MAT4120

Exercises for Mathematical Optimization

August Femtehjell

Autumn 2025

Contents

Abstract

This document contains my solutions to the exercises for the course MAT4120–Mathematical Optimization, taught at the University of Oslo in the spring of 2025. The code for everything, as well as this document, can be found at my GitHub repository: https://github.com/augustfe/MAT4120.

1 The basic concepts

Exercise 1.1. Let $x_1, x_2, y_1, y_2 \in \mathbb{R}^n$ and assume that $x_1 \leq x_2$ and $y_1 \leq y_2$. Verify that the inequality $x_1 + y_1 \leq x_2 + y_2$ holds. Let now λ be a non-negative real number. Explain why $\lambda x_1 \leq \lambda x_2$ holds. What happens if λ is negative?

Solution 1.1. With $x_1 \leq x_2$, we have that

$$(x_1)_i \leq (x_2)_i \qquad \forall i = 1, \ldots, n. \tag{1.1}$$

Component-wise, we then have

$$(x_1)_i + (y_1)_i \leq (x_2)_i + (y_2)_i \qquad \forall i = 1, \ldots, n, \tag{1.2}$$

and thus $x_1 + y_1 \leq x_2 + y_2$. Similarly, if $\lambda \geq 0$, we have

$$\lambda(x_1)_i \leq \lambda(x_2)_i \qquad \forall i = 1, \ldots, n, \tag{1.3}$$

and therefore $\lambda x_1 \leq \lambda x_2$. Finally, for $\lambda < 0$, the inequality reverses:

$$\lambda(x_1)_i \geq \lambda(x_2)_i \qquad \forall i = 1, \ldots, n, \tag{1.4}$$

giving $\lambda x_1 \geq \lambda x_2$.

Example 1.2.1 (*The non-negative real vectors*) The sum of two non-negative numbers is again a non-negative number. Similarly, we see that the sum of two non-negative vectors is a non-negative vector. Moreover, if we multiply a non-negative vector by a non-negative number, we get another non-negative vector. These two properties may be summarized by saying that $\mathbb{R}^n_+$ is closed under addition and multiplication by non-negative scalars. We shall see that this means that $\mathbb{R}^n_+$ is a convex cone, a special type of convex set.

Exercise 1.2. Think about the question in Exercise 1.1 again, now in light of the properties explained in Example 1.2.1.

Solution 1.2. We can now rewrite $x_1 \leq x_2$ as $x_2 - x_1 \in \mathbb{R}^n_+$. We can now easily consider the first question as

$$(x_2 + y_2) - (x_1 + y_1) = (x_2 - x_1) + (y_2 - y_1) \in \mathbb{R}^n_+, \tag{1.5}$$

as $\mathbb{R}^n_+$ is closed under addition. Similarly, we can use the fact that $\mathbb{R}^n_+$ is closed under multiplication by non-negative scalars to see that

$$(\lambda x_2 - \lambda x_1) = \lambda(x_2 - x_1) \in \mathbb{R}^n_+, \tag{1.6}$$

for $\lambda \geq 0$. As $\mathbb{R}^n_+$ is not closed under multiplication by negative scalars, we cannot conclude that $\lambda x_1 \leq \lambda x_2$ for $\lambda < 0$.

Exercise 1.3. Let $a \in \mathbb{R}^n_+$ and assume that $x \leq y$. Show that $a^T x \leq a^T y$. What happens if we do not require a to be non-negative here?

Solution 1.3. With $a \in \mathbb{R}^n_+$ and $x \leq y$, we have that

$$x_i \leq y_i \qquad \forall i = 1, \ldots, n, \tag{1.7}$$

and consequently

$$a_i x_i \leq a_i y_i \qquad \forall i = 1, \ldots, n, \tag{1.8}$$

as shown previously. Written in vector notation, we therefore have

$$a^T x \leq a^T y. \tag{1.9}$$

With a not necessarily non-negative, we may have neither $a^T x \geq a^T y$ nor $a^T x \leq a^T y$, as we could have $a_i x_i > a_i y_i$ for some i.

Exercise 1.4. Show that every ball $B(a, r) := \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$ is convex (where $a \in \mathbb{R}^n$ and $r \geq 0$).

Solution 1.4. Let $x, y \in B(a, r)$ for some $a \in \mathbb{R}^n$ and $r \geq 0$. Then, let $0 \leq \lambda \leq 1$ and consider $z = \lambda x + (1 - \lambda) y$. We then have

$$\begin{aligned} \|z - a\| &= \|\lambda(x - a) + (1 - \lambda)(y - a)\| \\ &\leq \lambda \|x - a\| + (1 - \lambda)\|y - a\| \\ &\leq \lambda r + (1 - \lambda) r = r, \end{aligned} \tag{1.10}$$

showing that $z \in B(a, r)$. $B(a, r)$ is therefore convex.

Exercise 1.5. Explain how you can write the LP problem $\max\{c^T x : Ax \leq b\}$ in the form $\max\{c^T x : Ax = b, x \geq O\}$

Solution 1.5. We introduce new slack variables $w \in \mathbb{R}^m_+$, where m is the number of rows/inequalities in A, defined by

$$w_j = b_j - (Ax)_j \quad \forall j = 1, \ldots, m. \tag{1.11}$$

We can then rewrite our system of equations by setting $\tilde{A} = \begin{bmatrix} A & I \end{bmatrix}$, $\tilde{x} = \begin{bmatrix} x \\ w \end{bmatrix}$, and $\tilde{c} = \begin{bmatrix} c \\ 0 \end{bmatrix}$. We then have

$$\tilde{A}\tilde{x} = \begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} x \\ w \end{bmatrix} = Ax + w = b, \tag{1.12}$$

and

$$\tilde{c}^T \tilde{x} = \begin{bmatrix} c \\ 0 \end{bmatrix}^T \begin{bmatrix} x \\ w \end{bmatrix} = c^T x. \tag{1.13}$$

Again, as we require $w \geq 0$, we then have $Ax \leq b$.

Exercise 1.6. Make a drawing of the standard simplices S_1, S_2, and S_3. Verify that each unit vector e_j lies in S_n (e_j has a one in position j, all other components are zero). Each $x \in S_n$ may be written as a linear combination $x = \sum_{j=1}^n \lambda_j e_j$ where each λ_j is non-negative and $\sum_{j=1}^n \lambda_j = 1$. How? Can this be done in several ways?

Solution 1.6. Fig. 1 shows the standard simplices S_1, S_2, and S_3. Clearly each unit vector e_j lies in S_n. Each $x \in S_n$ may be written as $\sum_{j=1}^n \lambda_j e_j$ where $\lambda_j = x_j$, i.e. the coordinate components of x.

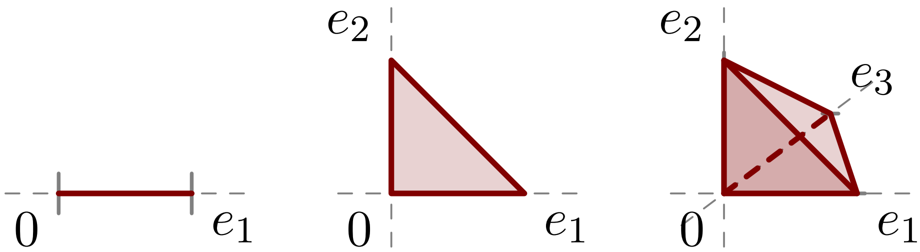

Figure 1: The simplices S_1, S_2, and S_3. For S_1, the standard simplex is the point e_1, for S_2, the standard simplex is the line segment between e_1 and e_2, and for S_3, the standard simplex is the triangle with vertices e_1, e_2, and e_3.

Exercise 1.7. Show that each convex cone is indeed a convex set.

Solution 1.7. To see that a convex cone is a convex set, let first $x_1, x_2 \in C$. Then let $0 \leq \lambda_1 \leq 1$ and $\lambda_2 = 1 - \lambda_1 \geq 0$. We then have by definition of the convex cone that

$$\lambda_1 x_1 + (1 - \lambda_1)x_1 = \lambda_1 x_1 + \lambda_2 x_2 \in C, \tag{1.14}$$

showing that the set is convex.

Exercise 1.8. Let $A \in \mathbb{R}^{m \times n}$ and consider the set $C = \{x \in \mathbb{R}^n : Ax \leq O\}$. Prove that C is a convex cone.

Solution 1.8. Let $x_1, x_2 \in C$ and $\lambda_1, \lambda_2 \in \mathbb{R}_+$. We then have

$$A(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 A x_1 + \lambda_2 A x_2 \leq \lambda_1 O + \lambda_2 O = O, \tag{1.15}$$

showing that the set is a convex cone.

Polyhedral cone A convex cone of the form $\{x \in \mathbb{R}^n : Ax \leq O\}$ where $A \in \mathbb{R}^{m\times n}$ is called a *polyhedral cone*. Let $x_1, \dots, x_t \in \mathbb{R}^n$ and let $C(x_1, \dots, x_t)$ be the set of vectors of the form

$$u = \sum_{j=1}^{t} \lambda_j x_j, \tag{1.16}$$

where $\lambda_j \geq 0$ for each $j = 1, \dots, t$.

Exercise 1.9. Prove that $C(x_1, \dots, x_t)$ is a convex cone.

Solution 1.9. Let $C = C(x_1, \dots, x_t)$ here for convenience. Let $u, v \in C$ with respective coefficients $\lambda_j, \mu_j \geq 0$ for $j = 1, \dots, t$. Then, for arbitrary coefficients $\alpha, \beta \geq 0$, we have

$$\begin{aligned} A(\alpha u + \beta v) &= A\left(\alpha \sum_{j=1}^{t} \lambda_j x_j + \beta \sum_{j=1}^{t} \mu_j x_j\right) \\ &= \alpha A \sum_{j=1}^{t} \lambda_j x_j + \beta A \sum_{j=1}^{t} \mu_j x_j \\ &\leq \alpha O + \beta O = O, \end{aligned} \tag{1.17}$$

showing that $\alpha u + \beta v \in C$, and that C is a convex cone.

Exercise 1.10. Let $S = \{(x, y, z) : z \geq x^2 + y^2\} \subset \mathbb{R}^3$. Sketch the set and verify that it is a convex set. Is S a finitely generated cone?

Solution 1.10. Let $u = (x_1, y_1, z_1)$ and $v = (x_2, y_2, z_2)$ be two points in S, and $0 \leq \lambda \leq 1$. We then seek to show that $\lambda u + (1-\lambda)v \in S$. Note that $f(x) = x^2$ is convex, i.e. that $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$.

Considering the right hand side of the inequality first, we have component-wise that

$$\begin{aligned} (\lambda x_1 + (1-\lambda)x_2)^2 &\leq \lambda x_1^2 + (1-\lambda)x_2^2 \\ (\lambda y_1 + (1-\lambda)y_2)^2 &\leq \lambda y_1^2 + (1-\lambda)y_2^2, \end{aligned} \tag{1.18}$$

and therefore

$$\begin{aligned} (\lambda x_1 + (1-\lambda)x_2)^2 + (\lambda y_1 + (1-\lambda)y_2)^2 &\leq \lambda(x_1^2 + y_1^2) + (1-\lambda)(x_2^2 + y_2^2) \\ &\leq \lambda z_1 + (1-\lambda)z_2, \end{aligned} \tag{1.19}$$

and so $\lambda u + (1-\lambda)v \in S$, showing that S is convex.

S is not a finitely generated cone, as it should then be closed under scaling. Consider e.g. $u = (1, 0, 1)$, which satisfies $1 \geq 1^2 + 0^2$ and so $u \in S$. Letting however $\lambda_1 = 2$ (and $\lambda_2 = 0$), we have $2u = (2, 0, 2) \notin S$ as $2 \ngeq 2^2 + 0^2$.

Exercise 1.11. Consider the linear system $0 \leq x_i \leq 1$ for $i = 1, \ldots, n$, and let P denote the solution set. Explain how to solve a linear programming problem

$$\max\{c^T x : x \in P\}. \tag{1.20}$$

What if the linear system was $a_i \leq x_i \leq b_i$ for $i = 1, \ldots, n$? Here we assume $a_i \leq b_i$ for each i.

Solution 1.11. We choose the solution component-wise, by maximizing each component $c_i x_i$. If $c_i > 0$, simply let $x_i = 1$. If $c_i < 0$, increasing x_i decreases the objective function, so we let $x_i = 0$. The same argument holds in the alternate case, just choose $x_i = b_i$ or $x_i = a_i$ respectively based on the sign of c_i.

Exercise 1.12. Is the union of two convex sets again convex?

Solution 1.12. No. Let $A = [-2, -1]$ and $B = [1, 2]$. Then $A \cup B = [-2, -1] \cup [1, 2]$ is not convex, as e.g. $\frac{1}{2}1 + (1 - \frac{1}{2})(-1) = 0 \notin A \cup B$.

Exercise 1.13. Determine the sum $A + B$ in each of the following cases:

$$\begin{array}{rll} (i) & A = \{(x, y) : x^2 + y^2 \leq 1\}, & B = \{(3, 4)\}; \\ (ii) & A = \{(x, y) : x^2 + y^2 \leq 1\}, & B = [0, 1] \times \{0\}; \\ (iii) & A = \{(x, y) : x + 2y = 5\}, & B = \{(x, y) : x = y, 0 \leq x \leq 1\}; \\ (iv) & A = [0, 1] \times [1, 2], & B = [0, 2] \times [0, 2]. \end{array}$$

Solution 1.13. Taking the cases in turn:

(i) The sum $A + B$ is given by the set of points $\{u + (3, 4) : u \in A\}$, that is, the unit disk centred around $(3, 4)$.

(ii) The sum $A + B$ is given by those points that are either in the rectangle with corners at $(0, -1)$, $(1, -1)$, $(1, 1)$ and $(0, 1)$, or in the unit disks centred about $(0, 0)$ or $(1, 0)$.

(iii) Let $B = \{(\lambda, \lambda) : 0 \leq \lambda \leq 1\}$. Then we can write a point $(x, y) \in A + B$ as

$$(x, y) = (x_0 + \lambda, y_0 + \lambda). \tag{1.21}$$

Then

$$x + 2y = (x_0 + 2y_0) + 3\lambda = 5 + 3\lambda, \tag{1.22}$$

which gives us that

$$A + B = \{(x, y) : 5 \leq x + 2y \leq 8\}. \tag{1.23}$$

(iv) The sum $A + B$ is simply given by $[0, 3] \times [1, 4]$.

Exercise 1.14. More enumerated exercises...

(*i*) Prove that, for every $\lambda \in \mathbb{R}$ and $A, B \subseteq \mathbb{R}^n$, it holds that $\lambda(A+B) = \lambda A + \lambda B$.

(*ii*) Is it true that $(\lambda + \mu)A = \lambda A + \mu A$ for every $\lambda, \mu \in \mathbb{R}$ and $A \subseteq \mathbb{R}^n$? If not, find a counterexample.

(*iii*) Show that, if $\lambda, \mu \geq 0$ and $A \subseteq \mathbb{R}^n$ is convex, then $(\lambda + \mu)A = \lambda A + \mu A$.

Solution 1.14. Taking the exercises in turn again...
(*i*) We have that

$$\begin{aligned}
\lambda(A+B) &= \{\lambda(a+b) : a \in A, b \in B\} \\
&= \{\lambda a + \lambda b : a \in A, b \in B\} \\
&= \{a + b : a \in \lambda A, b \in \lambda B\} \\
&= \lambda A + \lambda B.
\end{aligned}$$

(*ii*) No, it is not true. Consider $A = [1, 2]$, $\lambda = 1$ and $\mu = -1$. Then we have

$$(\lambda + \mu)A = 0A = \{0\} \quad \text{and} \quad \lambda A + \mu A = [1, 2] + [-2, -1] = [-1, 1]. \tag{1.24}$$

(*iii*) Let $\lambda, \mu \geq 0$. For any $a \in A$, we have that

$$(\lambda + \mu)a = \lambda a + \mu a \in \lambda A + \mu A, \tag{1.25}$$

so $(\lambda + \mu)A \subseteq \lambda A + \mu A$. For the reverse inclusion, let $u \in \lambda A + \mu A$. Then $u = \lambda a + \mu b$ for some $a, b \in A$. Scaling the factors, we have that

$$u = (\lambda + \mu)\left(\frac{\lambda}{\lambda + \mu}a + \frac{\mu}{\lambda + \mu}b\right). \tag{1.26}$$

As both the inner coefficients are non-negative and sum to one, and $a, b \in A$, we have that

$$\frac{\lambda}{\lambda + \mu}a + \frac{\mu}{\lambda + \mu}b \in A, \tag{1.27}$$

and $u \in (\lambda + \mu)A$. Therefore, $(\lambda + \mu)A = \lambda A + \mu A$ with the given assumptions.

Exercise 1.15. Show that if $C_1, \ldots, C_t \subseteq \mathbb{R}^n$ are all convex sets, then $C_1 \cap \cdots \cap C_t$ is convex. Do the same when all sets are affine (or linear subspaces, or convex cones). In fact, a similar result for the intersection of any family of convex sets. Explain this.

Solution 1.15. Let $x, y \in C_1 \cap \cdots \cap C_t$ and $0 \leq \lambda \leq 1$. Then $x, y \in C_i$ for all $i = 1, \ldots, t$. Since each C_i is convex, we have that

$$\lambda x + (1 - \lambda) y \in C_i \tag{1.28}$$

for all $i = 1, \ldots, t$. Thus,

$$\lambda x + (1 - \lambda) y \in C_1 \cap \cdots \cap C_t, \tag{1.29}$$

proving that the intersection is convex.

Suppose there is a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$. Then the set

$$C = \{x \in \mathbb{R}^n : Ax = b\} \tag{1.30}$$

is an affine set. Then let $x \in \bigcap_{i=1}^t C_i$, meaning that $A_i x = b_i$ for all $i = 1, \ldots, t$. Thus, x satisfies

$$(A_1 + A_2 + \cdots + A_t) x = b_1 + b_2 + \cdots + b_t, \tag{1.31}$$

and $\bigcap_{i=1}^t C_i$ is itself affine.

A similar argument shows that the closure property of convex sets is preserved under finite intersections.

Exercise 1.16. Consider a family (possibly infinite) of linear inequalities $a_i^T x \leq b, i \in I$, and C be its solution set, i.e., C is the set of points satisfying all the inequalities. Prove that C is a convex set.

Solution 1.16. Let $x, y \in C$, and $0 \leq \lambda \leq 1$. Then, for each $i \in I$, we have

$$a_i^T (\lambda x + (1 - \lambda) y) = \lambda a_i^T x + (1 - \lambda) a_i^T y \leq \lambda b_i + (1 - \lambda) b_i = b_i. \tag{1.32}$$

Therefore, $\lambda x + (1 - \lambda) y \in C$, and C is convex.

Exercise 1.17. Consider the unit disc $S = \{(x_1, x_2) \in \mathbb{R}^2 : x_1^2 + x_2^2 \leq 1\}$ in $\mathbb{R}^2$. Find a family of linear inequalities as in the previous problem with solution set S.

Solution 1.17. Let $U = \{(\cos\theta, \sin\theta) : 0 \leq \theta < 2\pi\}$ be the unit circle in $\mathbb{R}^2$. Then, for each $(u, v) \in U$, we have the linear inequalities

$$u x_1 + v x_2 \leq 1. \tag{1.33}$$

This feels like a circular argument, but alright.

Exercise 1.18. Is the unit ball $B = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ a polyhedron?

Solution 1.18. I don't believe the unit ball is a polyhedron. A polyhedron requires a *finite* number of linear inequalities, i.e. constraints which can be written as $Ax \leq b$. The unit ball however is inherently smooth, without edges, and therefore requires an infinite number of linear constraints, as those applied in the previous exercise.

Exercise 1.19. Show that the unit ball $B_\infty = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$ is convex. Here $\|x\|_\infty = \max_j |x_j|$ is the max norm of x. Show that B_∞ is a polyhedron. Illustrate when $n = 2$.

Solution 1.19. Let $x, y \in B_\infty$ and $0 \leq \lambda \leq 1$. We then have

$$\|\lambda x + (1-\lambda)y\|_\infty \leq \lambda\|x\|_\infty + (1-\lambda)\|y\|_\infty \leq \lambda + (1-\lambda) = 1, \tag{1.34}$$

so $\lambda x + (1-\lambda)y \in B_\infty$, and B_∞ is convex. B_∞ is a polyhedron, as it is described by the constraints

$$x_j \leq 1 \quad \text{and} \quad -x_j \leq 1 \qquad j = 1, \dots, n. \tag{1.35}$$

When $n = 2$, B_∞ is the unit square, with corners $(\pm 1, \pm 1)$ and $(\pm 1, \mp 1)$, as seen in Fig. 2.

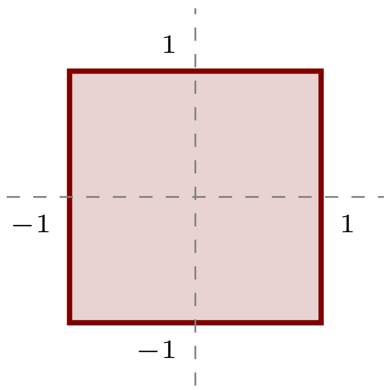

Figure 2: B_∞ when $n = 2$.

Exercise 1.20. Show that the unit ball $B_1 = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$ is convex. Here $\|x\|_1 = \sum_{j=1}^n |x_j|$ is the absolute norm of x. Show that B_1 is a polyhedron. Illustrate when $n = 2$.

Solution 1.20. Let $x, y \in B_1$ and $0 \le \lambda \le 1$. Then we have

$$\begin{aligned} \|\lambda x + (1-\lambda)y\|_1 &= \sum_{j=1}^{n} |\lambda x_j + (1-\lambda)y_j| \\ &\le \sum_{j=1}^{n} \lambda|x_j| + (1-\lambda)|y_j| = \lambda \sum_{j=1}^{n} |x_j| + (1-\lambda) \sum_{j=1}^{n} |y_j| \\ &\le \lambda + (1-\lambda) = 1. \end{aligned} \tag{1.36}$$

Therefore, $\lambda x + (1-\lambda)y \in B_1$, and B_1 is convex.

B_1 is a polyhedron, as it is described by the constraints

$$\sum_{j=1}^{n} \sigma_j x_j \le 1, \quad \forall(\sigma_1, \dots, \sigma_n) \in \{-1, 1\}^n. \tag{1.37}$$

When $n = 2$, B_1 is the unit diamond, with corners $(\pm 1, 0)$ and $(0, \pm 1)$, as seen in Fig. 3.

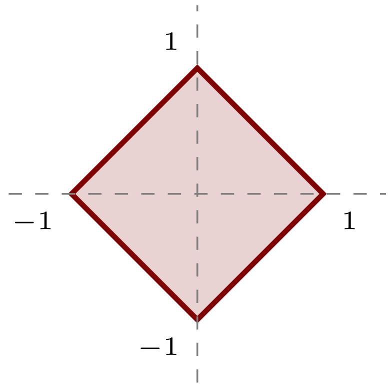

Figure 3: B_1 when $n = 2$.

Proposition 1.5.1 (*Affine sets*). Let C be a non-empty subset of $\mathbb{R}^n$. Then C is an affine set if an only if there is a matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$ for some m such that

$$C = \{x \in \mathbb{R}^n : Ax = b\}. \tag{1.38}$$

Moreover, C may be written as $C = L + x_0 = \{x + x_0 : x \in L\}$ for some linear subspace L of $\mathbb{R}^n$. The subspace L is unique.

Exercise 1.21. Prove Proposition 1.5.1.

Solution 1.21. Let $x_0 \in C$, where C is affine, and $L = \{x - x_0 : x \in C\}$. For an arbitrary $x \in C$, we have that

$$\lambda(x - x_0) = \lambda x + (1 - \lambda)x_0 - x_0 \in L. \tag{1.39}$$

Finish later, this is getting long.

Exercise 1.22. Let C be a non-empty affine set in $\mathbb{R}^n$. Define $L = C - C$. Show that L is a subspace and that $C = L + x_0$ for some vector x_0.

Solution 1.22. We have that

$$L = C - C = \{x - y : x, y \in C\}. \tag{1.40}$$

We have from the previous exercise that $C = L + x_0$ for a subspace L. Letting $x, y \in C$, we can express these as

$$x = \alpha + x_0 \quad \text{and} \quad y = \beta + x_0 \tag{1.41}$$

for some $\alpha, \beta \in L$. Then $x - y = \alpha - \beta \in L$, showing that $C - C \subseteq L$. For the reverse inclusion, let $\alpha \in L$. Then $\alpha = x - x_0$ for some $x \in C$, showing that $L \subseteq C - C$, proving that $L = C - C$.

2 Convex hulls and Carathéodory's theorem

Exercise 2.1. Illustrate some combinations (linear, convex, non-negative) of two vectors in $\mathbb{R}^2$.

Solution 2.1. Let $\mathbf{u} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$.

- Linear combinations: $\alpha\mathbf{u} + \beta\mathbf{v}$ for $\alpha, \beta \in \mathbb{R}$.
- Convex combinations: $\alpha\mathbf{u} + \beta\mathbf{v}$ for $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$.
- Non-negative combinations: $\alpha\mathbf{u} + \beta\mathbf{v}$ for $\alpha, \beta \geq 0$.

The linear combinations fill the entire $\mathbb{R}^2$ plane, the convex combinations fill the line segment between $\mathbf{u}$ and $\mathbf{v}$, and the non-negative combinations fill the first quadrant of the plane.

Exercise 2.2. Choose your favourite three points x_1, x_2, x_3 in $\mathbb{R}^2$, but make sure that they do not all lie on the same line. Thus, the three points form the corners of a triangle C. Describe those points that are convex combinations of two of the three points. What about the interior of the triangle C, i.e., those points that lie in C but not on the boundary (the three sides): can these points be written as convex combinations of x_1, x_2 and x_3? If so, how?

Solution 2.2. Let's choose the points $x_1 = (0, 0)$, $x_2 = (1, 0)$, and $x_3 = (0, 1)$, i.e., the corners of a right triangle. The points that are convex combinations of two of the three points are those that lie on the edges of the triangle. The interior points, which do not lie on the boundary, can be expressed as

$$c = \lambda_1 x_1 + \lambda_2 x_2 + \lambda_3 x_3, \tag{2.1}$$

with $0 < \lambda_i < 1$ for $i = 1, 2, 3$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$.

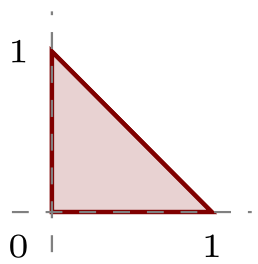

Figure 4: The unit triangle with points x_1, x_2, and x_3.

Exercise 2.3. Show that conv(S) is convex for all $S \subseteq \mathbb{R}^n$. (Hint: look at two convex combinations $\sum_j \lambda_j x_j$ and $\sum_j \mu_j y_j$, and note that both these points may be written as a convex combination of the same set of vectors.)

Solution 2.3. Let $u, v \in \text{conv}(S)$. Then there exists points $x_1, x_2, \ldots, x_n \in S$ and $y_1, y_2, \ldots, y_m \in S$ and coefficients $\lambda_i, \mu_j \geq 0$ with $\sum_i \lambda_i = 1$ and $\sum_j \mu_j = 1$ such that

$$u = \sum_{i=1}^{n} \lambda_i x_i, \quad v = \sum_{j=1}^{m} \mu_j y_j. \tag{2.2}$$

For any $0 \leq \theta \leq 1$, we have that

$$\begin{aligned} \theta u + (1-\theta)v &= \theta \sum_{i=1}^{n} \lambda_i x_i + (1-\theta) \sum_{j=1}^{m} \mu_j y_j \\ &= \sum_{i=1}^{n} (\theta \lambda_i) x_i + \sum_{j=1}^{m} ((1-\theta)\mu_j) y_j. \end{aligned} \tag{2.3}$$

The new coefficients are non-negative, and furthermore we have

$$\sum_{i=1}^{n} (\theta \lambda_i) + \sum_{j=1}^{m} ((1-\theta)\mu_j) = \theta \cdot 1 + (1-\theta) \cdot 1 = 1, \tag{2.4}$$

showing that $\theta u + (1-\theta)v$ is a convex combination of points in S, and therefore lies in conv(S). conv(S) is thus convex.

Exercise 2.4. Give an example of two distinct sets S and T having the same convex hull. It makes sense to look for a smallest possible subset S_0 of a set S such that $S = \text{conv}(S_0)$. We study this question later.

Solution 2.4. Let $S = \{(0,0), (1,0), (0,1)\}$ and $T = \{(0,0), (1,0), (0,1), (0.5, 0.5)\}$. Both sets have the same convex hull, which is the triangle formed by the points $(0,0)$, $(1,0)$, and $(0,1)$. The point $(0.5, 0.5)$ in set T lies within this triangle and does not change the convex hull.

Exercise 2.5. Prove that if $S \subseteq T$, then conv$(S) \subseteq$ conv(T).

Solution 2.5. Let $u \in \text{conv}(S)$. Then there exist points $x_1, x_2, \ldots, x_n \in S$ and coefficients $\lambda_i \geq 0$ with $\sum_i \lambda_i = 1$ such that

$$u = \sum_{i=1}^{n} \lambda_i x_i. \tag{2.5}$$

Since $S \subseteq T$, we have that $x_i \in T$ for all i. Therefore, u is also a convex combination of points in T, and thus lies in conv(T). Hence, conv$(S) \subseteq$ conv(T).

Exercise 2.6. If S is convex, then $\operatorname{conv}(S) = S$. Show this!

Solution 2.6. Since S is convex, for any $u, v \in S$ and any $0 \leq \lambda \leq 1$, we have $\lambda u + (1-\lambda)v \in S$. By the definition of convex hull, $\operatorname{conv}(S)$ is the smallest convex set containing S. Since S is already convex and contains itself, it follows that $\operatorname{conv}(S) = S$.

Exercise 2.7. Let $S = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$, this is the unit circle in $\mathbb{R}^2$. Determine $\operatorname{conv}(S)$ and $\operatorname{cone}(S)$.

Solution 2.7. The convex hull $\operatorname{conv}(S)$ of the unit circle is the unit disk, i.e., the set $\{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$. This is because any point inside the unit circle can be expressed as a convex combination of points on the unit circle.

The conical hull $\operatorname{cone}(S)$ of the unit circle is the entire $\mathbb{R}^2$ plane. This is because any point in $\mathbb{R}^2$ can be expressed as a non-negative combination of points on the unit circle, scaled appropriately. This can be seen simply by looking at the problem in polar coordinates.

Exercise 2.8. Does affine independence imply linear independence? Does linear independence imply affine independence? Prove or disprove!

Solution 2.8. Affine independence does not imply linear independence. Consider a set of linearly independent vectors $x_2, \ldots, x_t \in \mathbb{R}^n$ and let $x_1 = 0$. Then clearly the set $\{x_1, x_2, \ldots, x_t\}$ is affinely independent, but not linearly independent since x_1 is the zero vector.

Linear independence does imply affine independence. Suppose $\{x_1, x_2, \ldots, x_t\}$ is linearly independent. To show affine independence, we need to show that the only solution to

$$\sum_{i=1}^{t} \lambda_i x_i = 0 \quad \text{and} \quad \sum_{i=1}^{t} \lambda_i = 0 \tag{2.6}$$

is $\lambda_i = 0$ for all i. Since the vectors are linearly independent, the first equation implies that all λ_i must be zero. Thus, the set is affinely independent.

Exercise 2.9. Let $x_1, \ldots, x_t \in \mathbb{R}^n$ be affinely independent and let $w \in \mathbb{R}^n$. Show that $x_1 + w, \ldots, x_t + w$ are also affinely independent.

Solution 2.9. As $x_1, \ldots, x_t$ are affinely independent, we have that

$$\{x_2 - x_1, x_3 - x_1, \ldots, x_t - x_1\} \tag{2.7}$$

is linearly independent. We then have that

$$\{(x_2+w)-(x_1+w), \ldots, (x_t+w)-(x_1+w)\} = \{x_2-x_1, x_3-x_1, \ldots, x_t-x_1\}, \tag{2.8}$$

which is linearly independent. Thus, $x_1 + w, \ldots, x_t + w$ are affinely independent.

Dimension of a set. The *dimension* of a set $S \subseteq \mathbb{R}^n$, denoted by $\dim(S)$, is the maximal number of affinely independent points in S minus 1. So, for example in $\mathbb{R}^3$, the dimension of a point and a lines is 0 and 1 respectively, and the dimension of the plane $x_3 = 0$ is 2.

Exercise 2.10. Let L be a linear subspace of dimension (in the usual linear algebra sense) t. Check this this coincides with our new definition of dimension above. (Hint: add O to a "suitable" set of vectors).

Solution 2.10. Let $\{v_1, v_2, \ldots, v_t\}$ be a basis for the linear subspace L. Then the set $\{0, v_1, v_2, \ldots, v_t\}$ contains $t+1$ affinely independent points in L. To see this, note that the vectors $v_1, v_2, \ldots, v_t$ are linearly independent by definition of a basis. Therefore, the maximal number of affinely independent points in L is $t+1$, and thus $\dim(L) = t+1-1 = t$, which coincides with the usual definition of dimension in linear algebra.

Exercise 2.11. Consider a convex set C of dimension d. Then there are (and no more than) $d+1$ affinely independent points in C. Let $S = \{x_1, \ldots, x_{d+1}\}$ denote a set of such points. Then the set of all convex combinations of these vectors, i.e., $\mathrm{conv}(S)$ is a polytope contained in C and $\dim(S) = \dim(C)$. Moreover, let A be the set of all vectors of the form $\sum_{j=1}^t \lambda_j x_j$ where $\sum_{j=1}^t \lambda_j = 1$ (no sign restrictions of the λ's). Then A is an affine set containing C, and it is the smallest affine set with this property. A is called the *affine hull* of C.

Prove the last statements in the previous paragraph.

Solution 2.11. Consider two points $u, v \in A$ where

$$u = \sum_{j=1}^{t} \lambda_j x_j \quad \text{and} \quad v = \sum_{j=1}^{t} \mu_j x_j, \tag{2.9}$$

where $\sum_{j=1}^t \lambda_j = 1$ and $\sum_{j=1}^t \mu_j = 1$. Then, we have that

$$(1-\theta)u + \theta v = \sum_{j=1}^{t} ((1-\theta)\lambda_j + \theta \mu_j) x_j \in A \tag{2.10}$$

for any $\theta \in \mathbb{R}$ as the coefficients still sum to 1. A is therefore affine. Choosing $\lambda_j = \delta_{ij}$ (the Kronecker delta) shows that $x_i \in A$ for all i, and thus $\mathrm{conv}(C) = C \subseteq A$.

Exercise 2.12. Construct a set which is neither open nor closed.

Solution 2.12. Consider the interval $S = (0, 1]$ in $\mathbb{R}$. This set is not open because it contains the point 1, which is a limit point of the set. It is not closed because it does not contain the point 0, which is also a limit point of the set. Therefore, S is neither open nor closed.

Exercise 2.13. Show that $x^k \to x$ if and only if $x_j^k \to x_j$ for $j = 1, \ldots, n$. Thus convergence of a point sequence simply means that all the component sequences are convergent.

Solution 2.13. ($\Rightarrow$) Suppose $x^k \to x$. Then, as

$$|x_j^k - x_j| \leq \|x^k - x\|, \tag{2.11}$$

we have that $x_j^k \to x_j$ for each component j.

($\Leftarrow$) Conversely, suppose $x_j^k \to x_j$ for each component j. Then,

$$\|x^k - x\|^2 = \sum_{j=1}^{n} (x_j^k - x_j)^2 \to 0 = 0, \tag{2.12}$$

showing that $x^k \to x$. This is all rather informal, but the details are easy to fill in.

Exercise 2.14. Show that every simplex cone is closed.

Solution 2.14. Let $x_1, \ldots, x_t \in \mathbb{R}^n$ be linearly independent vectors, and consider the simplex cone spanned by these vectors. Let X be the matrix with columns $x_1, \ldots, x_t$. Then any point in the simplex cone can be written as $X\boldsymbol{\lambda}$ for some $\boldsymbol{\lambda} \geq 0$. X then has full column rank, and thus $X^T X$ is invertible, so we define the pseudo-inverse $X^\dagger = (X^T X)^{-1} X^T$. We then obtain $\boldsymbol{\lambda} = X^\dagger x$, such that the simplex cone can be written as the inverse image of the closed set $\{\boldsymbol{\lambda} : \boldsymbol{\lambda} \geq 0\}$ under the continuous map $x \mapsto X^\dagger x$. The simplex cone is therefore closed.

Boundary of a set. The *boundary* $\mathrm{bd}(S)$ of S is defined by $\mathrm{bd}(S) = \mathrm{cl}(S) \setminus \mathrm{int}(S)$. For instance, we have that $\mathrm{bd}(B(a, r)) = \{x \in \mathbb{R}^n : \|x - a\| = r\}$.

Exercise 2.15. Prove that $x \in \mathrm{bd}(S)$ if and only if each ball with center x intersects both S and the complement of S.

Solution 2.15. For $x \in \mathrm{bd}(S)$, we have that $x \in \mathrm{cl}(S)$ and $x \notin \mathrm{int}(S)$. Since $x \in \mathrm{cl}(S)$, every ball centred at x intersects S. Since $x \notin \mathrm{int}(S)$, every ball centred at x also intersects the complement of S.

Affine hull. The *affine hull* of a set S, denoted by $\mathrm{aff}(S)$, is the smallest affine set containing S.

Relative interior and relative boundary. We say that x is a *relative interior point* of S if there is an $r > 0$ such that

$$B^\circ(x, r) \cap \operatorname{aff}(S) \subseteq S. \tag{2.13}$$

This means that x is the center of some open ball whose intersection with $\operatorname{aff}(S)$ is contained in S. We let the *relative interior* of S, denoted by $\operatorname{int}(S)$, be the set of all such relative interior points of S. Similarly, we define the *relative boundary* of S by $\operatorname{rbd}(S) = \operatorname{cl}(S) \setminus \operatorname{rint}(S)$.

Exercise 2.16. Consider again the set $C = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 \leq 1\}$. Verify that:

(*i*) C is closed,

(*ii*) $\dim(C) = 2$,

(*iii*) $\operatorname{int}(C) = \emptyset$,

(*iv*) $\operatorname{bd}(C) = C$,

(*v*) $\operatorname{rint}(C) = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\}$,

(*vi*) $\operatorname{rbd}(C) = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 = 1\}$.

Solution 2.16. (*i*): Assume that $x^k \to x$ where $x^k = (x_k, y_k, 0) \in C$ for all k. As $x_k^2 + y_k^2 \leq 1$ for all k, we have that $x^2 + y^2 \leq 1$ by taking limits, showing that $x \in C$. Similarly, as $z_k = 0$ for all k, we have that $z = 0$, and thus $x \in C$. C is therefore closed.

(*ii*): We find only two linearly independent vectors in C, for instance $(1, 0, 0)$ and $(0, 1, 0)$. Thus, $\dim(C) = 2$. A similar argument is possible through the affinely independent points $(0, 0, 0)$, $(1, 0, 0)$, and $(0, 1, 0)$.

(*iii*): Any ball centred around a point $x = (x_1, x_2, 0) \in C$ will contain points of the form (x_1, x_2, ε) for $\varepsilon > 0$, which are not in C. Thus, $\operatorname{int}(C) = \emptyset$.

(*iv*): As $\operatorname{int}(C) = \emptyset$, we have that $\operatorname{bd}(C) = \operatorname{cl}(C) \setminus \operatorname{int}(C) = C \setminus \emptyset = C$.

(*v*): We have that $\operatorname{aff}(C)$ is the plane $z = 0$. For a point $x = (x_1, x_2, 0) \in C$ with $x_1^2 + x_2^2 < 1$, we can find a ball $B^\circ(x, r) \subset B^\circ(0, 1)$. Then

$$\begin{aligned} B^\circ(x, r) \cap \operatorname{aff}(C) &= \{(y_1, y_2, 0) : (y_1 - x_1)^2 + (y_2 - x_2)^2 < r^2\} \\ &= (B^\circ((x_1, x_2), r), 0) \subseteq C, \end{aligned} \tag{2.14}$$

so $\{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\} \subseteq \operatorname{rint}(C)$. If we have a point $x = (x_1, x_2, 0)$ with $x_1^2 + x_2^2 = 1$, any ball centred at x will contain points of the form (x_1, x_2, ε) for $\varepsilon > 0$, which are not in C. Thus, $\operatorname{rint}(C) = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\}$.

(v): We have simply

$$\begin{aligned}\operatorname{rbd}(C) &= \operatorname{cl}(C) \setminus \operatorname{rint}(C) = C \setminus \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\} \\ &= \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 = 1\}.\end{aligned} \tag{2.15}$$

Exercise 2.17. Show that every polytope in $\mathbb{R}^n$ is bounded. (Hint: use the properties of the norm: $\|x + y\| \leq \|x\| + \|y\|$ and $\|\lambda x\| = \lambda \|x\|$ for $\lambda \geq 0$.)

Solution 2.17. For a polytope $P = \operatorname{conv}(\{x_1, \ldots, x_t\})$, we can express any point $x \in P$ as

$$x = \sum_{i=1}^{t} \lambda_i x_i, \tag{2.16}$$

where $\lambda_i \geq 0$ and $\sum_{i=1}^{t} \lambda_i = 1$. We then have that

$$\begin{aligned}\|x\| &= \left\| \sum_{i=1}^{t} \lambda_i x_i \right\| \leq \sum_{i=1}^{t} \|\lambda_i x_i\| = \sum_{i=1}^{t} \lambda_i \|x_i\| \\ &\leq \max_{i=1,\ldots,t} \|x_i\| \sum_{i=1}^{t} \lambda_i = \max_{i=1,\ldots,t} \|x_i\|.\end{aligned} \tag{2.17}$$

Thus, P is bounded.

Exercise 2.18. Consider the standard simplex S_t. Show that it is compact, i.e., closed and bounded.

Solution 2.18. The standard simplex S_t is given by

$$S_t = \{x \in \mathbb{R}^t : x_i \geq 0, \sum_{i=1}^{t} x_i = 1\}, \tag{2.18}$$

or equivalently, $S_t = \operatorname{conv}(\{e_1, e_2, \ldots, e_t\})$ where e_i are the standard basis vectors in $\mathbb{R}^t$. As S_t is a polytope, it is bounded by the previous exercise. It is closed as it is the inverse of the closed set $\{1\}$ under the continuous map $x \mapsto \sum_{i=1}^{t} x_i$ intersected with the closed set $\{x : x_i \geq 0 \text{ for all } i\}$. Thus, S_t is compact.

Exercise 2.19. Give an example of a convex cone which is not closed.

Solution 2.19. Consider the set $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_2 > 0\}$. This set is a convex cone because for any $x, y \in C$ and any $\alpha, \beta \geq 0$, we have $\alpha x + \beta y \in C$. However, C is not closed because it does not contain the boundary line $x_2 = 0$. Thus, C is a convex cone that is not closed.

Exercise 2.20. Let $S \subseteq \mathbb{R}^2$ and let W be the set of all convex combinations of points in S. Prove that W is convex.

Solution 2.20. Let $u, v \in W$ be two convex combinations of points in S, i.e.,

$$u = \sum_{i=1}^{m} \lambda_i x_i \quad \text{and} \quad v = \sum_{j=1}^{n} \mu_j y_j. \tag{2.19}$$

where $x_i, y_j \in S$, $\lambda_i, \mu_j \geq 0$, and $\sum_{i=1}^{m} \lambda_i = 1$, $\sum_{j=1}^{n} \mu_j = 1$. For any $0 \leq \theta \leq 1$, we have that

$$\begin{aligned} \theta u + (1-\theta)v &= \theta \sum_{i=1}^{m} \lambda_i x_i + (1-\theta) \sum_{j=1}^{n} \mu_j y_j \\ &= \sum_{i=1}^{m} (\theta \lambda_i) x_i + \sum_{j=1}^{n} ((1-\theta)\mu_j) y_j. \end{aligned} \tag{2.20}$$

The new coefficients are non-negative, and furthermore we have

$$\sum_{i=1}^{m} (\theta \lambda_i) + \sum_{j=1}^{n} ((1-\theta)\mu_j) = \theta \cdot 1 + (1-\theta) \cdot 1 = 1, \tag{2.21}$$

showing that $\theta u + (1-\theta)v$ is a convex combination of points in S, and therefore lies in W. W is thus convex.

Proposition 2.1.1 (Convex sets). A set $C \subseteq \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its points. A set $C \subseteq \mathbb{R}^n$ is a convex cone if and only if it contains all non-negative combinations of its points.

Exercise 2.21. Prove the second statement of Proposition 2.1.1.

Solution 2.21. This is the definition of a convex cone?

Exercise 2.22. Give a geometrical interpretation of the induction step in the proof of Proposition 2.1.1.

Solution 2.22. This essentially traces out a Bezier curve.

Exercise 2.23. Let $S = \{(0,0), (1,0), (0,1)\}$. Show that $\text{conv}(S) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$.

Solution 2.23. A point in $\text{conv}(S)$ can be written as

$$x = \lambda_1(0,0) + \lambda_2(1,0) + \lambda_3(0,1) = (\lambda_2, \lambda_3), \tag{2.22}$$

where $\lambda_i \geq 0$ and $\lambda_1 + \lambda_2 + \lambda_3 = 1$. Thus, we have that $\lambda_2, \lambda_3 \geq 0$ and $\lambda_2 + \lambda_3 \leq 1$, showing that $\text{conv}(S) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$.

Exercise 2.24. Let S consist of the points $(0,0,0)$, $(1,0,0)$, $(0,1,0)$, $(0,0,1)$, $(1,1,0)$, $(1,0,1)$, $(0,1,1)$, and $(1,1,1)$. Show that $\mathrm{conv}(S) = \{(x_1,x_2,x_3) \in \mathbb{R}^3 : 0 \le x_i \le 1 \text{ for } i = 1,2,3\}$. Also determine $\mathrm{conv}(S \setminus \{(1,1,1)\})$ as the solution set of a system of linear inequalities. Illustrate these cases geometrically.

Solution 2.24. Consider instead the points $(0,0)$, $(1,0)$, $(0,1)$, and $(1,1)$ in $\mathbb{R}^2$, as the argument is similar and easier to visualize. A point in $\mathrm{conv}(S)$ can be written as

$$x = \lambda_1(0,0) + \lambda_2(1,0) + \lambda_3(0,1) + \lambda_4(1,1) = (\lambda_2 + \lambda_4, \lambda_3 + \lambda_4), \tag{2.23}$$

where $\lambda_i \ge 0$ and $\sum_{i=1}^4 \lambda_i = 1$. Thus we are bounded by the constraints

$$\begin{aligned} 0 \le x_1 = \lambda_2 + \lambda_4 \le 1, \\ 0 \le x_2 = \lambda_3 + \lambda_4 \le 1, \end{aligned}$$

showing that $\mathrm{conv}(S) = \{(x_1,x_2) \in \mathbb{R}^2 : 0 \le x_i \le 1 \text{ for } i = 1,2\}$. The same argument holds in $\mathbb{R}^3$, where we end up with the unit cube. For $\mathrm{conv}(S \setminus \{(1,1)\})$, we end up in the same scenario as the previous exercise, spanning out the unit triangle, as in Fig. 4

Exercise 2.25. Let $A, B \subseteq \mathbb{R}^n$. Prove that $\mathrm{conv}(A+B) = \mathrm{conv}(A) + \mathrm{conv}(B)$. (Hint: it is useful to consider the sum $\sum_{j,k} \lambda_j \mu_k (a_j + b_k)$ where $a_j \in A$, $b_k \in B$ and $\lambda_j, \mu_k \ge 0$ with $\sum_j \lambda_j = 1$ and $\sum_k \mu_k = 1$.)

Solution 2.25. We have that

$$\begin{aligned} \sum_{j,k} \lambda_j \mu_k (a_j + b_k) &= \sum_{j,k} \lambda_j \mu_k a_j + \sum_{j,k} \lambda_j \mu_k b_k \\ &= \sum_k \mu_k \left(\sum_j \lambda_j a_j \right) + \sum_j \lambda_j \left(\sum_k \mu_k b_k \right) \\ &= \sum_j \lambda_j a_j + \sum_k \mu_k b_k. \end{aligned}$$

As $\sum_{j,k} \lambda_j \mu_k = 1$, the left-hand side is a convex combination of points in $A+B$, while the right-hand side is an element in $\mathrm{conv}(A) + \mathrm{conv}(B)$. Therefore, $\mathrm{conv}(A+B) \subseteq \mathrm{conv}(A) + \mathrm{conv}(B)$. As $\mathrm{conv}(A) + \mathrm{conv}(B)$ is a convex set, it equals its convex hull, and

$$\mathrm{conv}(A+B) \subseteq \mathrm{conv}(\mathrm{conv}(A) + \mathrm{conv}(B)) = \mathrm{conv}(A) + \mathrm{conv}(B). \tag{2.24}$$

Exercise 2.37. Let $S \subseteq \mathbb{R}^n$. Show that either $\mathrm{int}(S) = \mathrm{rint}(S)$ or $\mathrm{int}(S) = \emptyset$.

Solution 2.37. If $\mathrm{int}(S) \neq \emptyset$, then there exists a point $x \in \mathrm{int}(S)$ and an $r > 0$ such that $B^\circ(x, r) \subseteq S$. As $B^\circ(x, r) \subseteq \mathrm{aff}(S)$, we have that $x \in \mathrm{rint}(S)$. Thus, $\mathrm{int}(S) \subseteq \mathrm{rint}(S)$.

Conversely, if $x \in \mathrm{rint}(S)$, then there exists an $r > 0$ such that $B^\circ(x, r) \cap \mathrm{aff}(S) \subseteq S$. As $B^\circ(x, r) \cap \mathrm{aff}(S)$ is open in $\mathrm{aff}(S)$, we have that $x \in \mathrm{int}(S)$. Thus, $\mathrm{rint}(S) \subseteq \mathrm{int}(S)$.

Therefore, if $\mathrm{int}(S) \neq \emptyset$, we have $\mathrm{int}(S) = \mathrm{rint}(S)$. If $\mathrm{int}(S) = \emptyset$, then the statement holds trivially.

Exercise 2.38. Prove Theorem 2.4.3. (Hint: To prove that $\mathrm{rint}(C)$ is convex, use Theorem 2.4.2. Concerning $\mathrm{int}(C)$, use Exercise 2.37. Finally, to show that $\mathrm{cl}(C)$ is convex, let $x, y \in \mathrm{cl}(C)$ and consider the two point sequences that converge to x and y respectively. Then, look at a convex combination of x and y and construct a suitable sequence!)

Solution 2.38. Theorem 2.4.2 states that a convex set has a "thin boundary", i.e., for $C \subseteq \mathbb{R}^n$ non-empty and convex, and $x_1 \in \mathrm{rint}(C)$, $x_2 \in \mathrm{cl}(C)$, we have

$$(1 - \lambda)x_1 + \lambda x_2 \in \mathrm{rint}(C) \tag{2.25}$$

for all $0 \leq \lambda < 1$.

Theorem 2.4.3 on the other hand states as follows: If $C \subseteq \mathbb{R}^n$ is a convex set, then all sets $\mathrm{rint}(C)$, $\mathrm{int}(C)$ and $\mathrm{cl}(C)$ are convex. Therefore, assume $C \subseteq \mathbb{R}^n$ is convex.

Let $x, y \in \mathrm{rint}(C)$. As $\mathrm{rint}(C) \subseteq C \subseteq \mathrm{cl}(C)$, we can simply apply Theorem 2.4.2 to show that $\mathrm{rint}(C)$ is convex. By Exercise 2.37, we have that either $\mathrm{int}(C) = \mathrm{rint}(C)$ or $\mathrm{int}(C) = \emptyset$. In both cases, $\mathrm{int}(C)$ is convex.

Finally, let $x, y \in \mathrm{cl}(C)$, and let $x^k, y^k \in C$ be sequences such that $x^k \to x$ and $y^k \to y$. As C is convex, $(1 - \lambda)x^k + \lambda y^k \in C$ for all $0 \leq \lambda \leq 1$, which converges to $(1 - \lambda)x + \lambda y$. Thus, $(1 - \lambda)x + \lambda y \in \mathrm{cl}(C)$, showing that $\mathrm{cl}(C)$ is convex.

3 Projection and separation

Exercise 3.1. Give an example where the nearest point is unique, and one where it is not. Find a point x and a set S such that every point of S is a nearest point to x!

Solution 3.1. Consider first the singleton set $S = \{1\} \subset \mathbb{R}$. Then, for any point $x \in \mathbb{R}$, the nearest point in S to x is simply 1, which is unique. Now, consider the set $S = \{-1, 1\} \subset \mathbb{R}$. If we take $x = 0$, then both -1 and 1 are nearest points to x, and thus the nearest point is not unique, and additionally every point in S is a nearest point to x.

Exercise 3.2. Let $a \in \mathbb{R}^n \setminus \{0\}$ and $x_0 \in \mathbb{R}^n$. Then there is a unique hyperplane H that contains x_0 and has a normal vector a. Verify this and find the value of the constant α in the definition of $H = \{x \in \mathbb{R}^n : a^T x = \alpha\}$.

Solution 3.2. The hyperplane H is defined by $a^T x = a^T x_0$. Thus, the constant α is given by $\alpha = a^T x_0$. To see that this hyperplane is unique, suppose there is another hyperplane H' with the same normal vector a that also contains x_0. Then H' must be defined by $a^T x = \beta$ for some constant β. Since $x_0 \in H'$, we have $a^T x_0 = \beta$. But we already have $a^T x_0 = \alpha$, so $\beta = \alpha$. Therefore, $H' = H$, proving the uniqueness of the hyperplane.

Exercise 3.3. Give an example of two disjoint sets S and T that cannot be separated by a hyperplane.

Solution 3.3. Consider the sets $S = \{x \in \mathbb{R}^2 : \|x\|_2 < r\}$ and $T = \{x \in \mathbb{R}^2 : r < \|x\|_2 < 2r\}$ for some $r > 0$. These sets are disjoint, as there are no points that belong to both S and T. However, they cannot be separated by a hyperplane, as there is no linear boundary that can separate the inner circle S from the annular region T without intersecting either set.

Exercise 3.4. In view of the previous remark, what about the separation of S and a point $p \notin \mathrm{aff}(S)$? Is there an easy way to find a separating hyperplane?

Solution 3.4.

Exercise 3.5. Let $C \subseteq \mathbb{R}^n$ be convex. Recall that if a point $x_0 \in C$ satisfies (3.2) for any $y \in C$, then x_0 is the (unique) nearest point to x in C. Now, let C be the unit ball in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n$ satisfy $\|x\| > 1$. Find the nearest point to x in C. What if $\|x\| \leq 1$?

Solution 3.5. (3.2) states

$$(x - x_0)^T (y - x_0) \leq 0 \quad \text{for all } y \in C. \tag{3.2}$$

When $\|x\| > 1$, we have that $x \notin C$. The nearest point to x in C is given by

$$x_0 = \frac{x}{\|x\|}. \tag{3.1}$$

Plugging this into (3.2), we have for $\|x\| > 1$ and any $y \in C$ that

$$\begin{aligned}
(x - x_0)^T (y - x_0) &= \left(x - \frac{x}{\|x\|}\right)^T \left(y - \frac{x}{\|x\|}\right) \\
&= \left(1 - \frac{1}{\|x\|}\right) x^T \left(y - \frac{x}{\|x\|}\right) \\
&\leq x^T y - \frac{x^T x}{\|x\|} = x^T y - \|x\| \\
&\leq \|x\| \|y\| - \|x\| \\
&\leq \|x\| - \|x\| = 0.
\end{aligned} \tag{3.2}$$

Therefore, x_0 is indeed the nearest point to x in C when $\|x\| > 1$.

For $\|x\| \leq 1$, we have that $x \in C$, and thus the nearest point to x in C is simply x itself.

Exercise 3.6. Let L be a line in $\mathbb{R}^n$. Find the nearest point in L to a point $x \in \mathbb{R}^n$. Use your result to find the nearest point on the line $L = \{(x, y) : x + 3y = 5\}$ to the point $(1, 2)$.

Solution 3.6. Let L be the defined by the line $\{b + td : t \in \mathbb{R}\}$, where b is a point on the line and d is a direction vector. The nearest point on L to a point $x \in \mathbb{R}^n$ can be found by projecting the vector $x - b$ onto the direction vector d. The projection is given by

$$\text{proj}_d(x - b) = b + \frac{(x - b)^T d}{\|d\|^2} d. \tag{3.3}$$

With $b = (5, 0)$, we note that $(1, 3)$ is orthogonal to the line, so we choose the direction vector to be $d = (-3, 1)$. Then, the nearest point on the line L to the point $(1, 2)$ is given by

$$\begin{aligned}
\text{proj}_d((1, 2) - (5, 0)) = \text{proj}_d((-4, 2)) &= (5, 0) + \frac{(-4, 2)^T (-3, 1)}{\|(-3, 1)\|^2} (-3, 1) \\
&= (5, 0) + \frac{14}{10}(-3, 1) = \left(\frac{25}{5}, 0\right) + \left(-\frac{21}{5}, \frac{7}{5}\right) \\
&= \left(\frac{4}{5}, \frac{7}{5}\right).
\end{aligned} \tag{3.4}$$

Therefore, the nearest point on the line L to the point $(1, 2)$ is $\left(\frac{4}{5}, \frac{7}{5}\right)$.

Exercise 3.7. Let H be a hyperplane in $\mathbb{R}^n$. Find the nearest point in H to a point $x \in \mathbb{R}^n$. In particular, find the nearest point to each of the point $(0,0,0)$ and $(1,2,2)$ in the hyperplane $H = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 1\}$.

Solution 3.7. The hyperplane H can be defined by the equation $a^T x = \alpha$, where a is the normal vector to the hyperplane and α is a constant. As we can write a hyperplane on the form $H = x_0 + L$ for some point $x_0 \in H$. Similarly as before, we can find the nearest point on H to a point $x \in \mathbb{R}^n$ by projecting the vector $x - x_0$ onto the normal vector a. This gives us

$$\mathrm{proj}_H(x) = (x - x_0) - \frac{(x - x_0)^T a}{\|a\|^2} a + x_0 = x - \frac{x^T a - \alpha}{\|a\|^2} a, \tag{3.5}$$

where the chosen point $x_0 \in H$ is arbitrary, as we can utilize the property $x_0^T a = \alpha$. For the hyperplane $H = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 1\}$, we can choose the normal vector $a = (1,1,1)$. Then, the nearest point on H to the point $(0,0,0)$ is given by

$$\begin{aligned} \mathrm{proj}_H((0,0,0)) &= (0,0,0) - \frac{(0,0,0)^T(1,1,1) - 1}{\|(1,1,1)\|^2}(1,1,1) \\ &= (0,0,0) - \frac{-1}{3}(1,1,1) \\ &= \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right). \end{aligned} \tag{3.6}$$

For the point $x = (1,2,2)$, we have $x^T a = 5$, so the nearest point is given by

$$\mathrm{proj}_H(x) = x - \frac{5-1}{3} a = (1,2,2) - \frac{4}{3}(1,1,1) = \left(-\frac{1}{3}, \frac{2}{3}, \frac{2}{3}\right). \tag{3.7}$$

Exercise 3.8. Let L be a linear subspace in $\mathbb{R}^n$ and let $q_1, \ldots, q_t$ be an orthonormal basis for L. Thus $q_1, \ldots, q_t$ span L, and $q_i^T q_j = \delta_{ij}$. Let Q be the $(n \times t)$-matrix whose j-th column is q_j. Define the associated matrix $P = QQ^T$. Show that Px is the nearest point in L to x. (The matrix P is called an orthogonal projector (or projection matrix)). Thus, performing the projection is simply to apply the linear transformation given by P. Let $L^\perp$ be the orthogonal complement of L. Explain why $(I - P)x$ is the nearest point in $L^\perp$ to x.

Solution 3.8. Note firstly that we have $Q^T Q = I_t$, such that we have $P^2 = QQ^T QQ^T = QI_tQ^T = P$, and thus P is indeed a projection matrix. Consider then the point $b = Px$. We have

$$b = Px = P^2 x = P(Px) = Pb. \tag{3.8}$$

Note next that for a vector $y \in L$, we have $y = Qz$ for some $z \in \mathbb{R}^t$. Then,

$$(x-b)^T(y-b) = (x-Px)^T(Qz-Px) = x^TQz - x^TPx - x^TPQz + x^TP^2x. \quad (3.9)$$

Using the properties of P, we have $PQ = Q$ and $P^2 = P$, such that

$$(x-b)^T(y-b) = x^TQz - x^TPx - x^TQz + x^TPx = 0. \quad (3.10)$$

Thus, $b = Px$ satisfies (3.2), and is therefore the nearest point in L to x.

For the orthogonal complement $L^\perp$, we have that any vector $w \in L^\perp$ satisfies $w^Tv = 0$ for all $v \in L$. Note that for any $x \in \mathbb{R}^n$, we can decompose x as $x = Px + (I-P)x$, where $Px \in L$. We have that $(I-P)x \in L^\perp$, since for any $v \in L$, we have

$$((I-P)x)^Tv = x^T(I-P)^Tv = x^T(I-P)v = x^T(v-Pv) = x^T(v-v) = 0. \quad (3.11)$$

Thus, $(I-P)x$ is orthogonal to every vector in L, and is therefore the nearest point in $L^\perp$ to x, as

$$\begin{aligned}(x-(I-P)x)^T(w-(I-P)x) &= (Px)^T(w-(I-P)x)\\ &= (Px)^Tw - x^TP(I-P)x\\ &= 0 - x^T(P-P)x\\ &= 0\end{aligned} \quad (3.12)$$

for any $w \in L^\perp$, as $Px \in L$.

Exercise 3.9. Let $L \subset \mathbb{R}^3$ be a subspace spanned by the vectors $(1,0,1)$ and $(0,1,0)$. Find the nearest point to $(1,2,3)$ in L using the results of the previous exercise.

Solution 3.9. We then have

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0\\ 0 & 1\\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix}, \quad (3.13)$$

such that P is given by

$$P = QQ^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & 0\\ 0 & 1\\ \frac{1}{\sqrt{2}} & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}}\\ 0 & 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}. \quad (3.14)$$

The nearest point to $(1,2,3)$ in L is then given by

$$Px = \begin{bmatrix} \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 1 & 0\\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1\\ 2\\ 3 \end{bmatrix} = \begin{bmatrix} 2\\ 2\\ 2 \end{bmatrix}. \quad (3.15)$$

Note: There is a mistake in the suggested solutions, as they haven't normalized the basis vectors.

Exercise 3.10. Show that the nearest point in $\mathbb{R}^n_+$ to $x \in \mathbb{R}^n$ is the point x^+ defined by $x_i^+ = \max(x_i, 0)$.

Solution 3.10. Let $C = \mathbb{R}^n_+$ and consider the point x^+ defined by $x_i^+ = \max(x_i, 0)$. We want to show that x^+ is the nearest point in C to x by verifying (3.2). For any $y \in C$, we have

$$(x - x^+)^T (y - x^+) = \sum_{i=1}^{n} (x_i - x_i^+)(y_i - x_i^+). \tag{3.16}$$

Note that if $x_i \geq 0$, then $x_i^+ = x_i$, and thus the term $(x_i - x_i^+)(y_i - x_i^+) = 0$. If $x_i < 0$, then $x_i^+ = 0$, and since $y_i \geq 0$ (as $y \in C$), we have $(x_i - x_i^+)(y_i - x_i^+) = x_i y_i \leq 0$. Therefore, each term in the sum is non-positive, and we conclude that

$$(x - x^+)^T (y - x^+) \leq 0, \tag{3.17}$$

for all $y \in C$. Thus, by (3.2), x^+ is indeed the nearest point in $\mathbb{R}^n_+$ to x.

Exercise 3.11. Find a set $S \subset \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$ with the property that every point of S is nearest to x in S!

Solution 3.11. Let $S = \{y \in \mathbb{R}^n : \|y\| = r\}$ for some fixed $r > 0$, which is the surface of a sphere centred at the origin. Let $x = 0 \in \mathbb{R}^n$. Then, for any point $y \in S$, we have $\|x - y\| = \|y\| = r$. Since all points in S are at the same distance r from x, every point of S is a nearest point to x in S.

Exercise 3.13. Let $C = [0, 1] \times [0, 1] \subset \mathbb{R}^2$ and let $y = (2, 2)$. Find *all* hyperplanes that separates C and y.

Solution 3.13. Let $H = \{x \in \mathbb{R}^2 : a^T x = \alpha\}$ be a hyperplane that separates C and y. Then, we must have either

$$\sup_{x \in C} a^T x < a^T y \quad \text{or} \quad \inf_{x \in C} a^T x > a^T y. \tag{3.18}$$

We have $\sup_{x \in C} a^T x = \max(a_1, 0) + \max(a_2, 0)$ and $\inf_{x \in C} a^T x = \min(a_1, 0) + \min(a_2, 0)$. Thus, the first condition becomes

$$\max(a_1, 0) + \max(a_2, 0) < 2(a_1 + a_2), \tag{3.19}$$

and the second condition becomes

$$\min(a_1, 0) + \min(a_2, 0) > 2(a_1 + a_2). \tag{3.20}$$

There are several cases where this holds. Firstly, if $a \geq 0$, then we require $b \in [a_1 + a_2, 2(a_1 + a_2))$. If $a_1 \geq 0 \geq a_2$, then we require $b \in [a_1, 2(a_1 + a_2))$, and similarly for $a_2 \geq 0 \geq a_1$, we require $b \in [a_2, 2(a_1 + a_2))$. If $a_1, a_2 \leq 0$, then we require $b \in (2(a_1 + a_2), a_1 + a_2]$.

Exercise 3.14. Let C be a unit ball in $\mathbb{R}^n$ and let $y \notin C$. Find a hyperplane that separates C and y.

Solution 3.14. Choosing x_0 to be the midpoint between y and the nearest point in C to y, i.e.,

$$x_0 = \tfrac{1}{2}\left(y + \frac{y}{\|y\|}\right), \tag{3.21}$$

we can then simply choose y to be the normal vector of the hyperplane. The hyperplane is then given by

$$H = \{x \in \mathbb{R}^n : y^T x = x_0\}. \tag{3.22}$$

Exercise 3.15. Find an example in $\mathbb{R}^2$ of two sets that have a unique separating hyperplane.

Solution 3.15. Let $S = \{(x, y) : x^2 + y^2 \leq 1, x \leq 0\}$ and $T = \{(x, y) : x^2 + y^2 \leq 1, x \geq 1\}$. The unique separating hyperplane is then given by $H = \{(x, y) : x = 0\}$.

Exercise 3.16. Let $S, T \subseteq \mathbb{R}^n$. Explain the following fact: there exists a hyperplane that separates S and T if and only if there is a linear function $l : \mathbb{R}^n \to \mathbb{R}$ such that $l(s) \leq l(t)$ for all $s \in S$ and $t \in T$. Is there a similar equivalence for the notion of strong separation?

Solution 3.16. A hyperplane that separates S and T can be defined by a normal vector a and a constant α such that $a^T s \leq \alpha$ for all $s \in S$ and $a^T t \geq \alpha$ for all $t \in T$. Defining the linear function $l(x) = a^T x$, we have that $l(s) \leq \alpha \leq l(t)$ for all $s \in S$ and $t \in T$, which shows one direction of the equivalence.

Conversely, if there exists a linear function $l : \mathbb{R}^n \to \mathbb{R}$ such that $l(s) \leq l(t)$ for all $s \in S$ and $t \in T$, we can express this linear function as $l(x) = a^T x$ for some vector a. Let $\alpha = \sup_{s \in S} l(s)$. Then, we have $a^T s \leq \alpha$ for all $s \in S$ and $a^T t \geq \alpha$ for all $t \in T$, which defines a hyperplane that separates S and T.

For strong separation, the equivalence holds with the additional requirement that there exists an $\varepsilon > 0$ such that $l(s) + \varepsilon < \alpha < l(t) - \varepsilon$ for all $s \in S$ and $t \in T$. This ensures that the hyperplane not only separates the sets but does so with a positive margin, thus achieving strong separation.

Exercise 3.17. Let C be a non-empty closed convex set in $\mathbb{R}^n$. Then the associated projection operator p_C is Lipschitz continuous with Lipschitz constant 1, i.e.,

$$\|p_C(x) - p_C(y)\| \leq \|x - y\| \quad \text{for all } x, y \in \mathbb{R}^n. \tag{3.23}$$

(Such an operator is called non-expansive). You are asked to prove this using the following procedure: Define $a = x - p_C(x)$ and $b = y - p_C(y)$. Verify that $(a-b)^T(p_C(x)-p_C(y)) \geq 0$. (Show first that $a^T(p_C(y)-p_C(x)) \leq 0$ and $b^T(p_C(x)-p_C(y)) \leq 0$ using (3.2). Then consider $\|x-y\|^2 = \|(a-b)+(p_C(x)-p_C(y))\|^2$ and do some calculations.)

Solution 3.17. Let's follow the helpful suggestions. Let $a = x - p_C(x)$ and $b = y - p_C(y)$. As $p_C(x)$ is the nearest point in C to x, we have by (3.2) that

$$(x - p_C(x))^T(z - p_C(x)) \leq 0 \quad \text{for all } z \in C, \tag{3.24}$$

and therefore specifically for $z = p_C(y)$ as well. Similarly for b and $p_C(y)$, we have

$$(p_C(y) - y)^T(p_C(y) - z) = (y - p_C(y))^T(z - p_C(y)) \leq 0 \quad \text{for all } z \in C, \tag{3.25}$$

and thus specifically for $z = p_C(x)$. Adding these two inequalities, we get

$$(a - b)^T(p_C(y) - p_C(x)) = a^T(p_C(y) - p_C(x)) + b^T(p_C(x) - p_C(y)) \leq 0. \tag{3.26}$$

This gives us

$$\begin{aligned} \|x - y\|^2 &= \|(a - b) + (p_C(x) - p_C(y))\|^2 \\ &= \|a - b\|^2 + \|p_C(x) - p_C(y)\|^2 + 2(a - b)^T(p_C(x) - p_C(y)) \\ &\geq \|a - b\|^2 + \|p_C(x) - p_C(y)\|^2 \\ &\geq \|p_C(x) - p_C(y)\|^2, \end{aligned} \tag{3.27}$$

as desired.

Exercise 3.18. Consider the outer description of closed convex sets given in Corollary 3.2.4. What is this description for each of the following sets:

(i) $C_1 = \{x \in \mathbb{R}^n : \|x\| \leq 1\}$,

(ii) $C_2 = \text{conv}(\{0, 1\}^n)$,

(iii) C_3 is the convex hull of the points $(1, 1)$, $(1, -1)$, $(-1, 1)$, and $(-1, -1)$ in $\mathbb{R}^2$.

(iv) C_4 is the convex hull of all vectors in $\mathbb{R}^n$ having components that are either 1 or -1.

Solution 3.18. I don't quite see the point of having almost identical subproblems, but alright.

(i) The half-planes are the tangents to the unit ball, as used in a previous exercise.

(ii) One half of the half-planes are given by $x_i = 0$ for $i = 1, \ldots, n$, and the other half by $x_i = 1$ for $i = 1, \ldots, n$.

(iii) Similarly, here the half-planes are given by $x_1 = 1$, $x_1 = -1$, $x_2 = 1$, and $x_2 = -1$, almost equivalently to (ii).

(iv) Again, the half-planes are given by $x_i = 1$ and $x_i = -1$ for $i = 1, \ldots, n$, almost equivalently to (ii) and (iii).

4 Representation of convex sets

Exercise 4.1. Consider the polytope $P \subset \mathbb{R}^2$ being the convex hull of the point $(0,0)$, $(1,0)$ and $(0,1)$ (so P is a simplex in $\mathbb{R}^2$).

(i) Find the unique face of P that contains the point $(1/3, 1/2)$.

(ii) Find all faces of P that contain the point $(1/3, 2/3)$.

(iii) Determine all the faces of P.

Solution 4.1. Beginning with (iii), we note the the dimension 0 faces of P are the vertices. The dimension 1 faces are the edges between the vertices, and the dimension 2 face is P itself. (i): The point $(1/3, 1/2)$ clearly does not lie on any of the edges or vertices, so the only face containing it is P itself. (ii): The point $(1/3, 2/3)$ lies on the edge between $(0,1)$ and $(1,0)$ $(\lambda = 1/3)$, so the faces containing it are this edge and P itself.

Exercise 4.2. Explain why an equivalent definition of face is obtained using the condition: if whenever $x_1, x_2 \in C$ and $(1/2)(x_1 + x_2) \in F$, then $x_1, x_2 \in F$.

Solution 4.2. Let F be a face of C. Then, by definition, if $x_1, x_2 \in C$ and $\lambda x_1 + (1-\lambda)x_2 \in F$ for some $\lambda \in (0,1)$, then $x_1, x_2 \in F$. In particular, this holds for $\lambda = 1/2$, so the condition is satisfied.

For the converse, assume that F satisfies the midpoint condition: whenever $x_1, x_2 \in C$ and $(1/2)(x_1 + x_2) \in F$, then $x_1, x_2 \in F$. We can essentially perform a binary search to find each point on the line segment between x_1 and x_2. More precisely, for a point $y = \lambda x_1 + (1-\lambda)x_2$ where $\lambda \in (0,1)$, we can express λ in its binary representation. This means that there is a sequence $\{\sigma_i (1/2)^i\}_{i=1}^{\infty}$ where $\sigma_i \in \{-1, 1\}$, $\sigma_1 = 1/2$, such that

$$\lambda_n = \sum_{i=1}^{n} \sigma_i (1/2)^i, \tag{4.1}$$

and $\lambda_n \to \lambda$ as $n \to \infty$. We can then define a sequence of points $\{y_n\}_{n=1}^{\infty}$ where $y_n = \lambda_n x_1 + (1-\lambda_n)x_2$. At each step n, we are only considering a subinterval $[x_l^n, x_r^n] \subseteq [x_1, x_2]$ where $y_n = \frac{1}{2}(x_l^n + x_r^n)$. By the midpoint condition, if $y_n \in F$, then $x_l^n, x_r^n \in F$. If $\lambda > \lambda_n$, we set $x_l^{n+1} = y_n$ and $x_r^{n+1} = x_r^n$. If $\lambda < \lambda_n$, we set $x_l^{n+1} = x_l^n$ and $x_r^{n+1} = y_n$. In either case, we have that $y_{n+1} = \frac{1}{2}(x_l^{n+1} + x_r^{n+1})$. Since $y_1 \in F$, we have that $x_l^1, x_r^1 \in F$. By induction, we have that $x_l^n, x_r^n \in F$ for all n. Since y_n is the midpoint of x_l^n and x_r^n, we have that $y_n \in F$ for all n. Finally, since $y_n \to y$ as $n \to \infty$ and F is closed, we have that $y \in F$. Thus, F is a face of C.

5 Convex functions

Lemma 5.1.1. Let $x_1 < x_2 < x_3$. Then the following statements are equivalent:

1. P_{x_2} is below the line segment P_{x_1}, P_{x_3}.
2. $\text{slope}(P_{x_1}, P_{x_2}) \leq \text{slope}(P_{x_2}, P_{x_3})$.
3. $\text{slope}(P_{x_1}, P_{x_3}) \leq \text{slope}(P_{x_2}, P_{x_3})$.

Exercise 5.1. Prove this lemma.

Solution 5.1. For the first point, we have that

$$f(x_2) \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} + f(x_1) \tag{5.1}$$

which is equivalent to

$$\frac{f(x_2) - f(x_1)}{x_2 - x_1} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1}. \tag{5.2}$$

This shows the equivalence of the first two points. The first point is also equivalent to

$$f(x_2) \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1} + f(x_3), \tag{5.3}$$

which is again equivalent to

$$\frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_3) - f(x_2)}{x_3 - x_2}, \tag{5.4}$$

exactly the third point.

Exercise 5.2. Show that the sum of convex functions is a convex function, and that λf is convex if f is convex and $\lambda \geq 0$.

Solution 5.2. Let $f_i : \mathbb{R} \to \mathbb{R}$ be convex functions for $i = 1, \ldots, n$. Then, for $f = \sum_{i=1}^n f_i$, we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &= \sum_{i=1}^n f_i(\lambda x + (1-\lambda)y) \\ &\leq \sum_{i=1}^n (\lambda f_i(x) + (1-\lambda) f_i(y)) \\ &= \lambda f(x) + (1-\lambda) f(y), \end{aligned}$$

showing that f is convex. For $\lambda \geq 0$ and f convex, we have

$$\begin{aligned}(\lambda f)(\alpha x + (1-\alpha)y) &= \lambda f(\alpha x + (1-\alpha)y) \\ &\leq \lambda\left(\alpha f(x) + (1-\alpha)f(y)\right) \\ &= \alpha(\lambda f)(x) + (1-\alpha)(\lambda f)(y),\end{aligned}$$

showing that λf is convex.

Exercise 5.3. Prove that the following functions are convex:

(*i*) $f(x) = x^2$,

(*ii*) $f(x) = |x|$,

(*iii*) $f(x) = x^p$ where $p \geq 1$,

(*iv*) $f(x) = e^x$,

(*v*) $f(x) = -\ln(x)$ on $\mathbb{R}_+$.

Solution 5.3. We show convexity as follow:

(*i*) As $f''(x) = 2 > 0$ for all $x \in \mathbb{R}$, the function is convex.

(*ii*) For $x, y \in \mathbb{R}$ and $\lambda \in [0,1]$, we have by the triangle inequality that

$$|\lambda x + (1-\lambda)y| \leq \lambda|x| + (1-\lambda)|y|, \tag{5.5}$$

showing that f is convex.

(*iii*) We have that $f''(x) = p(p-1)x^{p-2}$, which for even p is non-negative for all $x \in \mathbb{R}$, and for odd p is non-negative for all $x \geq 0$.

(*iv*) As $f''(x) = e^x > 0$ for all $x \in \mathbb{R}$, the function is convex.

(*v*) We have that $f''(x) = 1/x^2 > 0$ for all $x \neq 0$, so the function is convex on $\mathbb{R}_+$.

Exercise 5.4. Consider Example 5.1.2 again. Use the same technique as in the proof of arithmetic-geometric inequality except that you consider general weights $\lambda_1, \ldots, \lambda_r$ (non-negative with sum one). Which inequality do you obtain? It involves the so-called weighted arithmetic mean and the weighted geometric mean.

Solution 5.4. By convexity of $-\ln(x)$, we have for $\lambda_1, \dots, \lambda_r \geq 0$ with $\sum_{i=1}^r \lambda_i = 1$ that

$$-\ln\left(\sum_{i=1}^r \lambda_i x_i\right) \leq \sum_{i=1}^r \lambda_i(-\ln(x_i)) = -\ln\left(\prod_{i=1}^r x_i^{\lambda_i}\right), \tag{5.6}$$

which is equivalent to

$$\ln\left(\sum_{i=1}^r \lambda_i x_i\right) \geq \ln\left(\prod_{i=1}^r x_i^{\lambda_i}\right). \tag{5.7}$$

Exponentiating both sides gives the inequality

$$\sum_{i=1}^r \lambda_i x_i \geq \prod_{i=1}^r x_i^{\lambda_i}. \tag{5.8}$$

Exercise 5.5. Repeat Exercise 5.2, but now for convex functions defined on some convex set in $\mathbb{R}^n$.

Solution 5.5. I don't see what differs from Solution 5.2.

Exercise 5.6. Verify that every linear function from $\mathbb{R}^n$ to $\mathbb{R}$ is convex.

Solution 5.6. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a linear function. Then, for $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$, we have

$$f(\lambda x + (1-\lambda)y) = \lambda f(x) + (1-\lambda)f(y),$$

showing that f is convex.

Proposition 5.2.1 (Composition)**.** Assume that $f : \mathbb{R}^m \to \mathbb{R}$ is convex and $h : \mathbb{R}^m \to \mathbb{R}^n$ is affine. Then the composition $f \circ h$ is convex (where $(f \circ h)(x) = f(h(x))$).

Exercise 5.7. Prove Proposition 5.2.1.

Solution 5.7. As h is affine, we have for $x, y \in \mathbb{R}^n$ and $\lambda \in [0,1]$ that

$$h(\lambda x + (1-\lambda)y) = \lambda h(x) + (1-\lambda)h(y). \tag{5.9}$$

By convexity of f, we then have

$$\begin{aligned}
(f \circ h)(\lambda x + (1-\lambda)y) &= f(h(\lambda x + (1-\lambda)y)) \\
&= f(\lambda h(x) + (1-\lambda)h(y)) \\
&\leq \lambda f(h(x)) + (1-\lambda)f(h(y)) \\
&= \lambda(f \circ h)(x) + (1-\lambda)(f \circ h)(y),
\end{aligned}$$

showing that $f \circ h$ is convex.

Exercise 5.8. Let $f : C \to \mathbb{R}$ be convex and let $w \in \mathbb{R}^n$. Show that the function $x \mapsto f(x + w)$ is convex.

Solution 5.8. Let $g(x) = f(x + w)$. Then, for $x, y \in C$ and $\lambda \in [0, 1]$, we have

$$\begin{aligned} g(\lambda x + (1 - \lambda)y) &= f(\lambda x + (1 - \lambda)y + w) \\ &= f(\lambda(x + w) + (1 - \lambda)(y + w)) \\ &\leq \lambda f(x + w) + (1 - \lambda)f(y + w) \\ &= \lambda g(x) + (1 - \lambda)g(y), \end{aligned}$$

showing that g is convex.

Theorem 5.2.3 (Epigraph)**.** Let $f : C \to \mathbb{R}$ where $C \subseteq \mathbb{R}^n$ is a convex set. Then f is a convex function if and only if $\operatorname{epi}(f)$ is a convex set.

Exercise 5.9. Prove Theorem 5.2.3 (just apply the definitions).

Solution 5.9. Assume that f is convex. Let $(x, a), (y, b) \in \operatorname{epi}(f)$ and $\lambda \in [0, 1]$. Then, by definition of the epigraph, we have that $f(x) \leq a$ and $f(y) \leq b$. As f is convex, we have

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &\leq \lambda f(x) + (1 - \lambda)f(y) \\ &\leq \lambda a + (1 - \lambda)b, \end{aligned}$$

showing that $\lambda(x, a) + (1 - \lambda)(y, b) \in \operatorname{epi}(f)$. Thus, $\operatorname{epi}(f)$ is convex.

Similarly, assume that $\operatorname{epi}(f)$ is convex. As $(x, f(x)), (y, f(y)) \in \operatorname{epi}(f)$ for all $x, y \in C$, we have for $\lambda \in [0, 1]$ that

$$\lambda(x, f(x)) + (1 - \lambda)(y, f(y)) \in \operatorname{epi}(f), \tag{5.10}$$

implying that

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y), \tag{5.11}$$

showing that f is convex.

Exercise 5.10. By the result above we have that if f and g are convex functions, then the function $\max\{f, g\}$ is also convex. Prove this result directly from the definition of a convex function.

Solution 5.10. Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be convex functions, and define $h(x) = \max\{f(x), g(x)\}$. Then, for any $x, y \in \mathbb{R}^n$ and any $0 \leq \lambda \leq 1$, we have

$$\begin{aligned} h(\lambda x + (1 - \lambda)y) &= \max\{f(\lambda x + (1 - \lambda)y), g(\lambda x + (1 - \lambda)y)\} \\ &\leq \max\{\lambda f(x) + (1 - \lambda)f(y), \lambda g(x) + (1 - \lambda)g(y)\} \\ &\leq \lambda \max\{f(x), g(x)\} + (1 - \lambda)\max\{f(y), g(y)\} \\ &= \lambda h(x) + (1 - \lambda)h(y), \end{aligned}$$

showing that h is convex.

Exercise 5.11. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\alpha \in \mathbb{R}$. Show that the set $\{x \in \mathbb{R}^n : f(x) \leq \alpha\}$ is a convex set. Each such set is called a sublevel set.

Solution 5.11. Let $S = \{x \in \mathbb{R}^n : f(x) \leq \alpha\}$. For $x, y \in S$ and $\lambda \in [0, 1]$, we have by convexity of f that

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \lambda f(x) + (1-\lambda)f(y) \\ &\leq \lambda\alpha + (1-\lambda)\alpha = \alpha, \end{aligned}$$

showing that $\lambda x + (1-\lambda)y \in S$. Thus, S is a convex set.

Exercise 5.12. Verify that the function $x \mapsto \|x\|_p$ is positively homogenous.

Solution 5.12. Let $x \in \mathbb{R}^n$ and $\lambda \geq 0$. Then, we have

$$\|\lambda x\|_p = \left(\sum_{i=1}^n |\lambda x_i|^p\right)^{1/p} = \left(\sum_{i=1}^n \lambda^p |x_i|^p\right)^{1/p} = \lambda\left(\sum_{i=1}^n |x_i|^p\right)^{1/p} = \lambda\|x\|_p, \quad (5.12)$$

showing that the function is indeed positively homogenous.

Exercise 5.13. Consider the support function of an optimization problem with a linear objective function, i.e., let $f(c) \coloneqq \max\{c^T x : x \in S\}$ where $S \subseteq \mathbb{R}^n$ is a given non-empty set. Show that f is positively homogenous. Therefore (due to Example 5.2.2), the support function is convex and positively homogeneous when S is a compact convex set.

Solution 5.13. Let $c \in \mathbb{R}^n$ and $\lambda \geq 0$. Then, we have

$$\begin{aligned} f(\lambda c) &= \max\{(\lambda c)^T x : x \in S\} = \max\{\lambda(c^T x) : x \in S\} \\ &= \lambda \max\{c^T x : x \in S\} = \lambda f(c), \end{aligned}$$

showing that f is positively homogenous.

Exercise 5.14. Let $f(x) = x^T x = \|x\|^2$ for $x \in \mathbb{R}^n$. Show that the directional derivative $f'(x_0; z)$ exists for all x_0 and non-zero z and that $f'(x_0; z) = 2z^T x_0$.

Solution 5.14. We have

$$\begin{aligned} \lim_{t\to 0} \frac{f(x_0 + tz) - f(x_0)}{t} &= \lim_{t\to 0} \frac{(x_0 + tz)^T(x_0 + tz) - x_0^T x_0}{t} \\ &= \lim_{t\to 0} \frac{tz^T x_0 + tx_0^T z + t^2 z^T z}{t} \\ &= \lim_{t\to 0} 2z^T x_0 + tz^T z \\ &= 2z^T x_0, \end{aligned}$$

showing that the directional derivative $f'(x_0; z)$ exists and equals $2z^T x_0$.

Exercise 5.15. A quadratic function is a function of the form

$$f(x) = x^T Ax + c^T x + \alpha \tag{5.13}$$

for some (symmetric) matrix $A \in \mathbb{R}^{n\times n}$, a vector $c \in \mathbb{R}^n$, and a scalar $\alpha \in \mathbb{R}$. Discuss whether f is convex.

Solution 5.15. By Example 5.3.1, the hessian of $x^T Ax$ is $2A$. As the other terms are linear or constant, we have that the hessian of f is also $2A$. Therefore, f is convex if and only if A is positive semidefinite.

Exercise 5.16. Assume that f and g are convex functions defined on an interval I. Determine which of the following functions that are convex or concave:

(*i*) λf where $\lambda \in \mathbb{R}$,

(*ii*) $\min\{f, g\}$,

(*iii*) $|f|$.

Solution 5.16. The answers are as follows:

(*i*) If $\lambda \geq 0$, then λf is convex by Solution 5.2. If $\lambda \leq 0$, then λf is concave.

(*ii*) The function $\min\{f, g\}$ is in general neither convex nor concave. For example, let $f(x) = x$ and $g(x) = -x$ for $x \in \mathbb{R}$. Then, we have that $\min\{f, g\} = -|x|$, which is concave but not convex. On the other hand, let $f(x) = g(x) = x^2$ for $x \in \mathbb{R}$. Then, we have that $\min\{f, g\} = x^2$, which is convex but not concave.

(*iii*) The function $|f|$ is in general neither convex nor concave. For example, let $f(x) = x^2 - 1$ for $x \in \mathbb{R}$. Then, we have

$$|f(0)| = |-1| = 1$$
$$\frac{1}{2}|f(-1)| + \frac{1}{2}|f(1)| = \frac{1}{2}|0| + \frac{1}{2}|0| = 0,$$

showing that $|f|$ is not convex. On the other hand, with $f(x) = x$, we have $|f(x)| = |x|$, which is convex but not concave.

Exercise 5.17. Let $f, g : I \to \mathbb{R}$ where I is an interval. Assume that f and $f + g$ are both convex. Does this imply that g is convex? Or concave? What if $f + g$ is convex and f is concave?

Solution 5.17. Not generally, no. If $g = -f$, then $f + g = 0$ is convex, but g is concave. On the other hand, if $g = f$, then $f + g = 2f$ is convex, and g is convex. For the final question, let $h = f + g$. Then, we have $g = h - f$, showing that g is the sum of two convex functions, showing that g is convex.

Exercise 5.18. Let $f : [a, b] \to \mathbb{R}$ be a convex function. Show that

$$\max\{f(x) : x \in [a, b]\} = \max\{f(a), f(b)\}, \tag{5.14}$$

i.e., a convex function defined on a closed real interval attains its maximum in one of the endpoints.

Solution 5.18. Assume for contradiction that there exists some $x^* \in (a, b)$ such that $f(x^*) > \max\{f(a), f(b)\}$. By convexity of f, we have for $\lambda \in [0, 1]$ that

$$f(\lambda a + (1 - \lambda)b) \leq \lambda f(a) + (1 - \lambda)f(b). \tag{5.15}$$

Choosing λ such that $\lambda a + (1 - \lambda)b = x^*$ gives

$$f(x^*) \leq \lambda f(a) + (1 - \lambda)f(b) \leq \max\{f(a), f(b)\}, \tag{5.16}$$

a contradiction. Thus, the maximum is attained at one of the endpoints.

Theorem 5.3.1 (Continuity)**.** Let $f : C \to \mathbb{R}$ be a convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Then f is continuous on C.

Exercise 5.19. Let $f : I \to \mathbb{R}$ be a convex function defined on a bounded interval I. Prove that f must be bounded below (i.e., there is a number L such that $f(x) \geq L$ for all $x \in I$). Is f also bounded above?

Solution 5.19. As f is convex on the bounded interval $I = [a, b]$, we have by Theorem 5.3.1 that f is continuous on I. By the extreme value theorem, f attains its minimum on I, i.e., there exists some $x^* \in I$ such that $f(x^*) \leq f(x)$ for all $x \in I$. Thus, f is bounded below by $L = f(x^*)$. By Exercise 5.18, we have that f attains its maximum at one of the endpoints, so f is also bounded above.

Exercise 5.20. Let $f, g : \mathbb{R} \to \mathbb{R}$ be convex functions and assume that f is increasing. Prove that the composition $f \circ g$ is convex.

Solution 5.20. For $x, y \in \mathbb{R}$ and $\lambda \in [0, 1]$, we have by convexity of g that

$$g(\lambda x + (1 - \lambda)y) \leq \lambda g(x) + (1 - \lambda)g(y). \tag{5.17}$$

As f is increasing, we then have

$$\begin{aligned}
(f \circ g)(\lambda x + (1 - \lambda)y) &= f(g(\lambda x + (1 - \lambda)y)) \\
&\leq f(\lambda g(x) + (1 - \lambda)g(y)) \\
&\leq \lambda f(g(x)) + (1 - \lambda)f(g(y)) \\
&= \lambda(f \circ g)(x) + (1 - \lambda)(f \circ g)(y),
\end{aligned}$$

showing that $f \circ g$ is convex.

Exercise 5.21. Find the optimal solutions of the problem

$$\min\{f(x) : a \leq x \leq b\} \tag{5.18}$$

where $a < b$ and $f : \mathbb{R} \to \mathbb{R}$ is a differentiable convex function.

Solution 5.21. If there exists some $x^* \in [a, b]$ such that $f'(x^*) = 0$, then x^* is a global minimizer. Otherwise, if $f'(x) > 0$ for all $x \in [a, b]$, then f is strictly increasing on $[a, b]$ and the minimum is attained at $x = a$. If $f'(x) < 0$ for all $x \in [a, b]$, then f is strictly decreasing on $[a, b]$ and the minimum is attained at $x = b$.

Proposition 5.1.2 (Increasing slopes)**.** A function $f : \mathbb{R} \to \mathbb{R}$ is convex if and only if for each $x_0 \in \mathbb{R}$ the slope function

$$x \mapsto \frac{f(x) - f(x_0)}{x - x_0} \tag{5.19}$$

is increasing on $\mathbb{R} \setminus \{x_0\}$.

Exercise 5.22. Let $f : (0, \infty) \to \mathbb{R}$ and define the function $g : (0, \infty) \to \mathbb{R}$ by $g(x) = xf(1/x)$. Prove that f is convex if and only if g is convex. Hint: Prove that

$$\frac{g(x) - g(x_0)}{x - x_0} = f(1/x_0) - \frac{1}{x_0} \cdot \frac{f(1/x) - f(1/x_0)}{(1/x) - (1/x_0)}, \tag{5.20}$$

and use Proposition 5.1.2. Why is the function $x \mapsto xe^{1/x}$ convex?

Solution 5.22. We have that

$$\begin{aligned}
\frac{g(x) - g(x_0)}{x - x_0} &= \frac{xf(\frac{1}{x}) - x_0 f(\frac{1}{x_0})}{x - x_0} \\
&= \frac{xf(\frac{1}{x}) - xf(\frac{1}{x_0}) + xf(\frac{1}{x_0}) - x_0 f(\frac{1}{x_0})}{x - x_0} \\
&= \frac{x(f(\frac{1}{x}) - f(\frac{1}{x_0}))}{x - x_0} + \frac{(x - x_0)f(\frac{1}{x_0})}{x - x_0} \\
&= f(\tfrac{1}{x_0}) + \frac{x(f(\frac{1}{x}) - f(\frac{1}{x_0}))}{x - x_0} \\
&= f(\tfrac{1}{x_0}) + \frac{1}{x_0} \frac{xx_0(f(\frac{1}{x}) - f(\frac{1}{x_0}))}{x - x_0}.
\end{aligned}$$

Then, as

$$\frac{xx_0}{x - x_0} = \frac{1}{\frac{x}{xx_0} - \frac{x_0}{xx_0}} = \frac{1}{\frac{1}{x_0} - \frac{1}{x}},$$

Eq. (5.20) follows.

Assume f is convex. Let h be the slope function of f at x_0, i.e.,

$$h(x) = \frac{f(x) - f(x_0)}{x - x_0}. \tag{5.21}$$

As $1/x$ decreases as x increases, we have that $h(1/x)$ is decreasing. Therefore, $-h(1/x)$ is increasing, and offsetting this with the constant $f(1/x_0)$ maintains the increasing property. Thus, by Proposition 5.1.2, g is convex, as it's slope function is increasing.

With $f(x) = e^x$, we have $g(x) = xf(1/x) = xe^{1/x}$, which is therefore convex, as e^x is convex.

Corollary 5.1.8. Let $f : \mathbb{R} \to \mathbb{R}$ be a convex function and let $x_0 \in \mathbb{R}$. Then the following three statements are equivalent:

(i) x_0 is a local minimum for f,

(ii) x_0 is a global minimum for f,

(iii) $0 \in \partial f(x_0)$.

Theorem 5.1.9 (Mean value theorem)**.** Let $f : [a, b] \to \mathbb{R}$ be a convex function. Then there exists a $c \in (a, b)$ such that

$$\frac{f(b) - f(a)}{b - a} \in \partial f(c). \tag{5.22}$$

Exercise 5.23. Prove Theorem 5.1.9 as follows. Consider the function

$$g(x) = f(x) - f(a) - \frac{f(b) - f(a)}{b - a}(x - a). \tag{5.23}$$

Explain why g is convex and that it has a minimum point at some $c \in (a, b)$ (note that $g(a) = g(b) = 0$ and g is not constant). Then verify that

$$\partial g(c) = \partial f(c) - \frac{f(b) - f(a)}{b - a} \tag{5.24}$$

and use Corollary 5.1.8.

Solution 5.23. The function g is convex as it is the sum of a convex function and an affine function. As g is convex on the closed interval $[a, b]$, it attains its minimum at some point $c \in [a, b]$ by the extreme value theorem. As $g(a) = g(b) = 0$ and g is not constant, the minimum must be attained at some $c \in (a, b)$, as

$$g(\lambda a + (1 - \lambda)b) \leq \lambda g(a) + (1 - \lambda)g(b) = 0 \tag{5.25}$$

for all $\lambda \in [0, 1]$.

For each $x \in (a, b)$, we have

$$g'_{\pm}(x) = f'_{\pm}(x) - \frac{f(b) - f(a)}{b - a}, \tag{5.26}$$

showing that

$$\partial g(c) = \partial f(c) - \frac{f(b) - f(a)}{b - a}. \tag{5.27}$$

As there then exists a minimizer $c \in (a, b)$ of g, we have by Corollary 5.1.8 that $0 \in \partial g(c)$, implying that

$$\frac{f(b) - f(a)}{b - a} \in \partial f(c), \tag{5.28}$$

exactly the statement of the theorem.

Exercise 5.24. Let $f : \mathbb{R} \to \mathbb{R}$ be an increasing convex function and let $g : C \to \mathbb{R}$ be a convex function defined on a convex set C in $\mathbb{R}^n$. Prove that the composition $f \circ g$ is convex.

Solution 5.24. This is essentially Exercise 5.20 restricted to convex functions defined on convex sets in $\mathbb{R}^n$. The proof is identical, just replacing $\mathbb{R}$ with C where appropriate.

Exercise 5.25. Prove that the function given by $h(x) = e^{x^T A x}$ is convex when A is positive definite.

Solution 5.25. Let $f(x) = e^x$ and $g(x) = x^T A x$. As A is positive definite, we have by Exercise 5.15 that g is convex. As f is increasing and convex, we have by Exercise 5.20 that the composition $f \circ g$ is convex, i.e., h is convex.

Corollary 4.3.4 (Minkowski's theorem)**.** If $C \subseteq \mathbb{R}^n$ is a compact convex set, then C is the convex hull of its extreme points, i.e., $C = \operatorname{conv}(\operatorname{ext}(C))$.

Exercise 5.26. Let $f : C \to \mathbb{R}$ be a convex function defined on a compact convex set $C \subseteq \mathbb{R}^n$. Show that f attains its maximum in an extreme point. Hint: use Minkowski's theorem (Corollary 4.3.4).

Solution 5.26. As C is a compact convex set, we have by Corollary 4.3.4 that $C = \operatorname{conv}(\operatorname{ext}(C))$. Let $x^* \in C$ be a maximizer of f on C. Then, by Minkowski's theorem, we can write x^* as a convex combination of extreme points, i.e., there exist $r \in \mathbb{N}$, $\lambda_1, \ldots, \lambda_r \geq 0$ with $\sum_{i=1}^r \lambda_i = 1$ and $x_1, \ldots, x_r \in \operatorname{ext}(C)$ such that

$$x^* = \sum_{i=1}^{r} \lambda_i x_i. \tag{5.29}$$

By convexity of f, we then have

$$f(x^*) \leq \sum_{i=1}^{r} \lambda_i f(x_i) \leq \max_{1 \leq i \leq r} f(x_i). \tag{5.30}$$

As x^* is a maximizer of f on C, we must have equality throughout as each $x_i \in C$, showing that f attains its maximum at one of the extreme points $x_1, \ldots, x_r$.

Exercise 5.27. Let $C \subseteq \mathbb{R}^n$ be a convex set and consider the distance function d_C defined by

$$d_C(x) = \inf\{\|x - c\| : c \in C\}. \tag{5.31}$$

Show that d_C is a convex function.

Solution 5.27. Let $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. For any $\varepsilon > 0$, there exist $c_x, c_y \in C$ such that

$$\|x - c_x\| \leq d_C(x) + \varepsilon \quad \text{and} \quad \|y - c_y\| \leq d_C(y) + \varepsilon. \tag{5.32}$$

As C is convex, we have that $c = \lambda c_x + (1 - \lambda)c_y \in C$. Therefore, we have

$$\begin{aligned} d_C(\lambda x + (1 - \lambda)y) &= \inf\{\|\lambda x + (1 - \lambda)y - c\| : c \in C\} \\ &\leq \|\lambda(x - c_x) + (1 - \lambda)(y - c_y)\| \\ &\leq \lambda\|x - c_x\| + (1 - \lambda)\|y - c_y\| \\ &\leq \lambda(d_C(x) + \varepsilon) + (1 - \lambda)(d_C(y) + \varepsilon) \\ &= \lambda d_C(x) + (1 - \lambda)d_C(y) + \varepsilon. \end{aligned}$$

As $\varepsilon > 0$ was arbitrary, we have

$$d_C(\lambda x + (1 - \lambda)y) \leq \lambda d_C(x) + (1 - \lambda)d_C(y), \tag{5.33}$$

showing that d_C is convex.

Theorem 5.3.5 (Charachterization via gradients)**.** Let $f : C \to \mathbb{R}$ be a differentiable function defined on an open convex set $C \subseteq \mathbb{R}^n$. Then the following conditions are equivalent:

(i) f is convex,

(ii) $f(x) \geq f(x_0) + \nabla f(x_0)^T(x - x_0)$ for all $x, x_0 \in C$,

(iii) $(\nabla f(x) - \nabla f(x_0))^T(x - x_0) \geq 0$ for all $x, x_0 \in C$.

Corollary 6.1.1 (Global minimum)**.** Let $f : C \to \mathbb{R}$ be a differentiable convex function defined on an open convex set $C \subseteq \mathbb{R}^n$. Let $x^* \in C$. Then the following three statements are equivalent:

(i) x^* is a local minimum for f,

(ii) x^* is a global minimum for f,

(iii) $\nabla f(x^*) = 0$ (i.e., all partial derivatives at x^* are zero).

Exercise 5.28. Prove Corollary 6.1.1 using Theorem 5.3.5.

Solution 5.28. If x^* is a local minimum for f, then we have by Theorem 5.3.5 (ii) that

$$f(x) \geq f(x^*) + \nabla f(x^*)^T (x - x^*) \tag{5.34}$$

for all $x \in C$. If then $\nabla f(x^*) = 0$, we have that $f(x) \geq f(x^*)$ for all $x \in C$, showing that x^* is a global minimum.

Assume now that x^* is a local minimum for f. Consider any $y \in C$ and define the function $g : [0,1] \to \mathbb{R}$ by

$$g(\lambda) = f(x^* + \lambda(y - x^*)). \tag{5.35}$$

As f is convex and $\lambda \mapsto x^* + \lambda(y - x^*)$ is affine, we have by Proposition 5.2.1 that g is convex. As x^* is a local minimum for f, we have that $\lambda = 0$ is a local minimum for g, and $g'(0) = 0$ by Corollary 5.1.8. Therefore, we have that

$$0 = g'(0) = \nabla f(x^*)^T (y - x^*) \tag{5.36}$$

for all $y \in C$, showing that $\nabla f(x^*) = 0$.

We have thus shown that (iii) implies (ii) implies (i) implies (iii), and the statements are equivalent.

Exercise 5.29. Compare the notion of support for a convex function to the notion of supporting hyperplane of a convex set (see Section 3.2). Have in mind that f is convex if and only if $\mathrm{epi}(f)$ is a convex set. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and consider a supporting hyperplane of $\mathrm{epi}(f)$. Interpret the hyperplane in terms of functions, and derive a result saying that every convex function has a support at every point.

Solution 5.29. Let $a^T x + t = \alpha$ be a supporting hyperplane of $\mathrm{epi}(f)$ at the point $(y, f(y))$. The hyperplane is then given by

$$a^T x + t = a^T y + f(y), \tag{5.37}$$

which we can rewrite as

$$t = a^T (y - x) + f(y). \tag{5.38}$$

As $\mathrm{epi}(f)$ lies above the hyperplane, we have that

$$a^T x + f(x) \geq a^T y + f(y), \tag{5.39}$$

such that

$$f(x) \geq a^T(x - y) + f(y). \tag{5.40}$$

Thus, the hyperplane corresponds to the affine function $x \mapsto a^T(x - y) + f(y)$ which supports f at the point y.

6 Nonlinear and convex optimization

Exercise 6.1. Consider the least squares problem minimize $\|Ax - b\|$ over all $x \in \mathbb{R}^n$. From linear algebra we know that the optimal solutions to this problem are precisely the solutions to the linear system (called the normal equations)

$$A^T Ax = A^T b. \tag{6.1}$$

Show this using optimization theory by considering the function $f(x) = \|Ax - b\|^2$.

Solution 6.1. We have that the function f can be written as

$$f(x) = (Ax - b)^T (Ax - b) = x^T A^T Ax - 2b^T Ax + b^T b, \tag{6.2}$$

and we note that f is differentiable with gradient

$$\nabla f(x) = 2A^T Ax - 2A^T b. \tag{6.3}$$

f is convex, as the Hessian $A^T A$ is positive semidefinite for any matrix A (as $x^T A^T Ax = \|Ax\|^2 \geq 0$ for all x). Since f is convex we know that a point x^* is a global minimizer if and only if

$$\nabla f(x^*) = 0, \tag{6.4}$$

which is equivalent to

$$A^T Ax^* = A^T b. \tag{6.5}$$

Thus, the optimal solutions to the least squares problem are precisely the solutions to the normal equations.

Theorem 6.2.1 (Optimality condition)**.** Let $x^* \in C$. Then x^* is a (local and therefore global) minimum of f over C if and only if

$$\nabla f(x^*)^T (x - x^*) \geq 0 \quad \text{for all } x \in C. \tag{6.6}$$

Exercise 6.2. Prove the optimality condition is correct in Example 6.2.1.

Solution 6.2. Let's consider Theorem 6.2.1 in the case where $C = \mathbb{R}^n_+$. If $x^* \in \operatorname{int}(R^n_+)$, then the condition is clearly satisfied only when $\nabla f(x^*) = 0$, which is the standard unconstrained optimality condition. If $x^* \in \operatorname{bd}(\mathbb{R}^n_+)$, let $I = \{i : x^*_i = 0\}$ be the set of indices where x^* is on the boundary. Then Eq. (6.6) is reduces to

$$\begin{gathered}
\nabla f(x^*)^T (x - x^*) \geq 0 \\
\sum_{i=1}^{n} \frac{\partial f(x^*)}{\partial x_i}(x_i - x^*_i) \geq 0 \\
\sum_{i \in I} \frac{\partial f(x^*)}{\partial x_i} x_i + \sum_{i \notin I} \frac{\partial f(x^*)}{\partial x_i}(x_i - x^*_i) \geq 0.
\end{gathered}$$

Considering the sums separately, we see that the first sum is nonnegative for all $x \in \mathbb{R}^n_+$ if and only if $\frac{\partial f(x^*)}{\partial x_i} \geq 0$ for all $i \in I$. The second sum is nonnegative for all $x \in \mathbb{R}^n_+$ if and only if $\frac{\partial f(x^*)}{\partial x_i} = 0$ for all $i \notin I$. This is equivalent to the conditions stated in Example 6.2.1.

Exercise 6.3. Consider the problem to minimize a (continuously differentiable) convex function f subject to $x \in C = \{x \in \mathbb{R}^n : 0 \leq x \leq p\}$ where p is some nonnegative vector. Find the optimality conditions for this problem. Suggest a numerical algorithm for solving this problem.

Solution 6.3. We proceed similarly to the previous exercise. First, let

$$A = \{i : x_i^* = 0\}, \quad B = \{i : x_i^* = p_i\}, \quad C = \{i : 0 < x_i^* < p_i\}, \tag{6.7}$$

for an optimal point x^*. Then, the optimality condition from Theorem 6.2.1 reduces to

$$\sum_{i \in A} \frac{\partial f(x^*)}{\partial x_i} x_i + \sum_{i \in B} \frac{\partial f(x^*)}{\partial x_i}(x_i - p_i) + \sum_{i \in C} \frac{\partial f(x^*)}{\partial x_i}(x_i - x_i^*) \geq 0. \tag{6.8}$$

The sum over A is nonnegative for all $x \in C$ if and only if $\frac{\partial f(x^*)}{\partial x_i} \geq 0$ for all $i \in A$. As $x_i - p_i < 0$ for all $i \in B$, the sum over B is nonnegative for all $x \in C$ if and only if $\frac{\partial f(x^*)}{\partial x_i} \leq 0$ for all $i \in B$. Finally, the sum over C is nonnegative for all $x \in C$ if and only if $\frac{\partial f(x^*)}{\partial x_i} = 0$ for all $i \in C$.

In order to solve this numerically, we can use the Frank-Wolfe method.

Exercise 6.4. Consider the optimization problem minimize $f(x)$ subject to $x \geq 0$, where $f : \mathbb{R}^n \to \mathbb{R}$ is a differentiable convex function. Show that the KKT conditions for this problem are

$$\begin{aligned} x &\geq 0, \\ \nabla f(x) &\geq 0, \\ x_k \cdot \frac{\partial f(x)}{\partial x_k} &= 0, \quad k = 1, \ldots, n. \end{aligned} \tag{6.9}$$

Discuss the consequences of these conditions for optimal solutions.

Solution 6.4. The problem can be written as

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & g_j(x) \leq 0 \quad j = 1, \ldots, n, \end{array} \tag{6.10}$$

where $g_j(x) = -x_j$. This gives the Lagrangian

$$\mathcal{L}(x, \mu) = f(x) + \sum_{j=1}^{n} \mu_j g_j(x) = f(x) - \sum_{j=1}^{n} \mu_j x_j. \tag{6.11}$$

The KKT conditions are then

$$\begin{aligned} \nabla_x \mathcal{L}(x^*, \mu^*) &= \nabla f(x^*) - \mu^* = 0, \\ \mu_j^* &\geq 0, \quad j = 1, \ldots, n, \\ \mu_j^* &= 0, \quad j \in J(x^*), \end{aligned} \tag{6.12}$$

for a feasible point x^*, where $J(x^*) = \{j \leq n : g_j(x^*) = x^* = 0\}$ is the set of active constraints at x^*.

If $j \notin J(x^*)$, then $x_j^* > 0$ and thus $\mu_j^* = 0$ by complementary slackness, giving $\frac{\partial f(x^*)}{\partial x_j} = 0$. If $j \in J(x^*)$, then $x_j^* = 0$ and $\mu_j^* \geq 0$, giving $\frac{\partial f(x^*)}{\partial x_j} \geq 0$. In both cases we have that $x_j^* \cdot \frac{\partial f(x^*)}{\partial x_j} = 0$, $\frac{\partial f(x)}{\partial x_j} \geq 0$ and $x^* \geq 0$ by feasibility of x^*.

Exercise 6.5. Solve the problems

$$\begin{aligned} \text{minimize} \quad & (x + 2y - 3)^2 \\ \text{subject to} \quad & (x, y) \in \mathbb{R}^2, \end{aligned} \tag{6.13}$$

and

$$\begin{aligned} \text{minimize} \quad & (x + 2y - 3)^2 \\ \text{subject to} \quad & (x - 2)^2 + (y - 1)^2 \leq 1. \end{aligned} \tag{6.14}$$

Solution 6.5. The first problem is unconstrained, and we find

$$\nabla f(x, y) = \begin{pmatrix} 2(x + 2y - 3) \\ 4(x + 2y - 3) \end{pmatrix}, \tag{6.15}$$

which is zero when $x + 2y - 3 = 0$. Thus, the optimal solutions are all points on the line $x + 2y = 3$.

For the second problem, we get the gradient equation

$$\nabla f(x, y) + \mu \nabla g(x, y) = \begin{pmatrix} 2(x + 2y - 3) \\ 4(x + 2y - 3) \end{pmatrix} - \mu \begin{pmatrix} 2(x - 2) \\ 2(y - 1) \end{pmatrix} = 0. \tag{6.16}$$

The gradient of the constraint is zero at $(2, 1)$, which gives the candidate solution $f(2, 1) = 1$. Alternatively, we see by close inspection that $(1, 1)$ also satisfies the constraint with $f(1, 1) = 0$, the global minimum we found previously.

Exercise 6.6. Solve the problem

$$\begin{aligned} \text{minimize} \quad & x^2 + y^2 - 14x - 6y \\ \text{subject to} \quad & x + y \leq 2 \\ & x + 2y \leq 3. \end{aligned} \tag{6.17}$$

Solution 6.6. We consider first the unconstrained problem. Here, we have

$$\nabla f(x,y) = \begin{pmatrix} 2x - 14 \\ 2y - 6 \end{pmatrix}, \tag{6.18}$$

which is zero at $(7,3)$, with $f(7,3) = -58$. This point is not feasible, so we must consider the constraints.

We then get the gradient equation

$$\begin{pmatrix} 2x - 14 \\ 2y - 6 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \mu_2 \begin{pmatrix} 1 \\ 2 \end{pmatrix} = 0. \tag{6.19}$$

With both constraints active, we have

$$\begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \tag{6.20}$$

giving $(x,y) = (1,1)$ with $f(1,1) = -18$.

If only the first is active, we have $\nabla f(x,y) - \mu_1(1,1)^T = 0$ and $x + y = 2$, giving

$$\begin{bmatrix} 1 & 1 \\ 2 & -2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 2 \\ 8 \end{pmatrix}, \tag{6.21}$$

which yields

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{4} \begin{bmatrix} 2 & 1 \\ 2 & -1 \end{bmatrix} \begin{pmatrix} 2 \\ 8 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} 12 \\ 4 \end{pmatrix} = \begin{pmatrix} 3 \\ -1 \end{pmatrix}, \tag{6.22}$$

which gives $f(3,-1) = -26$.

If only the second is active, we have $\nabla f(x,y) - \mu_2(1,2)^T = 0$ and $x + 2y = 3$, giving

$$\begin{bmatrix} 1 & 2 \\ 4 & -2 \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 3 \\ 22 \end{pmatrix}, \tag{6.23}$$

which yields

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{1}{10} \begin{bmatrix} 2 & 2 \\ 4 & -1 \end{bmatrix} \begin{pmatrix} 3 \\ 22 \end{pmatrix} = \frac{1}{10} \begin{pmatrix} 50 \\ -10 \end{pmatrix} = \begin{pmatrix} 5 \\ -1 \end{pmatrix}, \tag{6.24}$$

which gives $f(5,-1) = -38$. This however violates the first constraint, so it is not feasible.

The optimal solution is thus $(3,-1)$ with objective value -26.

Exercise 6.7. Solve the problem

$$\begin{array}{rrl} \text{minimize} & x^2 - y & \\ \text{subject to} & y - x & \geq -2 \\ & y^2 & \leq x \\ & y & \geq 0. \end{array} \tag{6.25}$$

Solution 6.7. Considering the unconstrained problem firstly, have

$$\nabla f(x, y) = \begin{pmatrix} 2x \\ -1 \end{pmatrix}, \tag{6.26}$$

which is never zero.

Let's then consider the constraints closer. We have

$$\begin{aligned} x &\leq y + 2, \\ y^2 &\leq x, \\ 0 &\leq y. \end{aligned} \tag{6.27}$$

Combining the first two, we see that we must have $y^2 \leq y + 2$, or equivalently $y^2 - y - 2 \leq 0$. This factors as $(y-2)(y+1) \leq 0$, so we have $-1 \leq y \leq 2$. Together with the third constraint, we have $0 \leq y \leq 2$. We wish to choose the smallest possible x, so we set $x = y^2$. The problem then becomes

$$\begin{array}{rc} \text{minimize} & y^4 - y \\ \text{subject to} & 0 \leq y \leq 2. \end{array} \tag{6.28}$$

The derivative is $g'(y) = 4y^3 - 1$, which is zero at $y = 4^{-1/3}$. As $g''(y) = 12y^2 \geq 0$, this is a minimum. This yields the optimal solution $(x, y) = (4^{-2/3}, 4^{-1/3})$ with objective value

$$f(4^{-2/3}, 4^{-1/3}) = 4^{-4/3} - 4^{-1/3} = 4^{-1/3}(4^{-1} - 1) = -\frac{3}{4}4^{-1/3}. \tag{6.29}$$

7 Notes on Combinatorial Optimization

Exercise 7.1. Consider the knapsack problem

$$\max\left\{\sum_{j=1}^{n} c_j x_j : \sum_{j=1}^{n} a_j x_j \leq b, 0 \leq x \leq 1\right\} \tag{7.1}$$

where all the data are positive integers. Define the knapsack relaxation polytope by

$$P = \{x \in \mathbb{R}^n : \sum_{j=1}^{n} a_j x_j \leq b, 0 \leq x \leq 1\}. \tag{7.2}$$

Assume that the variables have been ordered such that $c_1/a_1 \geq c_2/a_2 \geq \cdots \geq c_n/a_n$. Try to guess an optimal solution and prove the optimality by considering the dual problem. Use this result to characterize all the vertices of P. What about the cover inequalities in relation to these vertices?

Solution 7.1.