diagonal matrix*: the main diagonal consists of non-zero blocks, and zeros outside of these blocks.

There is another useful property of the matrix multiplication, called *block multiplication*. Consider a matrix $\mathbf{A} \in \mathbb{R}^{n\times k}$ and a matrix $\mathbf{B} \in \mathbb{R}^{k\times m}$ that can be partitioned into conformable blocks:

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_1 & \mathbf{A}_2 \\ \hline \mathbf{A}_3 & \mathbf{A}_4 \end{array}\right), \quad \mathbf{B} = \left(\begin{array}{c|c} \mathbf{B}_1 & \mathbf{B}_2 \\ \hline \mathbf{B}_3 & \mathbf{B}_4 \end{array}\right).$$

Then,

$$\mathbf{AB} = \left(\begin{array}{c|c} \mathbf{A}_1\mathbf{B}_1 + \mathbf{A}_2\mathbf{B}_3 & \mathbf{A}_1\mathbf{B}_2 + \mathbf{A}_2\mathbf{B}_4 \\ \hline \mathbf{A}_3\mathbf{B}_1 + \mathbf{A}_4\mathbf{B}_3 & \mathbf{A}_3\mathbf{B}_2 + \mathbf{A}_4\mathbf{B}_4. \end{array}\right)$$

7.4 Rank, determinant, inverse

Definition 7.18. The *rank* of the matrix $\mathbf{A}$ is the number of linearly independent rows of $\mathbf{A}$, or the number of linearly independent columns of $\mathbf{A}$ (which are equal).

The usual way to compute the rank is to transform the matrix into the echelon form. If there are no rows that contain only zeros, the matrix is of *full rank*, the rank being equal to the minimum of rows and columns (e.g., if $\mathbf{A}$ is $n \times k$, $\text{rank}(\mathbf{A}) \leq \min\{n, k\}$ with equality if and only if $\mathbf{A}$ of full rank). Otherwise, the rank is equal to the number of non-zero rows.

Definition 7.19. Every square matrix $\mathbf{A}$ has a number associated with it, which we call a *determinant*. We denote it by $\det(\mathbf{A})$ or $|\mathbf{A}|$.

The determinant encodes a lot of information about the matrix; the matrix is invertible exactly when the determinant is non-zero.

Consider a 2×2 matrix $\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. Then its determinant is given as

$$|\mathbf{A}| = ad - bc.$$

Consider a 3×3 matrix $\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$. Then its determinant is given as

$$|\mathbf{B}| = b_{11}\begin{vmatrix} b_{22} & b_{23} \\ b_{32} & b_{33} \end{vmatrix} - b_{12}\begin{vmatrix} b_{21} & b_{23} \\ b_{31} & b_{33} \end{vmatrix} + b_{13}\begin{vmatrix} b_{21} & b_{22} \\ b_{31} & b_{32} \end{vmatrix}.$$

The general formula for computing the determinant using *cofactors* C_{ij} is

$$|\mathbf{A}| = a_{11}C_{11} + a_{12}C_{12} + \ldots + a_{1n}C_{1n},$$

where cofactor C_{ij} is the determinant of the submatrix obtained from $\mathbf{A}$ by crossing out the ith row and jth column, multiplied by $(-1)^{i+j}$.

Corollary 7.20. *Below, we list the properties of the determinant.*

(a) $|\mathbf{I}_n| = 1$.

(b) If two rows of the matrix are exchanged, the sign of the determinant is reversed.

(c)
- *If one row of the matrix is multiplied by t, the determinant is multiplied by t:* $\begin{vmatrix} ta & tb \\ c & d \end{vmatrix} = t\begin{vmatrix} a & b \\ c & d \end{vmatrix}$.
- *The determinant is a linear function of the rows of the matrix:*

$$\begin{vmatrix} a + a' & b + b' \\ c & d \end{vmatrix} = \begin{vmatrix} a & b \\ c & d \end{vmatrix} + \begin{vmatrix} a' & b' \\ c & d \end{vmatrix}.$$

(d) If two rows of a matrix are equal, its determinant is zero.

(e) If $i \neq j$, subtracting t times row i from row j does not change the determinant.

(f) If A has a row that is all zeros, then $|\mathbf{A}| = 0$.

(g) The determinant of a triangular matrix is the product of the diagonal entries (pivots) $d_1, d_2, \ldots, d_n$.

(h) $|\mathbf{A}| = 0$ *exactly when* $\mathbf{A}$ *is singular.*

(i) $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

(j) $|\mathbf{A}'| = |\mathbf{A}|$.

Definition 7.21. The *inverse* of a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is a matrix $\mathbf{B} \in \mathbb{R}^{n\times n}$ such that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}_n$. There is at most one such $\mathbf{B}$, and it is denoted by $\mathbf{A}^{-1}$:

$$\mathbf{A^{-1}A} = \mathbf{I}_n, \quad \mathbf{AA^{-1}} = \mathbf{I}_n.$$

If $\mathbf{A}^{-1}$ exists, $\mathbf{A}$ is called *invertible*.

Note that using properties *(a)* and *(i)* from Corollary 7.20, we have $|\mathbf{A}| = \dfrac{1}{|\mathbf{A}^{-1}|}$.

The general formula for the inverse of a matrix $\mathbf{A}$ is

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|}\mathbf{C}',$$

where $\mathbf{C}$ is a matrix of cofactors.

Example 7.22. Consider a matrix $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$. Then we have

$$\mathbf{C} = \begin{pmatrix} 4 & -3 \\ -2 & 1 \end{pmatrix}, \quad \mathbf{C}' = \begin{pmatrix} 4 & -2 \\ -3 & 1 \end{pmatrix}, \quad |\mathbf{A}| = -2,$$

from which follows

$$\mathbf{A}^{-1} = \begin{pmatrix} -2 & 1 \\ 1.5 & -0.5 \end{pmatrix}, \quad |\mathbf{A}^{-1}| = \frac{1}{|\mathbf{A}|} = -0.5.$$

Example 7.23. Consider a matrix $\mathbf{A} = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 3 & 4 \\ 5 & 8 & 5 \end{pmatrix}$. Then we have

$$\mathbf{C} = \begin{pmatrix} -17 & 20 & -15 \\ 16 & -5 & -8 \\ -6 & -4 & 3 \end{pmatrix}, \quad \mathbf{C}' = \begin{pmatrix} -17 & 16 & -6 \\ 20 & -5 & -4 \\ -15 & -8 & 3 \end{pmatrix}, \quad |\mathbf{A}| = -47,$$

from which follows

$$\mathbf{A}^{-1} = -\frac{1}{47}\begin{pmatrix} -17 & 16 & -6 \\ 20 & -5 & -4 \\ -15 & -8 & 3 \end{pmatrix}.$$

Another way to find an inverse is through the *Gauss-Jordan method*: take an augmented matrix $(\mathbf{A} \mid \mathbf{I}_n)$, and via row operations, transform it into $(\mathbf{I}_n \mid \mathbf{B})$. The inverse is then $\mathbf{A}^{-1} = \mathbf{B}$.

Example 7.24. Consider a matrix $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$. Then we have

$$(\mathbf{A} \mid \mathbf{I}_n) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & -1 & 0 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right)$$
$$\sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right) \sim \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 1 & 1 \end{array}\right).$$

Previously, we saw that partitioning the matrices can make multiplication easier. The same is true for inversion. Consider a partitioned matrix

$$\mathbf{A} = \left(\begin{array}{c|c} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \hline \mathbf{A}_{21} & \mathbf{A}_{22} \end{array}\right).$$

Its inverse is

$$\mathbf{A}^{-1} = \left(\begin{array}{c|c} (\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1} & -(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21})^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \hline -(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & (\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12})^{-1} \end{array}\right).$$

Finally, it is straightforward to show that $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ for square invertible $\mathbf{A}$ and $\mathbf{B}$.

7.5 Orthogonality, projections

Definition 7.25. Consider two vectors $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{y} \in \mathbb{R}^n$. We say that $\mathbf{x}$ and $\mathbf{y}$ are *orthogonal* if $\mathbf{x}'\mathbf{y} = 0$. We write $\mathbf{x} \perp \mathbf{y}$.

Consider two matrices $\mathbf{A} \in \mathbb{R}^{n\times k}$ and $\mathbf{B} \in \mathbb{R}^{n\times p}$. We say that $\mathbf{A}$ and $\mathbf{B}$ are *orthogonal* if $\mathbf{A}'\mathbf{B} = \mathbf{0}$. We write $\mathbf{A} \perp \mathbf{B}$.

Example 7.26. Consider the following vectors:

$$\mathbf{x} = \begin{pmatrix} 1 \\ 0 \\ 3 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} 2 \\ 5 \\ 1 \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 3 \\ 0 \\ -1 \end{pmatrix}.$$

We have that the vector $\mathbf{z}$ is orthogonal to all remaining vectors, and $\mathbf{x} \perp \mathbf{v}$.

Definition 7.27. Consider a vector space spanned by a vector $\mathbf{a} \in \mathbb{R}^n$, and let $\mathbf{b} \in \mathbb{R}^n$. We call $\mathbf{p} = \alpha\mathbf{a}$ a *projection* of $\mathbf{b}$ onto the space spanned by $\mathbf{a}$.

Let a vector $\mathbf{e} = \mathbf{b} - \mathbf{p}$ that is orthogonal to $\mathbf{a}$ by construction, $\mathbf{a}'\mathbf{e} = 0$. Then it follows that $\mathbf{a}'(\alpha\mathbf{a} - \mathbf{b}) = 0$ and $\alpha = \dfrac{\mathbf{a}'\mathbf{b}}{\mathbf{a}'\mathbf{a}}$, then

$$\mathbf{p} = \mathbf{a}\alpha = \mathbf{a}\frac{\mathbf{a}'\mathbf{b}}{\mathbf{a}'\mathbf{a}} = \frac{\mathbf{a}\mathbf{a}'}{\mathbf{a}'\mathbf{a}}\mathbf{b}.$$

A generalization to a general vector space V of dimension k is given by

$$\mathbf{p} = \mathbf{P}\mathbf{b}, \quad \mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}',$$

where $\mathbf{P} \in \mathbb{R}^{n\times n}$ is a *projection matrix*, and the columns of $\mathbf{X} \in \mathbb{R}^{n\times k}$ are the basis of V.

Corollary 7.28. *Let $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ be a projection matrix with a matrix $\mathbf{X} \in \mathbb{R}^{n\times k}$ having a full column rank. It holds that*

(a) $\mathbf{P}' = \mathbf{P}$ *(symmetric),*

(b) $\mathbf{P}^2 = \mathbf{P}$ *(idempotent),*

(c) $\operatorname{rank}(\mathbf{P}) = k$,

(d) $\mathbf{P}\mathbf{X} = \mathbf{X}$,

(e) k eigenvalues of $\mathbf{P}$ are ones, and the remaining $n - k$ are zero.

Definition 7.29. A matrix $\mathbf{M} \in \mathbb{R}^{n\times n}$ defined as $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$ is called an *annihilator* matrix. By multiplying it by a vector $\mathbf{b}$ we obtain a vector of residuals $\mathbf{e}$, $\mathbf{M}\mathbf{b} = \mathbf{b} - \mathbf{p} = \mathbf{e}$.

Corollary 7.30. *Let $\mathbf{M} = \mathbf{I}_n - \mathbf{P}$ be an annihilator matrix with a matrix $\mathbf{X} \in \mathbb{R}^{n\times k}$ having a full column rank. It holds that*

(a) $\mathbf{M}' = \mathbf{M}$ *(symmetric),*

(b) $\mathbf{M}^2 = \mathbf{M}$ *(idempotent),*

(c) $\mathbf{M}\mathbf{X} = \mathbf{0}$,

(d) $n - k$ eigenvalues of $\mathbf{M}$ are ones, and the remaining k are zero.

7.6 Eigenvalues

Definition 7.31. Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a square matrix. $\mathbf{A}$ has n eigenvalues, $\lambda_1, \ldots, \lambda_n$, and corresponding eigenvectors $\mathbf{x}$, such that for each pair $(\lambda_k, \mathbf{x})$ it holds that $\mathbf{A}\mathbf{x} = \lambda_k\mathbf{x}$ for $k = 1, \ldots, n$.

Example 7.32. Let $\mathbf{P}$ be a projection matrix onto some vector space V. Let λ be its eigenvalues with a corresponding vector $\mathbf{x}$. Then, by Definition 7.31 it holds that $\mathbf{P}\mathbf{x} = \lambda\mathbf{x}$. The right-hand side is a multiple of $\mathbf{x}$ (which has the same direction as $\mathbf{x}$), and the left-hand side is a projection of $\mathbf{x}$ on V. Thus, for any eigenvector $\mathbf{x}$ of $\mathbf{P}$, $\mathbf{x}$ is either in V (with an associated $\lambda = 1$) or $\mathbf{x}$ is orthogonal to V (with an associated $\lambda = 0$):

$$\begin{aligned} \mathbf{P}\mathbf{x} &= \mathbf{x} \quad \text{for} \quad \forall \mathbf{x} \in V, \\ \mathbf{P}\mathbf{x} &= \mathbf{0} \quad \text{for} \quad \forall \mathbf{x} \perp V. \end{aligned}$$

Usually, eigenvalues of the matrix $\mathbf{A}$ are found as the solution to the *characteristic equation* $|\mathbf{A} - \lambda \mathbf{I}_n| = 0$. The left-hand side is called the *characteristic equation*.

Example 7.33. Consider a matrix $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix}$. Then

$$\mathbf{A} - \lambda \mathbf{I}_2 = \begin{pmatrix} 3-\lambda & 1 \\ 1 & 3-\lambda \end{pmatrix}, \quad |\mathbf{A} - \lambda \mathbf{I}_2| = (3-\lambda)^2 - 1 = 0, \quad \lambda_1 = 2, \quad \lambda_2 = 4.$$

Using the definition $(\mathbf{A} - \lambda \mathbf{I}_n)\mathbf{x} = \mathbf{0}$, for $\lambda_1 = 2$,

$$\begin{pmatrix} 3-2 & 1 \\ 1 & 3-2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

implying $x = \begin{pmatrix} -1 & 1 \end{pmatrix}'$. Similarly, for $\lambda_2 = 4$, we obtain $x = \begin{pmatrix} 1 & 1 \end{pmatrix}'$. Note that the eigenvectors are orthogonal.

Take a matrix $\mathbf{A} \in \mathbb{R}^{2\times 2}$ with its eigenvalues λ_1, λ_2. A useful property (that holds, however, for larger matrices as well) is

$$\begin{aligned} \lambda_1 + \lambda_2 &= \text{trace}(\mathbf{A}), \\ \lambda_1 \cdot \lambda_2 &= |\mathbf{A}|. \end{aligned}$$

Example 7.34. Consider a matrix $\mathbf{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Then we have

$$\mathbf{A} - \lambda \mathbf{I}_2 = \begin{pmatrix} 0-\lambda & -1 \\ 1 & 0-\lambda \end{pmatrix}, \quad |\mathbf{A} - \lambda \mathbf{I}_2| = \lambda^2 + 1 = 0, \quad \lambda_1 = i^2, \quad \lambda_2 = -i^2.$$

Example 7.35. Consider a matrix $\mathbf{A} = \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}$.

$$\mathbf{A} - \lambda \mathbf{I}_2 = \begin{pmatrix} 3-\lambda & 1 \\ 0 & 3-\lambda \end{pmatrix}, \quad |\mathbf{A} - \lambda \mathbf{I}_2| = (3-\lambda)^2 = 0, \quad \lambda_1 = \lambda_2 = 3.$$

The (repeated) eigenvalue is 3, and there is only one independent eigenvector.

Theorem 7.36. *Suppose a matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ has n independent eigenvectors, let $\mathbf{S} \in \mathbb{R}^{n\times n}$ be a matrix of the corresponding eigenvectors, and let $\mathbf{\Lambda} \in \mathbb{R}^{n\times n}$ be a diagonal matrix with eigenvalues on the diagonal. Then*

$$\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}.$$

From the Theorem 7.36, we can easily obtain powers of $\mathbf{A}$,

$$\begin{aligned} \mathbf{A}^2 &= \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}\mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1} = \mathbf{S}\mathbf{\Lambda}^2\mathbf{S}^{-1}, \\ \mathbf{A}^k &= \mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}. \end{aligned}$$

It follows that $\mathbf{A}^k \to 0$ as $k \to \infty$ if $|\lambda_i| < 1$ for all i.

Definition 7.37. A matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is *diagonalizable* if it has n independent eigenvectors. If all eigenvalues of $\mathbf{A}$ are different, then $\mathbf{A}$ is surely diagonalizable; if there are repeated eigenvalues, $\mathbf{A}$ may or may not have n independent eigenvectors.

Corollary 7.38. *Consider a square matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ with λ_i and $\mathbf{x}_i$ its eigenvalues and eigenvectors respectively $(i = 1, \ldots, n)$. Then*

(a) $\sum_{i=1}^{n} \lambda_i = \text{trace}(\mathbf{A})$.

(b) $\prod_{i=1}^{n} \lambda_i = |\mathbf{A}|$.

(c) $\mathbf{A}$ *is non-singular if and only if all its eigenvalues are non-zero.*

(d) *Non-zero eigenvalues of* $\mathbf{AB}$ *and* $\mathbf{BA}$ *are identical.*

(e) *If* $\mathbf{B}$ *is non-singular, then* $\mathbf{A}$ *and* $\mathbf{B}^{-1}\mathbf{AB}$ *have the same eigenvalues.*

(f) *If* $\mathbf{A}\mathbf{x}_i = \lambda_i \mathbf{x}_i$*, then* $(\mathbf{I}_n - \mathbf{A})\mathbf{x}_i = (1 - \lambda_i)\mathbf{x}_i$*, so that* $\mathbf{I}_n - \mathbf{A}$ *has the eigenvalue* $1 - \lambda_i$ *with an associated vector* $\mathbf{x}_i$*.*

(g) *Eigenvalues of* $\mathbf{A}$ *coincide with eigenvalues of* $\mathbf{A}'$*.*

An important class of matrices is a class of real symmetric matrices. If $\mathbf{A}$ is symmetric, it is diagonalizable, its eigenvalues are real, and eigenvectors are orthogonal.