SOLUTIONS TO EXERCISES IN AN INTRODUCTION TO CONVEXITY

ØYVIND RYAN *

JUNE 2020

* University of Oslo, Centre of Mathematics for ApplicatiP.O.Box 1053, Blindern, 0316 Oslo, Norway (geird@math.uio.no)

Contents

Chapter 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* $x_1 + y_1 \leq x_2 + y_2$ *also holds. Let now* $\lambda\hat{A}$ *be a nonnegative real number. Explain why* $\lambda x_1 \leq \lambda x_2$ *holds. What happens if* λ*is negative?*

Solution: We have that $x_1 + y_1 \leq x_2 + y_1 \leq x_2 + y_2$, since adding the same number to the both sides does not alter the inequality. If $\lambda < 0$ we obtain that $\lambda x_1 \geq \lambda x_2$, since multiplication with a negative number changes the direction of the inequality.

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: The exercise adds to the observation that $\mathbb{R}^n_+$ is closed under addition and multiplication with positive scalars, that these two operations also respect the ordering in $\mathbb{R}^n_+$ defined by $x \leq y \iff x_i \leq y_i$ for all i.

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 nonnegative here?*

Solution: If $a \in \mathbb{R}^n_+$, since $x_i \leq y_i$ we have that $a_i x_i \leq a_i y_i$. We obtain that

$$a^T x = \sum_{i=1}^n a_i x_i \leq \sum_{i=1}^n a_i y_i = a^T y,$$

and the result follows. If a has some negative components, the corresponding inequality is reversed. If all entries of a are nonpositive, the sum inequality above is also reversed.

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: Assume that $x \in B(a, r)$, $y \in B(a, r)$, so that $\|x - a\| \leq r$, $y - a\| \leq r$. We have that

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

so that also $(1-\lambda)x + \lambda y \in B(a,r)$. It follows that $B(a,r)$ is 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: $Ax \leq b$ is equivalent to solving $Ax + z = b$, $z \geq 0$ for x and z. Now, write $x = x_1 - x_2$ where $x_1, x_2 \geq 0$. We have the system

$$A(x_1 - x_2) + z = Ax_1 - Ax_2 + z = \begin{pmatrix} A & -A & I \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ z \end{pmatrix} = b,$$

so that A is replaced with $\begin{pmatrix} A & -A & I \end{pmatrix}$.

Exercise 1.6. *Make a drawing of the standard simplices* S_1, $S_2\hat{A}$ *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* $\lambda_j \hat{A}$ *is nonnegative and* $\sum_{j=1}^n \lambda_j = 1$. *How? Can this be done in several ways?*

Solution: e_j lies in S_n since its components sum to one. In $x = \sum_{j=1}^n \lambda_j e_j$, we simply have that $\lambda_j = x_j$ (the j'th component of x). This decomposition is unique, due to linear independence of the standard basis.

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

Solution: Let C be a convex cone, and let $x_1 \in C$, $x_2 \in C$. Then $(1-\lambda)x_1 + \lambda x_2 \in C$ for $0 \leq \lambda \leq 1$, since $\lambda, 1-\lambda \geq 0$. It follows that C also is a convex set.

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

Solution: Let $x_1, x_2 \in C$, and $\lambda_1, \lambda_2 \geq 0$. Then we have that

$$A(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 A x_1 + \lambda_2 A x_2 \leq 0$$

since $Ax_i \leq 0$. It follows that $\lambda_1 x_1 + \lambda_2 x_2$, so that C is a convex cone.

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

Solution: Let $y, z \in C(x_1, \ldots, x_t)$, so that

$$y = a_1 x_1 + \ldots + a_t x_t \qquad\qquad z = b_1 x_1 + \ldots + b_t x_t,$$

where $a_1, b_1, \ldots, a_t, b_t \geq 0$. Let $\lambda_1, \lambda_2 \geq 0$. We have that

$$\lambda_1 y + \lambda_2 z = (\lambda_1 a_1 + \lambda_2 b_1) x_1 + \ldots + (\lambda_1 a_t + \lambda_2 b_t) x_t,$$

where all $\lambda_1 a_i + \lambda_2 b_i \geq 0$ (since all variables are ≥ 0). It follows that $\lambda_1 y + \lambda_2 z \in C(x_1, \ldots, x_t)$, so that $C(x_1, \ldots, x_t)$ 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 $S\hat{A}$ a finitely generated cone?*

Solution: Assume that $a_1 = (x_1, y_1, z_1)$, $a_2 = (x_2, y_2, z_2)$ are both in S, so that $z_1 \geq x_1^2 + y_1^2$, $z_2 \geq x_2^2 + y_2^2$. Convexity of S is the same as showing, for $0 \leq \lambda \leq 1$, that

$$(1-\lambda)a_1 + \lambda a_2 = ((1-\lambda)x_1 + \lambda x_2, (1-\lambda)y_1 + \lambda y_2, (1-\lambda)z_1 + \lambda z_2) \in S,$$

i.e., that

$$((1-\lambda)x_1 + \lambda x_2)^2 + ((1-\lambda)y_1 + \lambda y_2)^2 \leq (1-\lambda)z_1 + \lambda z_2.$$

We have that

$$((1-\lambda)x_1 + \lambda x_2)^2 \leq (1-\lambda)x_1^2 + \lambda x_2^2$$

(since this can be reorganized to $\lambda(1-\lambda)(x_1^2 + x_2^2 - 2x_1x_2 = \lambda(1-\lambda)(x_1 - x_2)^2 \geq 0$. Convexity of the function $f(x) = x^2$ is really what is at play here. We will return to this later), and similarly

$$((1-\lambda)y_1 + \lambda y_2)^2 \leq (1-\lambda)y_1^2 + \lambda y_2^2.$$

It follows that

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

and the result follows.

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

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

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

Solution: Since $c^T x = \sum_{i=1}^n c_i x_i$, we see that the maximum can be obtained by maximizing each $c_i x_i$ separately. If $c_i \geq 0$ this maximum is 0. It follows that the maximum is $\sum_{i|c_i>0} c_i$.

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

Solution: No. Take for instance the union of the two intervals $(-\infty, 1)$ and $(1, \infty)$.

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

(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) $A = [0,1] \times [1,2]$, $B = [0,2] \times [0,2]$.

Solution:

(i): The disk with center $(3,4)$ and radius 1.

(ii): The left half of the disk with center $(0,0)$ and radius 1, combined with the rectangle with corners $(0,1),(0,-1),(1,1),(1,-1)$, combined with the disk with center $(1,0)$ and radius 1.

(iii): We can write $A = \begin{pmatrix}5\\0\end{pmatrix} + y\begin{pmatrix}-2\\1\end{pmatrix}$, and $B = x\begin{pmatrix}1\\1\end{pmatrix}$, $0 \leq x \leq 1$. If we set $x=1$ we obtain the line

$$\begin{pmatrix}5\\0\end{pmatrix} + y\begin{pmatrix}-2\\1\end{pmatrix} + \begin{pmatrix}1\\1\end{pmatrix} = \begin{pmatrix}6\\1\end{pmatrix} + y\begin{pmatrix}-2\\1\end{pmatrix}$$

It follows that $A+B$ is the area between the parallel lines

$$\begin{pmatrix}5\\0\end{pmatrix} + y\begin{pmatrix}-2\\1\end{pmatrix} \text{ and } \begin{pmatrix}6\\1\end{pmatrix} + y\begin{pmatrix}-2\\1\end{pmatrix}.$$

(iv): The rectangle with vertices $(0,1),(0,4),(3,1),(3,4)$.

Exercise 1.14. *(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:

(i): A general element in $A+B$ is on the form $a+b$ for $a \in A$, $b \in B$. Since $\lambda(a+b) = \lambda a + \lambda b \in\in \lambda A + \lambda B$, it follows that $\lambda(A+B) \subseteq \lambda A + \lambda B$. The other way follows in the same way.

(ii): If $\lambda = -\mu$ the set on the left consists of only the origin, while the set on the right can be any set.

(iii): Let A be convex, and let $a_1, a_2 \in A$. Then

$$\lambda a_1 + \mu a_2 = (\lambda+\mu)\left(\frac{\lambda}{\lambda+\mu}a_1 + \frac{\mu}{\lambda+\mu}a_2\right) \in (\lambda+\mu)A$$

since A is convex. It follows that $\lambda A + \mu B \subseteq (\lambda + \mu)A$. On the other hand, if $a \in A$,

$$(\lambda + \mu)a = \lambda a + \mu a \in \lambda A + \mu A,$$

so that $(\lambda + \mu)A \subseteq \lambda A + \mu A$. It follows that $(\lambda + \mu)A = \lambda A + \mu A$.

Exercise 1.15. *Show that if $C_1, \ldots, C_t \subseteq \mathbb{R}^n\hat{A}$ are all convex sets, then $C_1 \cap \ldots \cap C_t\hat{A}$ 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: Assume that $x, y \in C_1 \cap \ldots \cap C_t$. Then, in particular $x, y \in C_i$. Since C_i is convex, $(1-\lambda)x + \lambda y \in C_i$. But then also $(1-\lambda)x + \lambda y \in C_1 \cap \ldots \cap C_t$, so that $C_1 \cap \ldots \cap C_t$ is also convex. Affine sets, linear subspaces and convex cones are all convex, so the result for them also follows. Note that the proof applied for any number of sets, regardless of cardinality, so the result also holds for any number of sets.

Exercise 1.16. *Consider a family (possibly infinite) of linear inequalities $a_i^T x \leq b_i$, $i \in I$, 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: Each set $a_i^T x \leq b_i$ is convex (in fact affine). C is the intersection of all these sets, so it is also convex, by the previous exercise.

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: For any $a \in S$, use the linear inequality $a^T x \leq 1$. This describes a plane with normal vector 1 which supports the unit disc at a. The intersection of all these half planes thus contains the entire unit disc. If x is a point outside the unit disc, then clearly the half plane obtained by setting $a = x/\|x\|$ does not contain x, so that the intersection is exactly S.

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

Solution: Assume that B equals the intersection of n sets on the form $a_i^T x \leq b_i$. Clearly we can assume that none of the a_i are parallel, that all $\|a_i\|_2 = 1$, and all $b_i = 1$. The plane $a_i^T x = 1$ is a tangent plane to the unit ball at a_i, and B is contained in $\{x : a_i^T x \leq b_i\}$.

Let x be so that $\|x\|_2 = 1$, and so that $x \neq a_i$ for all i. Clearly $a_i^T x < b_i$ for all i, since for each i $x = a_i$ is the unique point in B so that equality holds. By continuity there exists a point $x' \notin B$ so that $a_i^T x' < b_i$ for all i. Thus, the intersection of all those half planes must contain more than B. It follows that B is not a polyhedron.

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

Solution: It is straightforward to show that $\|\cdot\|_\infty$ is a norm. Convexity thus follows from Exercise 1.4, since the proof therein applies for any norm. That B_∞ is a polyhedron follows from writing $\|x\|_\infty \leq 1$ equivalently as $\{x_i \leq 1, -x_i \leq 1\}_{i=1}^n$. For $n = 2$ we obtain the square with vertices $(1,1), (-1,1), (1,-1), (-1,-1)$.

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

Solution: It is straightforward to show that $\|\cdot\|_1$ is a norm. Convexity thus follows as above. That B_1 is a polyhedron follows from writing $\|x\|_1 = \sum_{j=1}^n |x_j| \leq 1$ equivalently as $\sum_{j=1}^n \pm x_j \leq 1$, where the signs traverse all possible 2^n combinations. For $n = 2$ there are four possible sign choices, leadin to the polyhedron defined by $x + y \leq 1$, $x - y \leq 1$, $-x + y \leq 1$, $-x - y \leq 1$. This gives the square with vertices $(1,0), (0,1), (-1,0), (0,-1)$.

Exercise 1.21. *Prove Proposition* 1.5.1*.*

Solution: Let $x_0 \in C$ (C is assumed nonempty), and let $L = C - x_0 = \{x - x_0 : x \in C\}$ (so that $C = L + x_0$). Assume that C is affine. For $x \in C$ we have that

$$\lambda(x - x_0) = \lambda x + (1-\lambda)x_0 - x_0 \in L,$$

since $\lambda x + (1-\lambda)x_0 \in C$. If $x_1, x_2 \in C$ and $\lambda \in \mathbb{R}$, it follows that we can write $x_1 - x_0 = (1-\lambda)(x_3 - x_0)$ and $x_2 = \lambda(x_4 - x_0)$ for some $x_3, x_4 \in C$. Thus

$$x_1 - x_0 + x_2 - x_0 = (1-\lambda)(x_3 - x_0) + \lambda(x_4 - x_0) = (1-\lambda)x_3 + \lambda x_4 - x_0 \in L,$$

since $(1-\lambda)x_3 + \lambda x_4 \in C$. It follows that L is a vector space. From

$$C - x_1 = (C - x_0) - (x_1 - x_0)$$

it follows that $C - x_1 \subseteq C - x_0 = L$. It follows that this vector space is uniquely defined (i.e., independent of the choice of x_0).

Let now A be a matrix, and let C be the solution set of $Ax = b$. If $Ax_1 = b$, $Ax_2 = b$, we get

$$A((1-\lambda)x_1 + \lambda x_2) = (1-\lambda)Ax_1 + \lambda Ax_2 = (1-\lambda)b + \lambda b = b,$$

so that C is affine. The other way, if L is a linear subspace of $\mathbb{R}^n$, and $x_0 \in \mathbb{R}^n$, we can find a matrix A with null space L. If $Ax_0 = b$, the solution set of $Ax = b$ is then the affine set $C = L + x_0$.

Exercise 1.22. *Let C Â be a nonempty 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: From the previous exercise we can write $C = L + x_0$ (so that $L = C - x_0$) for a unique subspace L (it does not depend on the choice of $x_0 \in C$).

Let $x_1, x_2 \in C$, so that $x_1 = l_1 + x_0$, $x_2 = l_2 + x_0$ for some $l_1, l_2 \in L$. We get that

$$x_1 - x_2 = l_1 - l_2 \in L,$$

so that $C - C \subseteq L$. The other way, if $l \in L$, we can write $l = x - x_0$ for some $x \in C$, so that $l \in C - C$. Thus $L \subseteq C - C$, so that $L = C - C$.

Chapter 2

Convex hulls and Carathéodory's theorem

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

Exercise 2.2. *Choose your favorite 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* $x_1, x_2\hat{A}$ *and* x_3*? If so, how?*

Solution: Take for instance the three points $x_1 = (0, 1)$, $x_2 = (1, 1)$, $x_3 = (1, 0)$. The convex combinations of two points lie on the line between those points. Convex combinations of $(0, 1)$ and $(1, 1)$, as well as convex combinations of $(1, 0)$ and $(1, 1)$, and convex combinations of $(0, 1)$ and $(1, 0)$, constitute the three edges which together form the boundary. Clearly, the interior points lie on a line through x_1 and a point on the edge connecting x_2 and x_3. The latter can be written as $(1 - \lambda)x_2 + \lambda x_3$. An interior point is thus a convex combination of this and x_1, which can be written as

$$(1 - \mu)x_1 + \mu\left((1 - \lambda)x_2 + \lambda x_3\right) = (1 - \mu)x_1 + \mu(1 - \lambda)x_2 + \mu\lambda x_3.$$

This is a convex combination of the three points, since the coefficients sum to $1 - \mu + \mu(1 - \lambda + \lambda) = 1$.

Exercise 2.3. *Show that* $\text{conv}(S)$ *is convex for all* $S \subseteq \mathbb{R}^n$*. (Hint: look at two convex combinations* $\sum_j \lambda_j x_j \hat{A}$ *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: Let $x = \sum_{j=1}^{t} \lambda_j x_j \hat{\text{A}}$ and $y = \sum_{j=1}^{s} \mu_j y_j$ be convex combinations of points in S. By taking the union of the points x_j and y_j, both x and y can be written as convex combinations of the same set of points $\{z_j\}_{j=1}^{r}$ (some of the coefficients may now be zero). We obtain

$$(1-\lambda)x + \lambda y = (1-\lambda)\left(\sum_{j=1}^{r}\lambda_j z_j\right) + \lambda\left(\sum_{j=1}^{r}\mu_j z_j\right)$$
$$= \sum_{j=1}^{r}((1-\lambda)\lambda_j + \lambda\mu_j)z_j.$$

Clearly the coefficients $(1-\lambda)\lambda_j + \lambda\mu_j$ sum to one, so that this is also a convex combination of points in S. It follows that $\text{conv}(S)$ is 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 $S_0\hat{A}$ of a set S such that $S = \text{conv}(S_0)$. We study this question later.*

Solution: We can set $S = \{-1, 0, 1\}$, and $T = \{-1, 1\}$. Both have $[-1, 1]$ as convex hull.

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

Solution: This follows from the fact that a convex combination of points in S then also is a convex combination of points in T.

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

Solution: This is a compulsory exercise.

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

Solution: $\text{conv}(S)$ must contain any point inside the unit circle. This can be seen if you take any line through this point. This line will intersect the unit circle at two points, so that the point is a convex combination of two points on the unit circle. Therefore $D \subseteq \text{conv}(S)$, where $D = \{x \in \mathbb{R}^2 : \|x\|_2 \leq 1\}$. Since $\text{conv}(S)$ is the smallest convex set that contains S and since D is convex and contains S, we obtain that $\text{conv}(S) \subseteq D$. It follows that $\text{conv}(S) = D$.

If $x \in \mathbb{R}^2 \neq 0$ then $x = \|x\|_2 u$ with $u = x/\|x\|_2 \in S$. It follows that $u \in \text{cone}(S)$, so that $\text{conv}(S) = \mathbb{R}^2$.

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

Solution: Linear independence (of the columns of A) is the same as

$$Ax = 0 \Rightarrow x = 0.$$

Affine independence (of the columns of A) is the same as

$$\begin{pmatrix} A \\ 1 \cdots 1 \end{pmatrix} x = 0 \Rightarrow x = 0,$$

i.e., it is the same as linear independence of the columns of A with a last component with a one added. Clearly then linear independence implies affine independence (since equality in the first n components already implies that the coefficients must be zero). Affine idependence does not imply linear independence, however: If A has n rows, $\begin{pmatrix} A \\ 1\cdots 1 \end{pmatrix}$ can have rank $n+1$, but A can have rank at most n. In particular, there can be $n+1$ linearly independent column vectors in $\begin{pmatrix} A \\ 1\cdots 1 \end{pmatrix}$, but only n in A.

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:

$$\sum_i \lambda_i(x_i + w) = 0 \text{ and } \sum_i \lambda_i = 0$$

is equivalent to

$$\sum_i \lambda_i x_i = 0 \text{ and } \sum_i \lambda_i = 0$$

since $\sum_i \lambda_i w = 0$ when $\sum_i \lambda_i = 0$. The result follows.

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

Solution: Let $x_1, \ldots, x_t$ be a basis for L. $\{0, x_1, \ldots, x_t\}$ are affinely independent since $x_1 - 0, \ldots, x_t - 0$ are linearly independent. Therefore the affine dimension of L is $\geq t$. If the affine dimension of L was larger than t, we could find at least $t+2$ affinely independent points $x_1, \ldots, x_{t+2}$ in L, so that $x_2 - x_1, \ldots, x_{t+2} - x_1$ are linearly independent. There are $t+1$ vectors here, all in L, so that the dimension of L is $\geq t+1$. This is a contradiction. It follows that the affine dimension equals the dimension.

Exercise 2.11. *Prove the last statements in the previous paragraph.*

Solution: With A being the set of all $\sum_{j=1}^t \lambda_j x_i$ with $\sum_{j=1}^t \lambda_j = 1$, we have that

$$(1-\lambda)\sum_j \lambda_j x_j + \lambda \sum_j \mu_j x_j = \sum_j (1-\lambda)\lambda_j + \lambda\mu_j)x_j \in A,$$

since the coefficients sum to one (by expansion we can clearly assume the same base set x_j for the two linear combinations). It follows that A is affine. Choosing $\lambda_j = 0$, the other zero, we see that $x_1, \ldots, x_{d+1}$ are in A. Then A must in particular contain all the convex combinations of these points. Since $\text{conv}(C) = C$, A contains C. Any affine set that contains C must also contain these, so this is the smallest one.

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

Solution: An example is the half-open interval $A = [0, 1)$.

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: If $x^k \to x$ we can find an N so that $\|x^k - x\|_2 \leq \epsilon$ for $k \geq N$. But since $|x_j^k - x_j| \leq \|x^r - x\|_2$, it follows that also $|x_j^k - x_j| \leq \epsilon$ for $r \geq N$, so that $x_j^k \to x_j$ for $j = 1, \ldots, n$.

The other way, if $x_j^k \to x_j$ for $j = 1, \ldots, n$, we can find an N so that $|x_j^k - x_j| \leq \epsilon/\sqrt{n}$ for all $k \geq N$ and $j = 1, \ldots, n$. But then, for $k \geq N$,

$$\|x^k - x\|_2 = \sqrt{\sum_{j=1}^{n} |x_j^k - x_j|^2} \leq \sqrt{\sum_{j=1}^{n} \epsilon^2/n} = \epsilon,$$

so that $x^k \to x$. The result follows.

Exercise 2.14. *Show that every simplex cone is closed.*

Solution: Let X be the matrix with columns x_j (which define the simplex cone). x is in the simplex cone if and only if it can be written on the form $\sum_{j=1}^{k} \lambda_j x_j$ for some $\lambda_j \geq 0$, i.e., $x = X\lambda$ (λ is the column vector with components λ_i).

Since the x_j are linearly independent, X has linearly independent columns. Assume that $X^T X x = 0$. This implies that Xx is in the null space of X^T, which is orthogonal to the range of X. This implies that $Xx = 0$, so that $x = 0$ by linear independence of the columns of X. It follows that $X^T X$ is invertible, so that we can define $X^\dagger := (X^T X)^{-1} X^T$.

Myltiplying $x = X\lambda$ with $X^\dagger$ on each side we obtain that $\lambda = X^\dagger x$. The simplex cone can now be written as the inverse image of the closed set $\lambda : \lambda_j \geq 0\}$ under the continuous map $f(x) = X^\dagger x$. It follows that the simplex cone is closed.

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: If $x \in bd(S)$ is the same as $x \in \mathrm{cl}(S)$ and $x \notin \mathrm{int}(S)$. $x \notin \mathrm{int}(S)$ is the same as any ball with center x intersects the complement of S. $x \in \mathrm{cl}(S)$ is the same as any ball intersects S. The result follows.

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) $\mathrm{int}(C) = \emptyset$,

(iv) $\text{bd}(C) = C$,
(v) $\text{rint}(C) = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\}$ *and*
(vi) $\text{rbd}(C) = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 = 1\}$.

Solution:

(i): Assume that $x^k \to x$, and that $x^k \in C$. Since all $x_k^2 + y_k^2 \leq 1$, then also $x^2 + y^2 \leq 1$. Since all $z_k = 0$, then also $z = 0$. It follows that $x \in C$ also. It follows that C is closed.

(ii): It is possible to find only 3 affinely independent points in C, sinde the third component is always zero. The dimension is thus ≤ 2. To see that the dimension is exactly 3, choose for instance the three points $(0, 0, 0)$, $(1, 0, 0)$, and $(0, 1, 0)$.

(iii): Any ball around a point in C will contain points with nonzero third component, so that no ball can be entirely contained in C. It follows that $\text{int}(C) = \emptyset$.

(iv): $\text{bd}(C) = \text{cl}(C) \setminus \text{int}(C) = C \setminus \emptyset = C$.

(v): Clearly $\text{aff}(C)$ is the entire xy-plane.

If $x_1^2 + x_2^2 < 1$, then in $\mathbb{R}^2$ we can find a ball $B^o((x_1, x_2), r)$ contained in $B^o((0, 0), 1)$. We then obtain

$$\begin{aligned} B^o(x, r) \cap \text{aff}(C) &= \{(z_1, z_2, 0) \in \mathbb{R}^3 : (z_1 - x_1)^2 + (z_2 - x_2^2) < r\} \\ &= (B^o((x_1, x_2, r), 0) \subseteq C, \end{aligned}$$

This shows that $\{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 < 1\} \subseteq \text{rint}(C)$.

If $x_1^2 + x_2 = 1$, any ball around such a point contains points suct that $z_1^2 + z_2^2 > 1$, so that for such points we can't have $B^o(x, r) \cap \text{aff}(C) \subseteq C$.

(vi):

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

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 $\|\lambda x\| = \lambda \|x\|$ when $\lambda \geq 0$).*

Solution: Let $C = \text{conv}(x_1, \ldots, x_t)$, ann let $x = \sum_{j=1}^t \lambda_j x_j$ with the λ_j nonnegative and summing to 1. Using the triangle inequality we obtain

$$\|x\| = \|\sum_{j=1}^t \lambda_j x_j\| \leq \sum_{j=1}^t \lambda_j \|x_j\| \leq \max_{j=1}^t \|x_j\| \sum_{j=1}^t \lambda_j = \max_{j=1}^t \|x_j\|.$$

This proves that the polytope is bounded.

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

Solution: Since a simplex is in particular polytope, it is bounded because of the previous exercise. It is closed since it is the inverse of the closed set $\{1\}$ under the mapping $f(x) = x_1 + \ldots + x_n$ restricted to the first "quadrant".

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

Solution:

$$S = \{(0,0)\} \cup \{x \in \mathbb{R}^2 : x_1, x_2 > 0\}.$$

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

Solution: See exercise 2.3.

Exercise 2.21. *Prove the second statement of Proposition* 2.1.1.

Solution: Assume that C is a convex cone. By definition and induction it must contain all nonnegative combinations of its points.

Assume that C contains all nonnegative combinations of its points. That it is a convex cone follows by restricting this to any two points.

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

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: A convex combination $\lambda_1 a_1 + \lambda_2 a_2 + \lambda_3 a_3$ can be written as (λ_2, λ_1), and the only requiremnet on the λ_1, λ_2 is that they are nonnegative and summing to something less than 1. This shows that the stated set coincides with the set of convex combinations.

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* $\text{conv}(S) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_i \leq 1 \text{ for } i = 1,2,3\}$. *Also determine* $\text{conv}(S \setminus \{(1,1,1)\}$ *as the solution set of a system of linear inequalities. Illustrate all these cases geometrically.*

Solution: Any convex combination of values between 0 and 1 lies between 0 and 1. Since the coordinates of all the points lie between 0 and 1, $\text{conv}(S)$ must consist of points with coordinates between 0 and 1 as well, i.e., $\text{conv}(S) \subseteq \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_i \leq 1 \text{ for } i = 1,2,3\}$. The other way, since the points constitute all vertices of the unit cube, $\text{conv}(S)$ will contain all the edges of the unit cube (by taking convex combinations of adjacent vertices). By taking convex combinations of edges which on different sides of a face, we see that all faces of the cube are also contained in $\text{conv}(S)$. By taking convex combinations of different faces we obtain the entire cube, so that $\{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_i \leq 1 \text{ for } i = 1,2,3\} \subseteq \text{conv}(S)$. Equality now follows.

Now, let us exclude the point $(1,1,1)$. In all remaining seven points the coordinates sum to ≤ 2, so that this applies to $\text{conv}(S \setminus (1,1,1))$ as well, which is thus contained in $x_1 + x_2 + x_3 \leq 2$. Consider the polyhedron described by $0 \leq x_1, x_2, x_3 \leq 1$, and $x_1 + x_2 + x_3 \leq 2$. It is easily verified that the vertices of this set is $S \setminus (1,1,1)$. Since the polyhedron is bounded, and since the vertcies equal the extreme points (see Chapter 4), it follows that $\text{conv}(S \setminus (1,1,1))$ equals this polyhedron.

Exercise 2.25. *Let $A, B \subseteq \mathbb{R}^n$. Prove that* $\text{conv}(A + B) = \text{conv}(A) + \text{conv}(B)$. *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 $\lambda_j \geq 0$, $\mu_k \geq 0\hat{A}$ and $\sum_j \lambda_j = 1$ and $\sum_k \mu_k = 1$.*

Solution: We have that

$$\sum_{j,k} \lambda_j \mu_k (a_j + b_k) = (\sum_j \lambda_j)(\sum_k \mu_k b_k) + (\sum_k \mu_k)(\sum_j \lambda_j a_j) = \sum_k \mu_k b_k + \sum_j \lambda_j a_j.$$

The right hand side is a general element in $\text{conv}(A) + \text{conv}(B)$. Since the left hand side is a convex combination of elements in $A+B$, it follows that $\text{conv}(A) + \text{conv}(B) \subseteq \text{conv}(A + B)$. The other way, $\text{conv}(A) + \text{conv}(B)$ is clearly convex (the sum of two convex sets is always convex), so that it equals its convex hull. Therefore

$$\text{conv}(A + B) \subseteq \text{conv}(\text{conv}(A) + \text{conv}(B)) = \text{conv}(A) + \text{conv}(B),$$

and the result follows.

Exercise 2.26. *When $S \subset \mathbb{R}^n\hat{A}$ is a finite set, say $S = \{x_1, \ldots, x_t\}$, then we have*

$$\text{conv}(S) = \{\sum_{j=1}^{t} \lambda_j x_j : \lambda_j \geq 0 \text{ for each } j, \sum_j \lambda_j = 1\}.$$

Thus, every point in $\text{conv}(S)$ *is a convex combination of the points $x_1, \ldots, x_t$. What happens if, instead, $S\hat{A}$ has an infinite number of elements? Then it may not be possible to give a fixed, finite subset S_0 of $S\hat{A}$ such that every point in* $\text{conv}(S)$ *is a convex combination of elements in S_0. Give an example which illustrates this.*

Solution: Let S be the integers. Clearly $\text{conv}(S) = \mathbb{R}$, but, for any finite subset S_0, $\text{conv}(S_0) = [\min(S_0), \max(S_0)]$. This means that we can't use any finite subset to describe $\text{conv}(S)$.

Exercise 2.27. *Let $x_0 \in \mathbb{R}^n\hat{A}$ and let $C \subseteq \mathbb{R}^n$ be a convex set. Show that*

$$\text{conv}(C \cup \{x_0\}) = \{(1 - \lambda)x_0 + \lambda x : x \in C, \lambda \in [0,1]\}.$$

Solution: By definition it is clear that

$$\{(1-\lambda)x_0 + \lambda x : x \in C, \lambda \in [0,1]\} \subseteq \text{conv}(C \cup \{x_0\}).$$

The other way, a general element in $\text{conv}(C \cup \{x_0\})$ can be written as

$$(1-\lambda)x_0 + \sum_{i=1}^{n} \lambda_i x_i,$$

where $x_i \in C$ and $\sum_{i=1}^n \lambda_i = \lambda$. We rewrite this as

$$(1-\lambda)x_0 + \lambda \sum_{i=1}^{n} \frac{\lambda_i}{\lambda} x_i = (1-\lambda)x_0 + \lambda x,$$

where $x = \sum_{i=1}^n \frac{\lambda_i}{\lambda} x_i \in C$ since C is convex. It follows that

$$\text{conv}(C \cup \{x_0\}) \subseteq \{(1-\lambda)x_0 + \lambda x : x \in C, \lambda \in [0,1]\}.$$

The result follows.

Exercise 2.28. *Prove Proposition* 2.2.2.

Solution: Let W be the intersection of all convex cones containing S. If T is a convex cone containing S, then clearly T also contains $\text{cone}(S)$ by definition. It follows xthat $\text{cone}(S) \subseteq W$. The other way, $\text{cone}(S)$ is also a convex cone containing S, so that $W \subseteq \text{cone}(S)$. Thus $W = \text{cone}(S)$, and the result follows.

Exercise 2.29. *Confer Exercise* 2.9. *Give an example showing that a similar property for linear independence does not hold. Hint: consider the vectors* $(1,0)$ *and* $(0,1)$ *and choose some* w.

Solution: Set $x_1 = (1,0)$, $x_2 = (0,1)$, and $w = -x_1$. Then x_1 and x_2 are linearly independent, but $x_1 + w$ and $x_2 + w$ are not, since $x_1 + w = 0$.

Exercise 2.30. *If* $x = \sum_{j=1}^t \lambda_j x_j \hat{A}$ *and* $\sum_{j=1}^t \lambda_j = 1$ *we say that* x *is an affine combination of* $x_1, \ldots, x_t$. *Show that* $x_1, \ldots, x_t\hat{A}$ *are affinely independent if and only if none of these vectors may be written as an affine combination of the remaining ones.*

Solution: Affine independence of $x_1, \ldots, x_t$ is the same as independent columns in $\begin{pmatrix} X \\ 1 \cdots 1 \end{pmatrix}$. That x_i can be written as an affine combination of the remaining ones means that column i in $\begin{pmatrix} X \\ 1 \cdots 1 \end{pmatrix}$ can be written as a linear combinations of the other columns. The result now follows from the fact that, a set of vectors is linearly independent if and only if none of the vectors can be written as a non-trivial linear combination of the others.

Exercise 2.31. *Prove that* $x_1, \ldots, x_t \in \mathbb{R}^n\hat{A}$ *are affinely independent if and only if the vectors* $(x_1, 1), \ldots, (x_t, 1) \in \mathbb{R}^{n+1}$ *are linearly independent.*

Solution See Exercise 2.8.

Exercise 2.32. *Prove Proposition* 2.3.2*.*

Solution: Asume that $x = \sum_{j=1}^{t} \lambda_j x_j$ where $\sum \lambda_j = 1$. Then

$$\begin{pmatrix} X \\ 1 \cdots 1 \end{pmatrix} \lambda = \begin{pmatrix} x \\ 1 \end{pmatrix}.$$

The convex hull consist of the vectors x for which this has a solution. Affine independence means that the left hand side columns are linearly independent, which implies that the solution λ is unique when it exist. This in turn implies that x has a unique representation.

Exercise 2.33. *Prove that* $\mathrm{cl}(A_1 \cup \ldots \cup A_t) = \mathrm{cl}(A_1) \cup \ldots \cup \mathrm{cl}(A_t)$ *holds whenever* $A_1, \ldots, A_t \subseteq \mathbb{R}^n$*.*

Solution: $\mathrm{cl}(A_1) \cup \ldots \cup \mathrm{cl}(A_t)$ is a closed set containing $A_1 \cup \ldots \cup A_t$. Since $\mathrm{cl}(A_1 \cup \ldots \cup A_t)$ is the smallest such set, it follows that

$$\mathrm{cl}(A_1 \cup \ldots \cup A_t) \subseteq \mathrm{cl}(A_1) \cup \ldots \cup \mathrm{cl}(A_t).$$

The other way, since $A_i \subseteq A_1 \cup \ldots \cup A_t$, $\mathrm{cl}(A_i) \subseteq \mathrm{cl}(A_1 \cup \ldots \cup A_t)$ (since a closed set S containing $A_1 \cup \cdots \cup A_t$ also must contain A_i). But then also

$$\mathrm{cl}(A_1) \cup \ldots \cup \mathrm{cl}(A_t) \subseteq \mathrm{cl}(A_1 \cup \ldots \cup A_t),$$

and the result follows.

Exercise 2.34. *Prove that every bounded point sequence in* $\mathbb{R}^n$ *has a convergent subsequence.*

Solution: Since any subsequence of a convergent sequence also is convergent, it is enough to prove the result for $n = 1$. Due to boundedness there exists an M so that $|x_i| \leq M$ for all i. Partition the interval $[-M, M]$ into two parts of equal length. One of these parts must contain an infinite number of the x_i. Choose the first x_{i_1} taking value in this part, and split the new interval in 2 again. Pick an x_{i_2} from the new interval, and split in two again, and so on. The new sequence y defined as $y_n = x_{i_n}$ is clearly a Cauchy sequence, so that it is convergent.

Exercise 2.35. *Find an infinite set of closed intervals whose union is the open interval* $\langle 0, 1 \rangle$*. This proves that the union of an infinite set of closed intervals may not be a closed set.*

Solution: We have that $(0, 1) = \cup_{n=3}^{\infty} [1/n, 1 - 1/n]$.

Exercise 2.36. *Let* $S\hat{A}$ *be a bounded set in* $\mathbb{R}^n$*. Prove that* $\mathrm{cl}(S)\hat{A}$ *is compact.*

Solution: Sine cl(S) is closed, we only have to prove that cl(S) is bounded. Since S is bounded, we can find a (closed) ball $B(0,r)$ so that $S \subseteq B(0,r)$. Since $B(0,r)$ is a closed set containing S, it follows that cl$(S) \subseteq B(0,r)$. Since $B(0,r)$ is bounded, so is cl(S).

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

Solution: We know that we can write $\text{aff}(S) = L + x_0$, where L is a vector space, and where x_0 can be chosen to be any $x_0 \in \text{aff}(S)$ (in particular we can choose any $x_0 \in S$). If L has dimension n, $\text{aff}(S) = \mathbb{R}^n$. In this case clearly the relative interior equals the interior (their definitions coincide).

Assume now that L has dimension $m < n$. Let $x_0, \dots x_m$ be a maximum number of affinely independent points in S (i.e., the $x_i - x_0$ are linearly independent). Find a vector $x \neq 0$ so that $x - x_0$ can not be written as a linear combination of the $\{x_i - x_0\}_{i=1}^m$. I claim that, for any choice of λ_i, $x_0 + \sum_{i=1}^m \lambda_i(x_i - x_0) + \epsilon(x - x_0)$ can not be in aff(S). Otherwise we would have that

$$x_0 + \sum_{i=1}^m \lambda_i(x_i - x_0) + \epsilon(x - x_0) = x_0 + \sum_{i=1}^m \mu_i(x_i - x_0)$$

for some μ_i (since $x_0 + \sum_{i=1}^m \mu_i(x_i - x_0)$ describes a general element in aff(S)). From this one could write $x - x_0$ as a linear combination of the $x_i - x_0$, which is a contradiction. It follows that $\text{int}(\text{aff}(S)) = \emptyset$ (since ϵ was arbitrary). But then also $\text{int}(S) = \emptyset$.

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

Solution: This is a compulsory exercise.

Exercise 2.39. *Prove Theorem* 2.5.2.

Solution: Since $x \in \text{cone}(S)$ there are nonnegative numbers $\lambda_1, \dots, \lambda_t$Â and vectors $x_1, \dots, x_t \in S$ such that $x = \sum_j \lambda_j x_j$. In fact, we may assume that each λ_j is positive, otherwise we could omit some x_j from the representation. If $x_1, \dots, x_t$Â are linearly independent, we are done, so assume that they are not. Then there are numbers $\mu_1, \dots, \mu_t$ not all zero such that $\sum_{j=1}^t \mu_j x_j = O$. Since the μ_js are not all zero, pick a nonzero one, say $\mu_1 \neq 0$. We now multiply the equation $\sum_j \mu_j x_j = O$ by a number α and subtract the resulting equation from the equation $x = \sum_j \lambda_j x_j$. This gives

$$x = \sum_j (\lambda_j - \alpha\mu_j)x_j.$$

Note that, for small α, this is still a nonnegative linear combination, and when $\alpha = 0$ it is just the original representation of x. But now we gradually increase or decrease αÂ from zero until one of the coefficients $\lambda_j - \alpha\mu_j$ becomes zero, say this happens for $\alpha = \alpha_0$. Recall here that each λ_jÂ is positive and that $\mu_1 \neq 0$. Then each coefficient $\lambda_j - \alpha_0\mu_j$Â is nonnegative and at least one of them is zero. But this means that we have found a new representation of xÂ as a nonnegative combination of $t-1$ vectors from S. Clearly, this reduction process may be continued until we have xÂ written as a nonnegative combination of, say, m linearly independent points in S. Finally, there are at most n linearly independent points in $\mathbb{R}^n$, so $m \leq n$.

Alternative proof for Theorem 2.4.2

To me it is not completely clear that $y \in \text{aff}(C)$ (third last line in the proof), so I have put together a more constructive proof, which I think is clearer. The following prerequisite for the new proof is useful (see page 8 in Rockafellar's book on convex analysis): Let $\{b_0, \ldots, b_m\}$ and $\{c_0, \ldots, c_m\}$ be two sets of points in $\mathbb{R}^n$, both affinely independent. Let A be an invertible linear transformation from $\mathbb{R}^n$ to itself so that $A(b_i - b_0) = c_i - c_0$ for $i = 1, \ldots, m$ (such a linear transformation exists due to linear independence of the $b_i - b_0$, $c_i - c_0$). The affine transformation T defined by $Tx = Ax + c_0 - Ab_0$ satisfies $Tb_i = c_i$ for $i = 0, \ldots, m$. For $i = 0$ this is immediate. For $i \geq 1$ we obtain

$$Tb_i = Ab_i + c_0 - Ab_0 = A(b_i - b_0) + Ab_0 + c_0 - Ab_0 = c_i - c_0 + Ab_0 + c_0 - Ab_0 = c_i.$$

Clearly this affine transformation is unique if $m = n$ (its extension to the affine hull of the vectors is also unique). Since T is invertible it preserves open and closed sets, and also affine hulls (since the line through x and y is sent to the line between Tx and Ty). It follows that T also preserves relative interior. Now, let $\{c_0, c_1, \ldots, c_m\} = \{0, e_1, \ldots, e_m\}$. T then embeds a convex set of dimension m into $\mathbb{R}^m$, where $\mathbb{R}^m$ is obtained from $\mathbb{R}^n$ by setting the mast $n-m$ coordinates to 0. Also, the aff(S) can clearly be identified with $\mathbb{R}^m$ as a subset of $\mathbb{R}^n$ in the same way.

Now for the proof itself, which also can be found on page 45 in Rockafellar's book: Since we can assume that the last $n-m$ coordinates are zero, we can assume that we are in $\mathbb{R}^m$, that the relative interior equals the interior, and that the affine hull equals the entire $\mathbb{R}^m$.

Let $w = (1-\lambda)x_1 + \lambda x_2$ with $x \in \text{rint}(C) = \text{int}(C)$, $y \in \text{cl}(C)$. Let B be the unit ball of $\mathbb{R}^m$. We need to show that $(1-\lambda)x_1 + \lambda x_2 + \epsilon B \subseteq C$ for some $\epsilon > 0$. Since

$x_2 \in C + \epsilon B$ for every ϵ (since $x_2 \in \mathrm{cl}(C)$), we obtain

$$\begin{aligned}(1-\lambda)x_1 + \lambda x_2 + \epsilon B &\subseteq (1-\lambda)x_1 + \lambda(C + \epsilon B) + \epsilon B \\ &= (1-\lambda)x_1 + \lambda C + \lambda\epsilon B + \epsilon B \\ &= (1-\lambda)\left(x_1 + \epsilon\frac{1+\lambda}{1-\lambda}B\right) + \lambda C,\end{aligned}$$

where the equality between the first and second line results from Exercise 1.14(i), and the equality between the second and third line results from convexity of B combined with (iii) in the same exercise. For ϵ sufficiently small we have that $x_1 + \epsilon\frac{1+\lambda}{1-\lambda}B \subseteq C$, so that the above is in $(1-\lambda)C + \lambda C \subseteq C$, due to convexity. It follows that $(1-\lambda)x_1 + \lambda x_2 + \epsilon B \subseteq C$, so that $(1-\lambda)x_1 + \lambda x_2 \in \mathrm{int}(C) = \mathrm{rint}(C)$. The proof follows.

Note that this proof simplifies in the sense that the case $x_2 \in C$ needs not be handled separately first.

Chapter 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 $x\hat{A}$ and a set S such that every point of $S\hat{A}$ is a nearest point to x!*

Solution: Let S be the unit disk in $\mathbb{R}^2$, and let $x = (2, 0)$. Then clearly $s = (1, 0)$ is the unique nearest point.

Let S be the unit circle in $\mathbb{R}^2$, and let $x = (0, 0)$. Then there is no unique nearest point, since every point in S has the same distance to x, so that every point in S is a nearest point.

Exercise 3.2. *Let $a \in \mathbb{R}^n \setminus \{O\}$ and $x_0 \in \mathbb{R}^n$. Then there is a unique hyperplane H that contains $x_0\hat{A}$ and has normal vector a. Verify this and find the value of the constant α (see above).*

Solution: The hyperplane is $a^T x = a^T x_0$. This hyperplane is unique since any other value of α will exclude x_0 from the set.

Exercise 3.3. *Give an example of two disjoint sets $S\hat{A}$ and $T\hat{A}$ that cannot be separated by a hyperplane.*

Solution: Let S be the circle in $\mathbb{R}^2$ consisting of points with modulus 1, T the circle in $\mathbb{R}^2$ consisting of points with modulus 2.

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:

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 $x\hat{A}$ in C. Now, let C be the unit ball in $\mathbb{R}^n$ and let $x \in \mathbb{R}^n\hat{A}$ satisfy $\|x\| > 1$. Find the nearest point to x in C. What if $\|x\| \leq 1$?*

Solution: If $\|x\| > 1$ then clearly $x_0 = x/\|x\|$ is the nearest point. If $x \leq 1$ then x itself is the nearest point.

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: This is a compulsory exercise.

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 points $(0, 0, 0)$ and $(1, 2, 2)$ in the hyperplane $H = \{(x_1, x_2, x_3) : x_1 + x_2 + x_3 = 1\}$.*

Solution: A general hyperplane in $\mathbb{R}^n$ can be written on the form $x_0 + L$, where L is an $n-1$-dimensional space. Let a span the orthogonal complement of L. We subtract x_0 and compute the projection to obtain

$$x - x_0 - \frac{\langle x - x_0, a\rangle}{\langle a, a\rangle} a + x_0 = x - \frac{\langle x - x_0, a\rangle}{\langle a, a\rangle} a$$

(we here instead projected onto the orthogonal complement).

For the plane in question we obtain $\langle a, a\rangle = 3$, and we can use $x_0 = (1, 0, 0)$ For the point $(0, 0, 0)$ we obtain

$$x - \frac{\langle x - x_0, a\rangle}{\langle a, a\rangle} a = \frac{1}{3}(1, 1, 1),$$

while for the point $(1, 2, 2)$ we obtain

$$x - \frac{\langle x - x_0, a\rangle}{\langle a, a\rangle} a = (1, 2, 2) - \frac{4}{3}(1, 1, 1) = (-1/3, 2/3, 2/3).$$

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, $q_i^T q_j = 0\hat{A}$ when $i \neq j$ and $\|q_j\| = 1$ for each j. Let $q\hat{A}$ be the $n \times t$-matrix whose jth column is q_j, for $j = 1, \ldots, t$. 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: The orthogonal decomposition theorem states that the closest point is

$$\sum_{j=1}^{t} \langle x, q_j\rangle q_j = \sum_{j=1}^{t} (q^T x)_j q_j = qq^T x = px.$$

by the definition of matrix multiplication. Similary, if $q_{t+1}, \ldots, q_n$ is an orthonormal basis for $L^\perp$, and Q is the matrix with these $n - t$ column, we have

$$\sum_{j=t+1}^{n} \langle x, q_j\rangle q_j = \sum_{j=t+1}^{n} (Q^T x)_j q_j = QQ^T x =: Px.$$

That $p + P = I$ (so that $P = I - p$) follows now since, for $x \in L$, $px = x$ and $Px = 0$, while for $x \in L^\perp$, $px = 0$ and $Px = x$. This shows that $I - P$ is the projector onto the orthogonal complement.

Exercise 3.9. *Let $L \subset \mathbb{R}^3$ be the 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: We compute

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 4 \\ 2 \\ 4 \end{pmatrix}.$$

Exercise 3.10. *Show that the nearest point in $\mathbb{R}^n_+$ to $x \in \mathbb{R}^n\hat{A}$ is the point x^+ defined by $x_j^+ = \max\{x_j, 0\}$ for each j.*

Solution: Let $y \in \mathbb{R}^n_+$. We have that

$$\|x - y\| = \sqrt{(x_1 - y_1)^2 + \cdots + (x_n - y_n)^2}.$$

Now, if x_j is negative, the smallest value we can obtain for $(x_j - y_j)^2$ is by choosing $y_j = 0$ (we require that $y_j \in \mathbb{R}_+$). If x_j is nonnegative, the smallest value we can obtain for $(x_j - y_j)^2$ is by choosing $y_j = x_j$. Setting $y_j = \max(0, x_j) = x_j^+$ captures both these possibilities.

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 $S\hat{A}$ is nearest to $x\hat{A}$ in S!*

Solution: This is exercise 3.1

Exercise 3.12. *Verify that every hyperplane in $\mathbb{R}^n$ has dimension $n - 1$.*

Solution: Let the hyperplane be $x^t a = \alpha$, let x_1 be a point in the hyperplane, and let $y_2, \ldots, y_n$ span the orthogonal complement of a. Set $x_j = x_1 + y_j$. Then $x_1, \ldots, x_n$ are affinely independent since $x_j - x_1 = y_j$ are linearly independent. The dimension can't be larger, since the orthogonal complement does not have dimension larger than $n - 1$.

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

Solution: Let $0 \le a \le 1/2$. Then $y = ax + b$ separates the two sets if $b > 1$ and $2a + b < 2$.

Let $a \le 0$. Then we have separation if $a + b > 1$ and $2a + b < 2$.

Let $a \ge 2$. Then we have separation if $a + b < 0$ and $a + 2b > 2$.

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

Solution: Set $x_0 = \frac{1}{2}(a + a/\|a\|)$. we can use $a^T x = a^T x_0 = \alpha$.

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

Solution: The left and right half planes give an example.

Exercise 3.16. *Let* $S, T \subseteq \mathbb{R}^n$. *Explain the following fact: there exists a hyperplane that separates* $S\hat{A}$ *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* $s \in S\hat{A}$ *and* $t \in T$. *Is there a similar equivalence for the notion of strong separation?*

Solution: Separation means that $a^T x \leq \alpha$ on S, $a^T x \geq \alpha$ on T. Setting $l(x) = a^T x$ (which is linear), this means that $l(x) \leq \alpha$ on S, $l(x) \geq \alpha$ on T. This implies one way. The other way follows by setting α to be any number in the interval $[\max_{x \in S}(l(x)), \min_{y \in T}(l(y))]$.

For strong separation, there must be some ϵ so that $l(x) \leq \alpha - \epsilon$ on S, $l(x) \geq \alpha + \epsilon$ on T.

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

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

(Such an operator is called nonexpansive). 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: From (3.2) it follows that $(x - p_C(x))^T(y - p_C(x)) = a^T(y - p_C(x)) \leq 0$ for all $y \in C$. If we in particular set $y = p_C(y)$ we obtain $a^T(p_C(y) - p_C(x)) \leq 0$. Similarly, $(x - p_C(y))^T(y - p_C(y)) = b^T(x - p_C(y)) \leq 0$ for all $x \in C$. If we in particular set $x = p_C(x)$ we obtain $b^T(p_C(x) - p_C(y)) \leq 0$.

We now obtain

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

The result follows after taking square roots.

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) $C_2 = \mathrm{conv}(\{0,1\}^n)$,
(iii) $C_3 = \mathrm{conv}(\{-1,1\}^2)$
(iv) $C_4 = \mathrm{conv}(\{-1,1\}^n)$, $n > 2$.

Solution: This is a compulsory exercise.

Chapter 4

Representation of convex sets

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

(i) Find the unique face of $P\hat{A}$ that contains the point $(1/3, 1/2)$.
(ii) Find all the faces of $P\hat{A}$ that contain the point $(1/3, 2/3)$.
(iii) Determine all the faces of P.

Solution:

(i) The face is the entire polytope.
(ii) This is the line from $(1,0)$ to $(0,1)$.
(iii) The three vertices, the three edges, and the polytope 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: The old condition clearly implies that the new condition holds (simply choose $\lambda = 1/2$). The other way, assume that the new condition holds, and let $x_1, x_2 \in C$ be so that $x = (1-\lambda)x_1 + \lambda x_2 \in F$. If $\lambda \leq 1/2$, x is the midpoint between x_1 and $(1-2\lambda)x_1 + 2\lambda x_2$, and since both these are on the line segment between x_1 and x_2 (and hence in C), it follows from the condition that $x_1 \in F$ $(1-2\lambda)x_1 + 2\lambda x_2$. This can be repeated until we find a point in F which is on the second part of the line segment between x_1 and x_2. In the same way, if $\lambda \geq 1/2$, it follows that $x_2 \in F$. Combining these two proves the result.

Exercise 4.3. *Prove this proposition!*

Solution: Let $x_1, x_2 \in C$, and let $(1-\lambda)x_1 + \lambda x_2 \in F_2$. Since $F_2 \subseteq F_1$ and F_1 is a face of C, it follows that $x_1, x_2 \in F_1$. Since F_2 is a face of F_1 it follows that $x_1, x_2 \in F_2$ as well. It follows that F_2 is a face of C.

Exercise 4.4. *Define $C = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0,\ x_2 \geq 0,\ x_1 + x_2 \leq 1\}$. Why does C not have any extreme halfline? Find all the extreme points of C.*

Solution: C is bounded. Existence of an extreme halfline would imply C to be unbounded. The extreme points of C are clearly $(0,0)$, $(1,0)$, and $(0,1)$.

Exercise 4.5. *Consider a polytope $P \subset \mathbb{R}^n$, say $P = \text{conv}(\{x_1, \ldots, x_t\})$. Show that if $x\hat{A}$ is an extreme point of P, then $x \in \{x_1, \ldots, x_t\}$. Is every x_j necessarily an extreme point?*

Solution: Write an extreme point as $x = \sum_{j=1}^{y} \lambda_j x_j \in P$ with $\sum_{j=1}^{t} \lambda_j = 1$. Assume that $0 < \lambda_1 < 1$. We can rewrite

$$x = \lambda_1 x_1 + (1 - \lambda_1) \sum_{j=2}^{t} \frac{\lambda_j}{1 - \lambda_1} x_j.$$

Since $x_1 \in P$ and $\sum_{j=2}^{t} \frac{\lambda_j}{1-\lambda_1} x_j \in P$, the extreme point property implies that $x = x_1$.

Not every x_j needs to be an extreme point: simply add an x_j which already is in the convex hull of the previous ones to see this.

Exercise 4.6. *Show that* $\text{rec}(C, x)$ *is a closed convex cone. First, verify that* $z \in \text{rec}(C, x)$ *implies that* $\mu z \in \text{rec}(C, x)\hat{A}$ *for each* $\mu \geq 0$. *Next, in order to verify convexity you may show that*

$$\text{rec}(C, x) = \bigcap_{\lambda > 0} \frac{1}{\lambda}(C - x)$$

where $\frac{1}{\lambda}(C - x)\hat{A}$ *is the set of all vectors of the form* $\frac{1}{\lambda}(y - x)$ *where* $y \in C$.

Solution: Since $x + z = x + \frac{1}{\mu}(\mu z)$, we see that $\mu z \in \text{rec}(C, x)$ whenever $z \in \text{rec}(C, x)$ and $\mu > 0$. This applies also when $\mu = 0$, since trivially $0 \in \text{rec}(C, x)$.

An element in $\bigcap_{\lambda>0} \frac{1}{\lambda}(C - x)$ can for any $\lambda > 0$ be written on the form $z = \frac{1}{\lambda}(c - x)$ for some $c \in C$, so that $x + \lambda z = c$. This is clearly the same as $z \in \text{rec}(C, x)$. $\text{rec}(C, x)$ can thus be ritten as an intersection of closed convex sets, and is thus also closed and convex.

Another way to show convexity is as follows: Let $z_1, z_2 \in \text{rec}(C, x)$, $0 \leq \lambda \leq 1$, and $\lambda' \geq 0$. We have that

$$x + \lambda' \left((1 - \lambda) z_1 + \lambda z_2\right) = (1 - \lambda)(x + \lambda' z_1) + \lambda(x + \lambda z_2)$$

$z_1, z_2 \in \text{rec}(C, x)$ implies $x + \lambda' z_1 \in C$, $x + \lambda z_2 \in C$. By convexity of C it follows that the left hand side also is in C, so that $(1 - \lambda) z_1 + \lambda z_2 \in C$. Convexity of $\text{rec}(C, x)$ follows.

Exercise 4.7. *Show that a closed convex set* $C\hat{A}$ *is bounded if and only if* $\text{rec}(C) = \{O\}$.

Solution: Clearly C is unbounded if the recession cone is nontrivial. The other way, when C is unbounded, we will construct a nonzero vector in $\mathrm{rec}(C)$. The proof is long and involved, and builds on many small lemmas (the three stated below are not proved here, see the book of Rockafellar).

Assume that $C \subseteq \mathbb{R}^n$. Consider the convex cone $K \subseteq \mathbb{R}^{n+1}$ generated by $(1, x)$, with $x \in C$. Clearly $(0, 0)$ is the only point in K with first component zero. We will attempt to extend K to a cone $K' = K \cup K_0$, where $K' \subseteq \{(\lambda, x) : \lambda \geq 0\}$, $K_0 \subseteq \{(\lambda, x) : \lambda = 0\}$. K_0 is thus required to be a cone, and so that $K_0 + K \subseteq K$. In particular we must have that, for any $z \in C$,

$$(0, x) + (1, z) = \sum_i \lambda_i (1, x_i)$$

for some $x_i \in C$, $\lambda_i \geq 0$. The λ_i must sum to one, so that the right hand side can be written as $(1, w)$ for some $w \in C$ (by convexity). It follows that $x + z = w$. i.e., adding any element from C to x gives another element in C. We prove the following statement, which implies that $x \in \mathrm{rec}(C)$:

$$x \in \mathrm{rec}(C) \iff C + x \subseteq C.$$

$\Rightarrow$ is obvious. The other way, if $C + x \subseteq C$, then $C + 2x \subseteq C$, and in general $C + mx \subseteq C$ for any integer m. letting $\lambda \geq 0$ be any integer, $y + \lambda x$ lies on the between y and $a + mx$ where m is some integer larger than λ. By convexity $y + \lambda x \in C$ (it is one the line between y and $y + mx$, which both are in C). It follows that $x \in \mathrm{rec}(C)$.

Thus, $K_0 \subseteq (0, \mathrm{rec}(C))$. Since $(0, \mathrm{rec}(C))$ itself is a cone, it follows that $K_0 = (0, \mathrm{rec}(C))$ is the largest cone we can choose for K_0. This choice also satisfies $K_0 + K \subseteq K$, so that $K_0 \cup K$ is a cone. To see this, let $C_\lambda = \{x : (\lambda, x) \in K\}$ for $\lambda \geq 0$. Clearly $\lambda C \subset C_\lambda$. The other way, if $x \in C_\lambda$, then $(\lambda, x) \in K$ so that

$$(\lambda, x) = \sum_i \lambda_i (1, x_i) = (\sum_i \lambda_i, \sum_i \lambda_i x_i)$$

with $\lambda_i \geq 0$. It follows that

$$x = \sum_i \lambda_i x_i = \lambda \sum_i \frac{\lambda_i}{\lambda} x_i \in \lambda C$$

since $\sum_i \lambda_i = \lambda$ and due to convexity of C. It follows that $\lambda C = C_\lambda$. From this, since

$$(0, x) + (\lambda, z) = (0, x) + (\lambda, \lambda y) = (\lambda, \lambda(\frac{1}{\lambda} x + y)) \in K$$

(since $\frac{1}{\lambda}x \in \text{rec}(C)$). $K_0 + K \subseteq K$ follows.

Below we will prove that $\text{cl}(K) \subseteq K'$ (actually one can prove equality here as well with a bit more work). Let us see how we can use this to construct a nonzero vector in $\text{rec}(C)$ when C is unbounded, and thereby complete the proof. If C is unbounded we can find a sequence x_n from C so that $\|x_n\| \to \infty$. The elements $(1/\|x_n\|, x_n/\|x_n\|)$ are in K. One can find a convergent subsequence so that $x_{n_k}/\|x_{n_k}\| \to y$, where $\|y\| = 1$. But then $(1/\|x_{n_k}\|, x_{n_k}/\|x_{n_k}\|) \to (0, y)$. Since $\text{cl}(K) \subseteq K'$ it follows that $y \in \text{rec}(C)$.

Proving that $\text{cl}(K) \subseteq K'$ requires two additional results, both taken from the book of Rockafellar. Set M_λ to be the affine set $\{(\lambda, x) : x \in \mathbb{R}^n\}$:

1. (Theorem 6.6). If K is convex and A is a linear transformations, the $\text{rint}(AK) = A\text{rint}(K)$. We will use this as follows: Consider the projection $P : (\lambda, x) \to \lambda$ (from $\mathbb{R}^{n+1}$ to $\mathbb{R}$). P is linear, so that $P(\text{rint}(K)) = \text{rint}(P(K)) = \text{rint}(\mathbb{R}^+) = (0, \infty)$. It follows that $\text{rint}(K) \cap M_\lambda \neq \emptyset$ for any $\lambda > 0$.

2. (Corollary 6.5.1). If K is convex and M is affine and contains a point in $\text{rint}(K)$, then

$$\begin{aligned} \text{rint}(M \cap C) &= M \cap \text{rint}(C) \\ \text{cl}(M \cap C) &= M \cap \text{cl}(C). \end{aligned}$$

We will use this as follows, using $M = M_\lambda$ (due to 1., $\text{rint}(K) \cap M_\lambda \neq \emptyset$):

$$M_\lambda \cap \text{rint}(K) = \text{rint}(M_\lambda \cap K) = \text{rint}(\lambda, C_\lambda) = \text{rint}(\lambda, \lambda C) = \lambda(1, \text{rint}(C)).$$

This proves that

$$\text{rint}(K) = \cup_{\lambda > 0}(M_\lambda \cap \text{rint}(K)) = \cup_{\lambda > 0}\lambda(1, \text{rint}(C)).$$

Since $\text{rint}(K) \cap M_1 \neq \emptyset$ in particular, $\text{cl}(M_1 \cap K) = M_1 \cap \text{cl}(K)$ (second part of Corollary 6.5.1). Since $M \cap K = \{(1, x) : x \in C\}$, which is closed, we have that $M_1 \cap \text{cl}(K) = \{(1, x) : x \in C\}$. It follows that $\text{cl}(K)$ and K are equal on $\lambda > 0$. By maximality of K' it follows that $\text{cl}(K) \subseteq K'$.

Exercise 4.8. *Consider a hyperplane H. Determine its recession cone and lineality space.*

Solution: Let $a^T x = \alpha$ be a hyperplane. $\text{rec}(H, x)$ consists of all z so that $x + \lambda z \in H$ for all $x \in H$ and $\lambda \geq 0$, i.e., $\alpha = a^T(x + \lambda z) = \alpha + \lambda a^T z$ for all λ. This implies that z must be orthogonal to a, i.e., $\text{rec}(C, x) = \{a\}^\perp$. Since the recession cone here is a vector space, it equals its lineality space.

Exercise 4.9. *What is* $\text{rec}(P)$ *and* $\text{lin}(P)$ *when* $P\hat{A}$ *is a polytope?*

Solution: Since polytopes are bounded, we have that $\mathrm{rec}(P) = \mathrm{lin}(P) = \{0\}$.

Exercise 4.10. *Let $C\hat{A}$ be a closed convex cone in $\mathbb{R}^n$. Show that $\mathrm{rec}(C) = C$.*

Solution: If $x \in C$ then $(1+\lambda)x = x+\lambda x \in C$ for any $\lambda \geq 0$, so that $x \in \mathrm{rec}(C)$. The other way, if $x \in \mathrm{rec}(C)$ then $0 + 1 \cdot x = x \in C$ since $0 \in C$. The result follows.

Exercise 4.11. *Prove that $\mathrm{lin}(C)\hat{A}$ is a linear subspace of $\mathbb{R}^n$.*

Solution: $\mathrm{lin}(C)$ is closed under multiplication with scalars:

If $x \in \mathrm{rec}(C)$ then $-x \in -\mathrm{rec}(C)$. If $x \in -\mathrm{rec}(C)$ then $-x \in \mathrm{rec}(C)$. It follows that $\mathrm{lin}(C)$ is closed under multiplication with -1, and thus also with all scalars. Since both $\mathrm{rec}(C)$ and $-\mathrm{rec}(C)$ are closed under addition, the result follows.

Exercise 4.12. *Show that $\mathrm{rec}(\{x : Ax \leq b\}) = \{x : Ax \leq O\}$.*

Solution: Let $C = \{x : Ax \leq b\}$, and assume that $z \in \mathrm{rec}(C, x)$. Then $Ax \leq b$ and $A(x+\lambda z) \leq b$ for all $\lambda \geq 0$. Since $A(x+\lambda z) = Ax+\lambda Az$, this is clearly only possible if $Az \leq 0$. This implies the $\subseteq$-direction.

The other way, if $Az \leq 0$, then, for any $x \in C$, $\lambda \geq 0$, we have that $A(x + \lambda z) \leq b + 0 = b$, so that $x + \lambda z \in C$. It follows that $z \in \mathrm{rec}(C)$. This proves the other direction.

Exercise 4.13. *Let C be a line-free closed convex set and let $F\hat{A}$ be an extreme halfline of C. Show that then there is an $x \in C\hat{A}$ and a $z \in \mathrm{rec}(C)$ such that $F = x + \mathrm{cone}(\{z\})$.*

Solution: An extreme halfline can be written on the form $\{x + tz\}_{t\geq 0}$ for some $x \in C$, and some vector z. Since the extreme halfline is in C, it follows that $z \in \mathrm{rec}(C, x) = \mathrm{rec}(C)$.

Exercise 4.14. *Decide if the following statement is true: if $z \in \mathrm{rec}(C)$ then $x + \mathrm{cone}(\{z\})\hat{A}$ is an extreme halfline of C.*

Solution: No. Let $C = \mathbb{R}^n$. Then $\mathrm{rec}(C) = \mathbb{R}^n$, and C has no extreme half lines.

Exercise 4.15. *Consider again the set $C = \{(x_1, x_2, 0) \in \mathbb{R}^3 : x_1^2 + x_2^2 \leq 1\}$ from Exercise 2.16. Convince yourself that C equals the convex hull of its relative boundary. Note that we here have $\mathrm{bd}(C) = C\hat{A}$ so the fact that $C\hat{A}$ is the convex hull of its boundary is not very impressive!*

Solution: Its relative boundary has been shown to be $\{(x_1, x_2, 0) : x_1^2 + x_2^2 = 1\}$, and the convex hull of these is clearly C.

Exercise 4.16. *Let H be a hyperplane in $\mathbb{R}^n$. Prove that $H \neq \mathrm{conv}(\mathrm{rbd}(H))$.*

Solution: The relative boundary of a hyperplane is $\emptyset$, so that $\mathrm{conv}(\mathrm{rbd}(H)) = \emptyset \neq H$.

Exercise 4.17. *Consider a polyhedral cone $C = \{x \in \mathbb{R}^n : Ax \leq O\}$ (where, as usual, $A\hat{A}$ is a real $m \times n$-matrix). Show that O is the unique vertex of C.*

Solution: This is a compulsory exercise.

Exercise 4.18. *Let $F\hat{A}$ be a face of a convex set C in $\mathbb{R}^n$. Show that every extreme point of F is also an extreme point of C.*

Solution: This follows from Exercise 4.3

Exercise 4.19. *Find all the faces of the unit ball in $\mathbb{R}^2$. What about the unit ball in $\mathbb{R}^n$?*

Solution: Every point on the boundary is a face.

Exercise 4.20. *Let $F\hat{A}$ be a nontrivial face of a convex set C in $\mathbb{R}^n$. Show that $F \subseteq \mathrm{bd}(C)\hat{A}$ (recall that $\mathrm{bd}(C)$ is the boundary of C). Is the stronger statement $F \subseteq \mathrm{rbd}(C)$ also true? Find an example where $F = \mathrm{bd}(C)$.*

Solution: Assume that F contains an interior point x. Then clearly F must contain a small ball around x. Let y be any other point in C. We can draw a line through x starting at y and ending at the opposite side of x in another point in C. By the face property this implies that $y \in C$, so that $F = C$. This shows that any non-trivial face is contained in the boundary. $F = \mathrm{bd}(C)$ is possible when C is the left or right side of a hyperplane.

Assume that F contains a relative interior point x of C. Then $B^o(x,r) \cap \mathrm{aff}(C) \subseteq C$ for some ball. Let y be another point in C. Draw again a line through x starting at y and ending at the opposite side of x in another point in C. By the face property this implies that $y \in C$, so that $F = C$.

Exercise 4.21. *Consider the convex set $C = B + ([0,1] \times \{0\})$ where B is the unit ball (of the Euclidean norm) in $\mathbb{R}^2$. Find a point on the boundary of C which is a face of C, but not an exposed face.*

Solution: Choose the point $(0,1)$ for instance. This gives the line from $(0,1)$ to $(1,1)$ as an exposed face, but the point itself is a face.

Exercise 4.22. *Let $P \subset \mathbb{R}^2\hat{A}$ be the polyhedron being the solution set of the linear system*

$$\begin{array}{rcrl} x & - & y & \leq 0; \\ -x & + & y & \leq 1; \\ & & 2y & \geq 5; \\ 8x & - & y & \leq 16; \\ x & + & y & \geq 4. \end{array}$$

Find all the extreme points of P.

Solution: This is a compulsory exercise.

Exercise 4.23. *Find all the extreme halflines of the cone $\mathbb{R}^n_+$.*

Solution: Clearly any halfline where all except one component is nonzero is an extreme halfline. Assume that we have an extreme halfline where there are points with two nonzero components. Assume for simplicity that $(x_1, x_2, 0)$ is on the extreme halfline. Then clearly the halfline must contain all points with arbitrary values in the first two components, so that it can't be a halfline.

Exercise 4.24. *Determine the recession cone of the set* $\{(x_1, x_2) \in \mathbb{R}^2 : x_1 > 0,\ x_2 \geq 1/x_1\}$. *What are the extreme points?*

Solution: Clearly the recession cone is $\mathbb{R}^2_+$. All points in the first quadrant on the graph $y = 1/x$ are extreme points.

Exercise 4.25. *Let* B *be the unit ball in* $\mathbb{R}^n$ *(in the Euclidean norm). Show that every point in* $B\hat{A}$ *can be written as a convex combination of two of the extreme points of* C.

Solution: The extreme points of B are the boundary points of B. Clearly any interior point of B can be written as a convex combination of two boundary points.

Exercise 4.26. *Let* C *be a compact convex set in* $\mathbb{R}^n$ *and let* $f : C \to \mathbb{R}$ *be a function satisfying*

$$f(\sum_{j=1}^{t} \lambda_j x_j) \leq \sum_{j=1}^{t} \lambda_j f(x_j)$$

whenever $x_1, \ldots, x_t \in C$, $\lambda_1, \ldots, \lambda_t \geq 0\hat{A}$ *and* $\sum_{j=1}^{t} \lambda_j = 1$. *Such a function is called convex, see chapter* 5. *Show that* f *attains its maximum over* C *in an extreme point. Hint: Minkowski's theorem.*

Solution: According to Minkowski's theorem, C is the convex hull of its extreme points. Write $x \in C$ as $x = \sum \lambda_i x_j$, where the x_j are extreme points. But then

$$f(x) \leq \sum \lambda_j f(x_j) \leq \max f(x_j).$$

This implies that, if $x = \text{argmax}_{x \text{ extreme}}(f(x))$ is attained, then the maximum is attained in the extreme point x. It turns out, however, that this maximum may not be attained: Let $C = \{x \in \mathbb{R}^2 : \|x\| \leq 1\}$, and consider the function f defined to be zero on the interior of C, and defined on the boundary so that it both ≥ 0 and unbounded there. f is clearly convex, but the maximum x is not attained due to unboundedness on the boundary.

Exercise 4.27. *Prove Corollary* 4.4.5 *using the Main theorem for polyhedra.*

Solution: Clearly a polytope is a bounded polyhedron. Assume now that $P = \text{conv}(V) + \text{cone}(Z)$ is a bounded polyhedron, where V and Z are finite sets. Boundedness implies that Z is empty, so that $P = \text{conv}(V)$. It follows that P is a polytope.

Exercise 4.28. *Let $S \subseteq \{0,1\}^n$, i.e., $S\hat{A}$ is a set of $(0,1)$-vectors. Define the polytope $P = \text{conv}(S)$. Show that $x\hat{A}$ is a vertex of $P\hat{A}$ if and only if $x \in S$.*

Solution: For polytopes, vertices and extreme points coincide.

Clearly any $x \in S$ is an extreme point: A line which passes through $x \in S$ must either increase or decrease at least one component linearly, in which case one of the end components lies outside $[0,1]$. But any component of a vector in P must lie between 0 and 1. It follows that the points are extreme points. It is also clear that any vector outside S with 0,1 components must be outside P.

Assume that x is an extreme point of P with $0 < x_1 < 1$. Then x must be a convex combinations of two different points in S, contradicting that it is an extreme point. Therefore, all components are either 0 or 1 in an extreme point.

Exercise 4.29. *Let $S \subseteq \{0,1\}^3$ consist of the points $(0,0,0)$, $(1,1,1)$, $(0,1,0)$ and $(1,0,1)$. Consider the polytope $P = \text{conv}(S)$ and find a linear system defining it.*

Solution: Since all four points have equal first and third candidate, we can eliminate the third component by requiring $x_3 = x_1$. The convex hull of the four points $(0,0)$, $(0,1)$, $(1,0)$, and $(1,1)$ in $\mathbb{R}^2$ is clearly described by $0 \leq x_1, x_2 \leq 1$. A possible linear system defining the system is thus

$$\begin{aligned} x_1, x_2, x_3 &\geq 0 \\ x_1, x_2 &\leq 1 \\ x_3 - x_1 &= 0 \end{aligned}$$

Exercise 4.30. *Let $P_1\hat{A}$ and $P_2\hat{A}$ be two polytopes in $\mathbb{R}^n$. Prove that $P_1 \cap P_2\hat{A}$ is a polytope.*

Solution: Both P_1 and P_2 are bounded polyhedra. But then $P_1 \cap P_2$ is also a bounded polyhedron, hence a polytope.

Exercise 4.31. *Is the sum of polytopes again a polytope? The sum of two polytopes $P_1\hat{A}$ and P_2 in $\mathbb{R}^n$ is the set $P_1 + P_2 = \{p_1 + p_2 : p_1 \in P_1, p_2 \in P_2\}$.*

Solution: Let $P_1 = \text{conv}(A)$, $P_2 = \text{conv}(B)$, where A and B are finite sets. By exercise 2.25, $\text{conv}(A + B) = \text{conv}(A) + \text{conv}(B) = P_1 + P_2$. In the case of polytopes A and B are finite sets. But then $A + B$ is also a finite set, so that $\text{conv}(A + B)$ is also a polytope. It follows that $P_1 + P_2$ is a polytope.

Exercise 4.32. *Let $L = \text{span}(\{b_1, \ldots, b_k\})\hat{A}$ be a linear subspace of $\mathbb{R}^n$. Define $b_0 = -\sum_{j=1}^k b_j$. Show that $L = \text{cone}(\{b_0, b_1, \ldots, b_k\})$. Thus, every linear subspace is a finitely generated cone, and we know how to find a set of generators for L (i.e., a finite set with conical hull being L).*

Solution: Clearly $\text{cone}(\{b_0, b_1, \ldots, b_k\}) \subseteq L$, since also $b_0 \in L$. The other way, let $x = \sum_{j=1}^k c_j b_j \in L$. If all $c_k \geq 0$, then this is in $\text{cone}(\{b_0, b_1, \ldots, b_k\})$. Assume thus there are nonnegative c_j, and let c_l be the smallest such. Write

$$x = -c_l \left(-\sum_{j=1}^{k} b_j \right) + \sum_{j \neq l} (c_j - c_l) b_j.$$

Since $c_j \geq c_l$ and $c_l < 0$ this is a nonnegative combination, so that

$$x \in \text{cone}(\{b_0, b_1, \ldots, b_k\}).$$

Exercise 4.33. *Let $P = \text{conv}(\{v_1, \ldots, v_k\}) + \text{cone}(\{z_1, \ldots, z_m\}) \subseteq \mathbb{R}^n$. Define new vectors in $\mathbb{R}^{n+1}$ by adding a component which is 1 for all the $v_\cdot$-vectors and a component which is 0 for all the $z_\cdot$-vectors, and let C be the cone spanned by these new vectors. Thus,*

$$C = \text{cone}(\{(v_1, 1), \ldots, (v_k, 1), (z_1, 0), \ldots, (z_m, 0)\})$$

Prove that $x \in P$ if and only if $(x, 1) \in C$. The cone C is said to be obtained by homogenization of the polyhedron P. This is sometimes a useful technique for translating results that are known for cones into similar results for polyhedra, as in the proof of Theorem 4.4.4.

Solution: Let $x = \sum_j \lambda_j v_j + \sum_k \mu_k z_k \in P$, where $\sum_j \lambda_j = 1$, all $\mu_k \geq 0$. We have that

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_j \lambda_j \begin{pmatrix} v_j \\ 1 \end{pmatrix} + \sum_k \mu_k \begin{pmatrix} z_k \\ 0 \end{pmatrix} \in C.$$

The other way, suppose that $(x, 1) \in C$, so that we can write

$$\begin{pmatrix} x \\ 1 \end{pmatrix} = \sum_j \lambda_j \begin{pmatrix} v_j \\ 1 \end{pmatrix} + \sum_k \mu_k \begin{pmatrix} z_k \\ 0 \end{pmatrix}.$$

Clearly this forces the λ_j to sum to one, and also $x = \sum_j \lambda_j v_j + \sum_k \mu_k z_k$, which thus is in $\text{conv}(\{v_1, \ldots, v_k\}) + \text{cone}(\{z_1, \ldots, z_m\}) = P$.

Exercise 4.34. *Show that the sum of valid inequalities for a set P is another valid inequality for P. What about weighted sums? What can you say about the properties of the set*

$$\{(a, \alpha) \in \mathbb{R}^{n+1} : a^T x \leq \alpha \text{ is a valid inequality for } P\}.$$

Solution: If $c_1^T x \leq \alpha_1$ and $c_2^T x \leq \alpha_2$ for $x \in P$, then clearly

$$(c_1 + c_2)^T x = c_1^T x + c_2^T x \leq \alpha_1 + \alpha_2,$$

for $x \in P$ as well, so that valid inequalities can be summed to obtain new valid inequalities. For weighted sums,

$$(w_1c_1 + w_2c_2)^T x = w_1c_1^T x + w_2c_2^T x \leq w_1\alpha_1 + w_2\alpha_2,$$

so that we get new valid inequalities here as well, as long as the wights are nonnegative. It follows that the stated set is a cone.

Chapter 5

Convex functions

Exercise 5.1. *Prove this lemma.*

Solution: That $P_{x_2}\hat{\text{A}}$ is below the line segment $P_{x_1}P_{x_3}$ (i.e., (i)) is equivalent to

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

Reorganizing this gives that

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

which states that slope$(P_{x_1}, P_{x_2}) \leq$ slope(P_{x_1}, P_{x_3}) (i.e. (ii)). This proves that (i) and (ii) are equivalent.

(i) is also equivalent to

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

Reorganizing this gives that

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

which states that slope$(P_{x_1}, P_{x_3}) \leq$ slope(P_{x_2}, P_{x_3}) (i.e. (iii)).

Exercise 5.2. *Show that the sum of convex functions is a convex function, and that $\lambda f \hat{A}$ is convex if f is convex and $\lambda \geq 0$ (here λf is the function given by $(\lambda f)(x) = \lambda f(x)$).*

Solution: We have that

$$\sum_k f_k((1-\lambda)x+\lambda y) \leq \sum_k((1-\lambda)f_k(x)+\lambda f_k(y)) = (1-\lambda)\sum_k f_k(x)+\lambda\sum_k f_k(y),$$

whichs shows that $\sum_k f_k$ also is convex. We also have that

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

which shows that μf 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) $f(x) = -\ln(x)$ *defined on* $\mathbb{R}_+$*.*

Solution:

(i) $f(x) = x^2$ has a positive second derivative, so that it must be convex.
(ii) $f(x) = |x|$ is convex due to the triangle inequality: $|(1-\lambda)x+\lambda y| \leq (1-\lambda)|x|+\lambda|y|$.
(iii) $f(x) = x^p$ has second derivative $p(p-1)x^{p-2}$, which is nonnegative when $p \geq 1$.
(iv) $f(x) = e^x$ has positive second derivative e^x, so that it is convex.
(v) $f(x) = -\ln x$ has positive second derivative $1/x^2$, so that it is convex.

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* $\lambda_1, \ldots, \lambda_r \hat{A}$ *(nonnegative with sum one). Which inequality do you obtain? It involves the so-called weighted arithmetic mean and the weighted geometric mean.*

Solution: By convexity of $f(x) = -\ln x$ we obtain

$$-\ln(\sum_{j=1}^{r} \lambda_j x_j) \leq -\sum_{j=1}^{r} \lambda_j \ln(x_j) = -\ln(\prod_{j=1}^{r} x_j^{\lambda_j}),$$

which leads to

$$\prod_{j=1}^{r} x_j^{\lambda_j} \leq \sum_{j=1}^{r} \lambda_j x_j.$$

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

Solution: The proof goes in the same way.

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

Solution: If f is linear we have that

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

so that f is also convex and with equality holding in the definition of convexity.

Exercise 5.7. *Prove Proposition* 5.2.1.

Solution: We have that

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

where we used that h is affine and that f is convex. It follows that $f \circ h$ also is convex.

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

Solution: This follows from the previous exercise since $x \to x+w$ is affine.

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

Solution: Assume that f is convex. Let $(x,s) \in \text{epi}(f)$, $(y,t) \in \text{epi}(f)$. We have that

$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y) \leq (1-\lambda)s+\lambda t.$$

It follows that $(1-\lambda)(x,s)+\lambda(y,t) = ((1-\lambda)x+\lambda y, (1-\lambda)s+\lambda t) \in \text{epi}(f)$, so that $\text{epi}(f)$ is convex.

Assume now that $\text{epi}(f)$ is convex. Since $(x,f(x))$ and $(y,f(y))$ are in $\text{epi}(f)$, also $((1-\lambda)x+\lambda y, (1-\lambda)f(x)+\lambda f(y)) \in \text{epi}(f)$. This implies that $f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y)$, so that f is convex.

Exercise 5.10. *By the result above we have that if* $f\hat{A}$ *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: This is a compulsory exercise.

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

Solution: Assume that $f(x) \leq \alpha$, $f(y) \leq \alpha$. Then

$$f((1-\lambda)x+\lambda y) \leq (1-\lambda)f(x)+\lambda f(y) \leq (1-\lambda)\alpha+\lambda\alpha = \alpha.$$

It follows that sublevel sets are convex.

Exercise 5.12. *Verify that the function* $x \to \|x\|_p$ *is positively homogeneous.*

Solution: We have that

$$\|\lambda x\|_p = ((\lambda x_1)^2+\ldots+(\lambda x_n)^2)^{1/p} = \lambda(x_1^2+\ldots+x_n^2)^{1/p} = \lambda\|x\|_p,$$

so that the p-norm is positively homogeneous.

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

Solution: We have that

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

so that f is positively homogeneous.

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 nonzero* z *and that* $f'(x_0; z) = 2z^T x_0$*.*

Solution: We have that

$$\begin{aligned} f'(x_0; z) &= \lim_{t \to 0} \frac{(x_0 + tz)^T (x_0 + tz) - x_0^T x_0}{t} \\ &= \lim_{t \to 0} \frac{2t x_0^T z + t^2 z^T z}{t} \\ &= \lim_{t \to 0} (2x_0^T z + t z^T z) = 2z^T x_0. \end{aligned}$$

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

$$f(x) = x^T A x + c^T x + \alpha$$

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

Solution: The Hessian of f is $2A$, so f is convex if and only if A is positive semidefinite.

Exercise 5.16. *Assume that* $f\hat{A}$ *and* g *are convex functions defined on an interval* I*. Determine which of the functions following functions that are convex or concave:*

(i) λf *where* $\lambda \in \mathbb{R}$*,*
(ii) $\min\{f, g\}$*,*
(iii) $|f|$*.*

Solution:

(i) λf is convex when $\lambda \geq 0$, otherwise it is concave.
(ii) $\min\{f, g\}$ can be neither convex or concave. As an example consider a convex parabola and a line which intersects it at two points.
(iii) $|f|$ can be either convex or concave. If $f(x) = x^2$, then $|f|$ is concave. If $f(x) = x^2 - 4$, defined on $(-2, 2)$, then $|f|$ is concave.

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

Solution: Assume that f is convex. If $g = -f$, $f + g = 0$ is clearly convex, but g is concave (not convex).

If $h = f + g$ is convex, and f is concave, then $g = h - f$ is a sum of two convex functions, so that it is convex.

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

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

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

Solution: Any point in $[a, b]$ can be written on the form $(1 - \lambda)a + \lambda b$, for some $0 \le \lambda \le 1$. We have that

$$\begin{aligned} f(x) &= f((1-\lambda)a + \lambda b) \le (1-\lambda)f(a) + \lambda f(b) \\ &\le (1-\lambda)\max\{f(a), f(b)\} + \lambda \max\{f(a), f(b)\} = \max\{f(a), f(b)\}. \end{aligned}$$

The result follows.

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

Solution: By the previous exercise f is bounded above by the maximum of the endpoint values. Since f is convex it has an increasing left-/right-derivative on (a, b). For $a \le x < z < y \le b$, since the slope function is increasing we have that

$$\frac{f(z) - f(x)}{z - x} \le f'_-(z) \qquad\qquad f'_+(z) \le \frac{f(y) - f(z)}{y - z},$$

so that

$$f(x) \ge -f_-(z)(z - x) + f(z) \qquad\qquad f(y) \ge f'_+(z)(y - z) + f(z).$$

It follows that f is bounded below on both sides of z by linear functions, so that f in total is bounded below.

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: Since $g((1-\lambda)x+\lambda y) \leq (1-\lambda)g(x)+\lambda g(y)$ and f is increasing, we have that

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

which proves that $f \circ g$ is convex.

Exercise 5.21. *Find the optimal solutions of the problem* $\min\{f(x) : a \leq x \leq b\}\hat{A}$ *where* $a < b\hat{A}$ *and* $f : \mathbb{R} \to \mathbb{R}$ *is a differentiable convex function.*

Solution: This occurs when the derivative is 0.

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

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

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

Solution: This is a compulsory exercise

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

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

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

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

and use Corollary 5.1.8.

Solution: g is convex since it is a sum of a convex and an affine function. We know from the exercises then that it has a minimum c. We compute first

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

We then obtain

$$\begin{aligned}
\lim_{x\to c+} \frac{g(x)-g(c))}{x-c} &= \lim_{x\to c+}\left(\frac{f(x)-f(c))}{x-c} - \frac{f(b)-f(a)}{b-a}\frac{x-c}{x-c}\right) \\
&= f'_+(c) - \frac{f(b)-f(a)}{b-a} \\
\lim_{x\to c-} \frac{g(x)-g(c))}{x-c} &= \lim_{x\to c-}\left(\frac{f(x)-f(c))}{x-c} - \frac{f(b)-f(a)}{b-a}\frac{x-c}{x-c}\right) \\
&= f'_-(c) - \frac{f(b)-f(a)}{b-a},
\end{aligned}$$

so that

$$\partial g(c) = [f'_-(c), f'_+(c)] - \frac{f(b)-f(a)}{b-a} = \partial f(c) - \frac{f(b)-f(a)}{b-a}.$$

By Corollary 5.1.8 it follows that $0 \in \partial g(c)$, so that $0 \in \partial f(c) - \frac{f(b)-f(a)}{b-a}$, so that $\frac{f(b)-f(a)}{b-a} \in \partial f(c)$. This proves Theorem 5.1.9.

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 $f \circ g$ (defined on C) is convex.*

Solution: See previous exercise.

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

Solution: We know that $g(x) = x^T Ax$ is convex, and that $f(x) = e^x$ is both convex and increasing. The result thus follows from the previous exercise.

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: Corollary 4.3.4 says that C is the convex hull of its extreme points. Denote them by x_j. We have that

$$f(\sum_j \lambda_j x_j) \leq \sum_j \lambda_j f(x_j) \leq \sum_j \lambda_j \max_j(f(x_j)) = \max_j(f(x_j)),$$

it follows that any maximum value must be attained in an extreme point.

Exercise 5.27. *Let $C \subseteq \mathbb{R}^n$ be a convex set and consider the distance function d_C defined in (3.1), i.e., $d_C(x) = \inf\{\|x - c\| : c \in C\}$. Show that d_C is a convex function.*

Solution: Let x, y be given, and find $x_1, y_1 \in C$ so that $\|x - x_1\| \leq d_C(x) + \epsilon$, $\|y - y_1\| \leq d_C(y) + \epsilon$. Since C is convex, $(1-\lambda)x_1 + \lambda y_1 \in C$. We have that

$$\begin{aligned} d_C((1-\lambda)x + \lambda y) &= \inf\{\|(1-\lambda)x + \lambda y - c\| : c \in C\} \\ &\leq \|(1-\lambda)x + \lambda y - ((1-\lambda)x_1 + \lambda y_1)\| \\ &\leq (1-\lambda)\|x - x_1\| + \lambda\|y - y_1\| \\ &\leq (1-\lambda)(d_C(x) + \epsilon) + \lambda(d_C(y) + \epsilon) \\ &= (1-\lambda)d_C(x) + \lambda d_C(y) + \epsilon. \end{aligned}$$

Since this applies for all ϵ it follows that $d_C((1-\lambda)x+\lambda y) \leq (1-\lambda)d_C(x)+\lambda d_C(y)$ as well, so that d_C is convex.

Exercise 5.28. *Prove Corollary* 6.1.1 *using Theorem* 5.3.5.

Solution: Let x^* be a local minimum Since f is convex, Theorem 5.3.5 says that $f(x) \geq f(x^*) + \nabla f(x^*)^T(x - x^*))$ for all $x \in C$. If $\nabla f(x^*) = 0$ this says that $f(x) \geq f(x^*)$, i.e., x^* is a global minimum. Therefore (iii) implies (ii), and (ii) obviously implies (i).

Assume finally that x^* is a local minimum, and assume for contradiction that $\nabla f(x^*) \neq 0$. Note that $\nabla f(y)^T \nabla f(x^*) > 0$ for y in some neighbourhood of x^* (at least if the gradient is continuous).

It is now better to use the following Taylor formula

$$f(x) = f(x^*) + \nabla f(x^* + t(x - x^*))^T(x - x^*),$$

for some $0 < t < 1$. By choosing $x = x^* - \alpha \nabla f(x^*)$ and α small enough, we get

$$f(x) = f(x^*) - \alpha \nabla f(x^* + t(x - x^*))^T \nabla f(x^*) < f(x^*),$$

which contradicts that we have a local minimum. This proves (iii), and the proof is complete.

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* epi(f) *is a convex set. Let $f : \mathbb{R}^n \to \mathbb{R}$ be convex and consider a supporting hyperplane of* 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: Let $a^T x = \alpha$ be a supporting hyperplane of epi(f) at $(y, f(y))$ (at points (y, t) with $t > f(y)$ the epigraph can not have a supporting hyperplane). Assume that a is scaled so that its last component is 1 (this is always possible when the last component of a is nonzero. Note that the case of a zero last component is not interesting, since this rather constrains the domain of f). This means that the hyperplane can be written as

$$a^T x = (a_1, \ldots, a_n)^T(x_1, \ldots, x_n) + t = (a_1, \ldots, a_n)^T(y_1, \ldots, y_n) + f(y),$$

so that

$$t = -(a_1, \ldots, a_n)^T(x_1 - y_1, \ldots, x_n - y_n) + f(y).$$

Denote this affine function by $h(x)$. Since $a^T x \geq \alpha$ on epi(f), from the above it follows that

$$(a_1, \ldots, a_n)^T(x_1, \ldots, x_n) + f(x) \geq (a_1, \ldots, a_n)^T(y_1, \ldots, y_n) + f(y),$$

so that

$$f(x) \geq -(a_1, \ldots, a_n)^T (x_1 - y_1, \ldots, x_n - y_n) + f(y) = h(x).$$

All this can be more compactly explained in terms of graphs: The graph of the hyperplane lies below the graph of f. The supporting hyperplane is viewed as the tangent plane of the graph.

Chapter 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)*

$$A^T Ax = A^T b.$$

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

Solution: The gradient of $f(x) = \|Ax-b\|^2 = x^T A^T Ax - 2b^T Ax + b^T b$ is $\nabla f(x) = 2A^T A - 2A^T b$. We see that the gradient is zero if and only if $A^T Ax = A^T b$.

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

Solution: Assume that $x_i^* = 0$. Let $x = x^* + e_i \in C$. Then $\nabla f(x^*)^T (x - x^*) = \frac{\partial f}{\partial x_i}(x^*)$, which must be greater than zero.

Assume that $x_i^* > 0$. Then $x = x^* \pm \epsilon e_i \in C$ for ϵ small enough. These two choices give $\pm\epsilon \frac{\partial f}{\partial x_i}(x^*)$ as values for $\nabla f(x^*)^T (x - x^*)$. If both of these are ≥ 0 then clearly $\frac{\partial f}{\partial x_i}(x^*) = 0$.

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

Solution: The constraints can be written as $-x_i \leq 0$, $x_i - p \leq 0$, which have gradients $-e_i$ and e_i, respectively.

Go through all possibilities of active constraints.
If $0 < x_i < p_i$, the gradient equation says that $\frac{\partial f}{\partial x_i} = 0$ If for instance $x_i = 0$, we would add $-\mu_i e_i$ to the gradient equation. This is the same as $\frac{\partial f}{\partial x_i} = \mu_i \geq 0$.
If for instance $x_i = p_i$, we would add $\mu_i e_i$ to the gradient equation. This is the same as $\frac{\partial f}{\partial x_i} = -\mu_i \leq 0$.
We thus arrive at the following: $\frac{\partial f}{\partial x_i} \geq 0$ if $x_i = 0$, $\frac{\partial f}{\partial x_i} \leq 0$ if $x_i = p_i$.

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

$$x \geq O, \ \nabla f(x) \geq O, \text{ and } x_k \cdot \partial f(x)/\partial x_k = 0 \text{ for } k = 1, \ldots, n.$$

Discuss the consequences of these conditions for optimal solutions.

Solution: If $x_i = 0$ then the ith component of the gradient equation is $\frac{\partial f}{\partial x_i} - \mu_i = 0$, so that $\frac{\partial f}{\partial x_i} = \mu_i \geq 0$. Otherwise $\frac{\partial f}{\partial x_i} = 0$. In any case $x_k \cdot \partial f(x)/\partial x_k = 0$. In any case $\nabla f(x) \geq 0$.

Exercise 6.5. *Solve the problem: minimize $(x + 2y - 3)^2 \hat{A}$ for $(x, y) \in \mathbb{R}^2$ and the problem minimize $(x + 2y - 3)^2$ subject to $(x - 2)^2 + (y - 1)^2 \leq 1$.*

Solution: The minimum of the first is clearly 0. with the minimum obtained for any point on the line $x + 2y = 3$. For the second the gradient equation is

$$\begin{pmatrix} 2(x + 2y - 3) \\ 4(x + 2y - 3) \end{pmatrix} + \mu \begin{pmatrix} 2(x - 2) \\ 2(y - 1) \end{pmatrix} = 0.$$

Note first that the gradient of the constraint is zero if $x = 2$, $y = 1$. This gives the candidate $f(2, 1) = 1$ (this point also satisfies the constraint).

If the constraint is not active we must have $x + 2y - 3 = 0$. Any point on this line, and so that $(x - 2)^2 + (y - 1)^2 < 1$, is a candidate for the minumum (the point $(1, 1)$ for instance makes the constraint active, so that the line clearly intersects the interior of the circle satisfie). At all these points f is zero, so that the unconstrained global minimum is attained, so that we can stop here. Nevertheless, let us also see what happens when the constraint is active. We then have that $2(y-1) = 4(x-2)$, i.e., $y = 2x-3$. Inserting this in $(x-2)^2 + (y-1)^2 = 1$ gives

$$(x - 2)^2 + 4(x - 1)^2 = 5(x - 1)^2 = 1,$$

so that $x = 1 \pm \sqrt{5}$, $y = \pm 2/\sqrt{5} - 3$.

$x+2y-3 = \pm 5\sqrt{5}-5 = -5\pm\sqrt{5}$. The minimum value for the two new candidates is clearly $(-5 + \sqrt{5})^2$, but the minimum was already found to be 0 above.

Clearly the global minimum is attained in this exercise, since we minimize over a bounded region.

Exercise 6.6. *Solve the problem: minimize $x^2+y^2-14x-6y$ subject to $x+y \leq 2$, $x + 2y \leq 3$.*

Solution: The gradient equation is

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

Note first that the constraint gradients are linearly independent.

Assume that both constraints are active. This gives $x = y = 1$ and $f(1,1) = -18$.

Assume that only the first constraint is active. We get $2x - 14 = 2y - 6$ and $x + y = 2$, so that we get $x = 3$, $y = -1$, and $f(3,-1) = -26$.

Assume that only the second constraint is active. We get $y = 2x - 11$ and $x + 2y = 3$, so that we get $x = 5$, $y = -1$, and $f(5,-1) = -38$.

Assume finally that no constraints are active. Then $x = 7$, $y = 3$, and $f(7,3) = -58$. By comparing with the other candidates we see that this is the minimum.

We should finally comment that f goes to zero as $x, y \to \infty$. This implies that the minimum we have found also is global.

Exercise 6.7. *Solve the problem: minimize $x^2 - y$ subject to $y - x \geq -2$, $y^2 \leq x$, $y \geq 0$.*

Solution: The gradient equation is

$$\begin{pmatrix} 2x \\ -1 \end{pmatrix} + \mu_1 \begin{pmatrix} 1 \\ -1 \end{pmatrix} + \mu_2 \begin{pmatrix} -1 \\ 2y \end{pmatrix} + \mu_3 \begin{pmatrix} 0 \\ -1 \end{pmatrix} = 0.$$

We see that $x \geq 0$.

From the second component we see that the second constraint must be active and $y > 0$. The third constraint can thus never be active. We are left with two possiblities: The first constraint may or may not be active.

Assume first that the first constraint is active. We then have that $y = x - 2$ and $y^2 = x$, so that $y^2 - y - 2 = 0$. This gives that $y = 2$ or $y = -1$. $y = -1$ can be discarded, so that we obtain the candidate $(4,2)$. We have that $f(4,2) = 14$.

Finally assume that the first constraint is not active. The gradient equation can now be written as

$$\begin{pmatrix} 2x \\ -1 \end{pmatrix} = \mu_2 \begin{pmatrix} 1 \\ -2y \end{pmatrix}.$$

Taking ratios we see that $-2x = -1/2y$ so that $4xy = 1$ (and both x and y must be > 0). Inserting this in $y^2 = x$ gives that $4y^3 = 1$, so that $x = 4^{-2/3}$, $y = 4^{-1/3}$, and

$$f(4^{-2/3}, 4^{-1/3}) = 4^{-4/3} - 4^{-1/3} = -\frac{3}{4} 4^{-1/3} < 0.$$

It follows that this is the constrained minimum.

We should also comment on the possibility of having linearly dependent active constraint gradients. The only problematic part here can be when the first two constraints both are active, but this is covered by the calculations above, which lead to $f(4, 2) = 14$.

We should also comment that the area we minimize over is bounded, so that there actually exists a global minimum, which is the one we have found.

More on the proof of Theorem 1.2 in [2]

$F(G) \subseteq Q$ follows from Lemma 1.1: Since (1.4) holds for all incidence vectors of forests, it also holds for their convex hull, i.e., $F(G) \subseteq Q$. The other way, as Q is compact and convex, Minkowski's theorem yields that Q is the convex hull of its extreme points. According to Chapter 4 in [1], as Q is a polyhedron, vertices and extreme points coincide, and faces and exposed faces are the same. It follows that it is enough to show that any unique optimal solution to a problem on the form $\max\{c^T x : x \in Q\}$ is also in $F(G)$. We will actually show that such a unique optimal solution must be an incidence vector for a forest, which is an even stronger statement.

It is smart to consult section 1.4 in [21] here, where one learns the following greedy algorithm for constructing a maximum weight forest: At step k in the algorithm we have a forest (collection of subtrees), and we add an edge e_k of maximum overall weight (there may be more than one such) which preserves the forest property. The new forest results from joining two trees in the previous forest, and we denote by V_k the vertex set of the resulting tree. We terminate when there are no such edges left with positive weight. Since the later edges added also were candidates at previous iterations, the weights are decreasing: $c(e_1) \geq c(e_2) \geq \cdots c(e_r) > 0$ (the edges e_i are the ones found by the algorithm). This is a crucial point, which is not commented in the proof.

We define $y_S = 0$ for S different from the V_j. With this definition (i) in (1.6) takes the form

$$\sum_{i:e \text{ joins vertices from } V_i} y_{V_i} \geq c_e.$$

In particular, if $e = e_j$, we get

$$y_{V_j} + \sum_{i:V_j \subseteq V_i} y_{V_i} \geq c(e_j) \tag{6.1}$$

We now define the y_{V_j} so that we have equality here (this gives 0 slack for the dual constraints corresponding to e_j). For each j, this gives an upper triangular system in the $\{y_{V_i}\}_{i:V_j \subseteq V_i}$ which defines the y_{V_j} uniquely. As x' is the incidence vector of

$e_1, \ldots, e_r$ we have complementary slackness between the primal variables and the dual slack variables.

Complementary slackness between the dual variables and the primal slack variables: since $y_S = 0$ for all $S \neq V_j$, we only need to show that the primal slack corresponding to the V_j constraint is 0, i.e., that $x(E[V_j) = |V_j| - 1$ for all i. But this follows from the fact that the components constructed by the greedy algorithm are trees.

It remains to argue why we have primal and dual feasibility. Primal feasibility is obvious, so let us prove dual feasibility.

$y_{V_j} \geq 0$ for all j: After all iterations of the greedy algorithm, we end up with a forest where the nodes in any connected component equal a V_j which is maximal in the sense that it is not a subset of another V_i (this means that $I(j) = \emptyset$). If V_j is such, then $y_{V_j} = c(e_j) > 0$ due to (6.1). Now, each V_j is constructed by the algorithm by adding edges, providing a subsystem of (6.1), where only subsets of V_j occur in the summations. If we eliminate $y_{V_j} = c(e_j) > 0$ (back-substitution) in this subsystem, we get a system where the right hand side is $c(e_k) - c(e_j) \geq 0$ (and still decreasing), and where e_j joins V_{j_1} and V_{j_2} to V_j. The sketched procedure can now be repeated to show that $y_{V_{j_1}} \geq 0$, $y_{V_{j_2}} \geq 0$, and so on, proving that all $y_{V_j} \geq 0$.

For dual feasibility we also need to explain why

$$\sum_{i:e \text{ joins vertices from } V_i} y_{V_i} \geq c_e$$

when e is different from the e_i added by the algorithm (when e equals some e_i we have equality here). There are two different cases to consider:

1. e joins two different components of the forest we end up with, we must have that $c(e) \leq 0$ (otherwise the forest is not maximum weight), and the equation follows since all $y_{V_i} \geq 0$.
2. e joins two vertices in the same component of the forest we end up with. At some point in the algorithm, the two end nodes of e are joined into the same component by means of one edge e_k. As e was also a candidate for this join, we must have that $c(e_k) \geq c(e)$ by maximality of $c(e_k)$ in the algorithm. But then

$$\sum_{i:e \text{ joins vertices from } V_i} y_{V_i} = \sum_{i:e_k \text{ joins vertices from } V_i} y_{V_i} = c(e_k) \geq c(e).$$

   where we used that e joints vertices from V_i if and only if e_k does.

Dual feasibility follows.

From complementary slackness it now follows that x' is optimal for the primal problem. As $\bar{x}$ also is optimal, they must be equal due to the uniqueness assumption. As x' is the incidence vector of a forest, the result follows.

Chapter 2 in [2]

Regarding the final statement in proposition 2.2, I believe it should be that, if P is pointed **and** rational, nonempty polyhedron, then P is integral if and only if each vertex is integral. To see why: When it is pointed, we can write $P = \text{conv}(\text{ext}(P)) + \text{cone}(Z)$ according to Proposition 4.3.3. Since P also is assumed rational, we can choose integral generators for Z as in Theorem 2.1.

$\Rightarrow$: IF P is integral, $P = \text{conv}(P \cap Z^n)$. According to Exercise 4.5, all vertices of P are contained in $P \cap Z^n$, so that they are integral (this direction does not require the polyhedron to be rational).

$\Leftarrow$: If each vertex is integral, then $\text{conv}(\text{ext}(P)) \subseteq P_I$. Since P and P_I have the same characteristic cone according to Theorem 2.1, we have that

$$P = \text{conv}(\text{ext}(P)) + \text{cone}(Z) \subseteq P_I + \text{cone}(Z) = P_I.$$

Thus $P \subseteq P_I$, so that P is integral.

Add details in Proposition 2.3: Note that the vertices are rational since the polyhedron is rational. Since the polyhedron is not integral, one of the vertices is not integral. Since it is rational, it must be fractional. That there is a $\bar{c}$ as described is not explained. Note that each vertex is the unique intersection of n planes. Consider a supporting hyperplane at the optimal solution. We can slightly change its normal vector $\bar{c}$ to a rational vector (and therefore also an integral vector), while maintaining the supporting hyperplane property. $\bar{c}$ must be changed so that its inner products with the normal vectors of the planes do not change in sign. For the same reason we can find c'.

Chapter 4 in [2]

The deductions on the bottom of page 54

Let us go carefully through the deductions on the bottom of page 54. The first two lines follow by definitions, while the third line follows frm the fact that the maximum of a convex function in a polytope can be found in one of the vertices of the polytope. The fourth line, introducing η, is a simple rewriting, while the fifth line simply moves the terms from side to side. The fifth line is an LP in the variables η and λ. The coefficient vector in the objective is $(1, 0, \ldots, 0)$. With $X = \begin{pmatrix} x_1 & \cdots & x_{|K|} \end{pmatrix}^T$, and 1_n the column vector with all ones, the inequalities can be rephrased as

$$\begin{pmatrix} 1_n & (A^2 X - \begin{pmatrix} b^2 & \cdots b^2 \end{pmatrix})^T \end{pmatrix} \begin{pmatrix} \eta \\ \lambda \end{pmatrix} \geq X^T c.$$

The dual LP is thus (with variables denoted by μ_k) to maximize $(X^T c)^T \mu = c^T \sum_{k \in K} \mu^k x^k$ subject to

$$\begin{pmatrix} 1_n^T \\ A^2 X - \begin{pmatrix} b^2 & \cdots b^2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mu^1 \\ \mu^2 \\ \vdots \end{pmatrix} \leq \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

where there is equality in the first inequality since there was no requirement on η to be nonnegative. This first equality can be rephrased as $\sum_{k \in K} \mu^k = 1$. The other inequalities can be rephrased as

$$\begin{pmatrix} A^2 X - \begin{pmatrix} b^2 & \cdots b^2 \end{pmatrix} \end{pmatrix} \begin{pmatrix} \mu^1 \\ \mu^2 \\ \vdots \end{pmatrix} \leq 0,$$

i.e.,

$$A^2 \sum_{k \in K} \mu^k x^k \leq b^2 \sum_{k \in K} \mu_k = b^2.$$

This is the statement on the sixth line. Since $\sum_{k \in K} \mu^k x^k$ is a general element in P_I^1, the last line follows.

The details at the top of page 65

Let us also clarify the details at the top of page 65. By adding the degree constraints we have that $\sum_{v\in H} x(\delta(v)) = 2|H|$. Now, each edge in $E(H)$ contributes twice in this sum, while each $x \in \delta(H)$ contributes once, so that $2x(E(H)) + x(\delta(H)) = 2|H|$, so that $x(E(H)) + \frac{1}{2}x(\delta(H)) = |H|$. Adding that $-\frac{1}{2}x_e \leq 0$ for $e \in \delta(H) \setminus (\cup_i E(T_i))$ we obtain

$$x(E(H)) + \frac{1}{2}\sum_{i=1}^{k} x(E(T_i) \cap \delta(H)) \leq |H|, \tag{6.2}$$

where we used that we have the disjoint union

$$\delta(H) = (\delta(H) \setminus \cup_i E(T_i)) \cup (\cup_i E(T_i) \cap \delta(H)).$$

Now note that we also have the disjoint union

$$E(T_i) = E(T_i \cap H) \cup E(T_i \setminus H) \cup \delta(H) \cap E(T_i),$$

and that

1. An edge in $E(T_i \cap H)$ contributes both in (ii) and (iv).
2. An edge in $E(T_i \setminus H)$ contributes both in (ii) and (iii).
3. An edge in $\delta(H) \cap E(T_i)$ contributes both in (ii) and (6.2).

Therefore, if (i), (ii), and (iii) are scaled by $\frac{1}{2}$, and these three are added for all i with (6.2), the left hand side will become $x(E(H)) + \sum_{i=1}^{k} x(E(T_i))$. On the right side we obtain

$$\begin{aligned} &|H| + \frac{1}{2}\left(\sum_{i=1}^{k}(|T_i| - 1 + |T_i \setminus H| - 1 + |T_i \cap H| - 1)\right) \\ &= |H| + \frac{1}{2}\sum_{i=1}^{k}(2|T_i| - 3) \\ &= |H| + \sum_{i=1}^{k}(|T_i| - 1) - k/2. \end{aligned}$$

The use of Farkas lemma

Let us also comment on how Farkas lemma is used in (4.17). Farkas lemma (Theorem 3.2.5) states that. $Ax = b$ has a solution $x \geq 0$ if and only if for each

y, $y^T A \geq 0$ implies $y^T b \geq 0$. Alternatively this says that $Ax = b$ has a solution $x \geq 0$ if and only if for each y, $y^T A \leq 0$ implies $y^T b \geq 0$.

Applied to the matrix $\begin{pmatrix} A \\ e \end{pmatrix}$ and vector $\begin{pmatrix} \bar{x} \\ 1 \end{pmatrix}$, the first statement is equivalent to $x \in \text{conv}(T)$. $x \notin \text{conv}(T)$ is thus equivalent to that there exists a y so that $y^T \begin{pmatrix} A \\ e \end{pmatrix} \leq 0$ and $y^T \begin{pmatrix} \bar{x} \\ 1 \end{pmatrix} > 0$. Writing $y = \begin{pmatrix} a \\ -b \end{pmatrix}$, this says that $a^T A - be \leq 0$ and $a^T \bar{x} - b > 0$, which is the staement in the book.

The matrix A here has nonnegative entries, as well as $\bar{x}$. The reason for using $-b$ instead of b is that the equations secure equivalence between positivity of a and that of b.

Exercises from [2]

Exercise 4.1: The coordinate change $x_j = \frac{b}{a_j} y_j$ turns the problem into maximizing $b \sum_{j=1}^n \frac{c_j}{a_j} y_j$ under the constraints $\sum_{j=1}^n y_j \leq 1$, and $0 \leq y_j \leq \frac{a_j}{b}$. Clearly we must choose $y_1 = \min\left(1, \frac{a_1}{b}\right)$, i.e., $x_1 = \min\left(1, \frac{b}{a_1}\right)$.

If $a_1 \geq b$ we are done and must choose $x_2 = \cdots = x_n = 0$. Otherwise $x_1 = 1$ and the problem can be rephrased as minimizing subject to $\sum_{j=2}^n a_j x_j \leq b - a_1$.

If $a_2 \geq b - a_1$ we are again done. Otherwise $x_2 = 1$, and we get the system $\sum_{j=3}^n a_j x_j \leq b - a_1 - a_2$. We continue this procedure until for some k,

$$a_k \geq b - a_1 - \cdots - a_{k-1}, \tag{6.3}$$

i.e., $\sum_{i=1}^{k-1} a_i < b \leq \sum_{i=1}^k a_i$. We obtain the optimal solution

$$\left(1, \ldots, 1, \frac{b - \sum_{i=1}^{k-1} a_i}{a_k}, 0, \ldots, 0\right), \tag{6.4}$$

and the optimal value is $c_1 + \ldots + c_{k-1} + \frac{c_k(b - \sum_{i=1}^{k-1} a_i)}{a_k}$

The dual problem is: Minimize $by + y_1 + \cdots + y_n$ subject to $a_i y + y_i \geq c_i$, $y \geq 0$.

This can also be written as $y_i \geq c_i - a_i y$.

With $y \geq 0$ fixed, clearly the minimum is obtained by choosing $y_j = \max(c_j - a_j y, 0)$. Let now y be chosen so that $\frac{c_i}{a_i} \geq y \geq \frac{c_{i+1}}{a_{i+1}}$.

Since $c_j - a_j y = a_j(\frac{c_j}{a_j} - y)$ we have that

$$\max(c_j - a_j y, 0) = \begin{cases} c_j - a_j y & j \leq i \\ 0 & j > i. \end{cases}$$

The minimum is thus

$$by + \sum_{j=1}^{i} (c_j - a_j y) = \left(b - \sum_{j=1}^{i} a_j\right) y + \sum_{j=1}^{i} c_j.$$

This states that the minimization problem can be stated as finding the minimum of the piecewise linear function defined as $\left(b - \sum_{j=1}^{i} a_j\right) y + \sum_{j=1}^{i} c_j$ on $\left(\frac{c_{i+1}}{a_{i+1}}, \frac{c_i}{a_i}\right)$.

Thus the minimum occurs when $\sum_{j=1}^{k-1} a_j \leq b \leq \sum_{j=1}^{k} a_{j+1}$, and at $y = \frac{c_k}{a_k}$, which gives

$$\left(b - \sum_{j=1}^{k-1} a_j\right) \frac{c_k}{a_k} + \sum_{j=1}^{k-1} c_j,$$

which is the same expression we found when solving the primal problem.

The vertices are obtained by collecting all possible maxima when the c_i are varied (this gives all exposed faces, which consitute the vertices for polyhedra). By permuting the c_i in particular we see that all vectors obtained by permuting the entries in (6.4) also are vertices. Let us take a look at how many possible such permutations there are. The actual number depends on the k we found in (6.3). If the middle number in (6.4) is in $(0, 1)$, the number of vertices is $n\binom{n-1}{k}$, where the binomial coefficient comes from the number of ways to place k zeros among $n - 1$ numbers.

Exercise 4.2: No. That $v(R(u)) > v(Q)$ implies that that the LP relaxation has an optimal value which is not integral. The optimal node in the enumeration tree may still be below u, so that we cannot prune.

Exercise 4.3: Assume that x represents a Hamilton tour. Then clearly (i) holds. Also, if W is as described in (ii), along the Hamilton tour we must pass at least twice between an element in W and an element outside W, so that $x(\delta(W)) \geq 2$.

On the other hand, suppose (i) and (ii) are fulfilled. (i) secures that x passes through each vertex twice, so that the only possibility is to have one tour, or several subtours. Assume the latter, and let W be the node set of one of those subtours. Then no edges are entering or leaving W, so that $x(\delta(W)) = 0$, contradicting (ii). It follows that x represents a Hamilton tour.

Exercise 4.4:

Exercise 4.5: The constraints $x(E[S]) \leq |S| - 1$ (for all $S \subseteq V$, $S \neq \emptyset$, $S \neq V$), $x_e \geq 0$ forces x to be the incidence vector of a forest.

We should now add the degree constraints $x(\delta(v)) \leq b_v$ (b_v is the constrained degree at v), for all $v \in V$.

Finally we should add constraints enforcing a tree. For this we can add the constraints $x(\delta(W)) \geq 1$ (for all $W \subseteq V$, $W \neq \emptyset$, $W \neq V$).

Exercise 4.6:

Exercise 4.7:

Exercise 4.8:

Exercises from [3]

Exercise 1: We have that

$$\begin{aligned}\sum_{v\in V}\mathrm{div}_x(v) &= \sum_{v\in V}\left(\sum_{e\in\delta^+(v)} x(e) - \sum_{e\in\delta^-(v)} x(e)\right)\\ &= \sum_{e=(u,v)\in E}(x(e)-x(e)) = 0,\end{aligned}$$

since $(u,v)\in\delta^+(u)$, and $(u,v)\in\delta^-(v)$.

Exercise 2: This does not give meaning, since one sums together edges sets. I suppose what is meant is to show that $x(\delta^-(S)) = x(\delta^+(S))$, i.e., that the total inflow to S equals the total outflow from S. The total inflow to S can be written $\sum_{v\in S}\sum_{e\in\delta^-(v)} x(e)$, while the total outflow can be written $\sum_{v\in S}\sum_{e\in\delta^+(v)} x(e)$. The flow balance equations say that $\sum_{e\in\delta^-(v)} x(e) = \sum_{e\in\delta^+(v)} x(e)$ for any v. Adding for $v\in S$ we obtain the desired equality.

We have that $\sum_{e\in\delta^+(v)} x(e) = \sum_{e\in\delta^-(v)} x(e)$ for any $v\in V$. Summing over $v\in S$ we obtain

$$\sum_{v\in S}\sum_{e\in\delta^+(v)} x(e) = \sum_{e\in\delta^-(v)} x(e) = \sum_{e\in\delta^+(S)} x(e) = \sum_{e\in\delta^-(S)} x(e).$$

Exercise 3: Clearly x satisfies the bounds $l\le x\le u$ if and only if $0\le x'\le u-l$. If $\mathrm{div}_x(v)=0$, then

$$\begin{aligned}\mathrm{div}'_x(v) &= \sum_{e\in\delta^+(v)} x'(e) - \sum_{e\in\delta^-(v)} x'(e)\\ &= \sum_{e\in\delta^+(v)} x(e) - \sum_{e\in\delta^-(v)} x(e) - \sum_{e\in\delta^+(v)} l(e) + \sum_{e\in\delta^-(v)} l(e)\\ &= -\sum_{e\in\delta^+(v)} l(e) + \sum_{e\in\delta^-(v)} l(e),\end{aligned}$$

which gives the required divergence of x'.

Exercise 4: We have that $\sum_{x \in V} \text{div}_x(v) = 0$. Also $\text{div}_x(v) = 0$ for $v \neq s, t$. It follows that $\text{div}_x(s) = -\text{div}_x(t)$. If one assumes (as one usually does) that there is no edge entering s and no edge leaving t, we obtain

$$\begin{aligned} \text{div}_x(s) &= \sum_{e \in \delta^+(s)} x(e) - \sum_{e \in \delta^-(s)} x(e) = \sum_{e \in \delta^+(s)} x(e) = \text{val}(x) \\ -\text{div}_x(t) &= \sum_{e \in \delta^-(t)} x(e) - \sum_{e \in \delta^+(t)} x(e) = \sum_{e \in \delta^-(t)} x(e), \end{aligned}$$

and the result follows.

Exercise 5: The value of a flow is a continuous (in fact linear) function. The constraints $0 \leq x \leq c$ gives a closed and bounded (i.e., compact) set. It follows from the extreme value theorem that there is a maximum flow. Any linear programming problem with box constraints has a maximum.

Exercise 6: Let x be a maximum flow. D_x thus contains no x-augmenting path. As in the proof of Theorem 1.5 we define $S(x)$ as the set of all vertices v to which we can find an augmenting sv-path in D_x, and we define the cut $K = \delta^+(S(x)$. The proof shows that the capacity of this cut equals the value of the (flow), and the result follows from Lemma 1.3.

Exercise 7:

Exercise 8:

(a) The outgoing edges from s are (s, v) with $b(v) > 0$, and their capacities are $b(v)$. It follows that $\text{val}(x) = \sum_{v \in V^+} x(s, v) \leq \sum_{v \in V^+} b(v)$.

(b) Assume that a flow in D satisfies $\text{div}_x = b$ and $0 \leq x \leq c$. Define a flow x' in D' by expanding x so that $x'(s, v) = b(v)$ for $v \in V^+$, and $x'(v, t) = -b(v)$ for $v \in V^-$. The value of x' is $\sum_{v \in V^+} b(v) = M$, so that the value of the maximum flow is M. x' also satisfies the balancing equations for $v \in V^+$, and for $v \in V-$,a dn thus for all v.

The other way, if the value of a maximum st-flow in D' is M, all the edges $x'(s, v)$ must be at capacity $c'(s, v) = b(v)$ for $v \in V^+$. From the flow balance equations it follows that $\text{div}(x)$ is $b(v)$ for all v, and the result follows.

(c) One simply restricts x' from E' to E.

Exercise 9: The Ford Fulkerson algorithm starts with the zero flow, which is integral. Due to integral capacities, the ϵ found by the algorithm will at each step be integral, so that each step produces a new integral flow. After all steps we thus end with an integral flow.

Exercise 10: Since the capacities are integral, there exists a maximum flow which is integral (Theorem 1.6). This implies that all edges have either unit- or zero flow. Due to flow conservation, each vertex has the same number of incoming unit flow edges as outgoing unit flow edges. Start by following edge-disjoint edges with unit flow from s, all the way to t (if this was impossible, we would have a contradiction to flow balancing). If one removes this st-path from D, one still have a flow of the same type. One can in this way take out one edge-disjoint st-path at a time, unitl there are noe edges left.

Exercise 11: From the previous exercise it is clear that the maximum number of edge-disjoint st-paths equals the maximum flow, which again equals the capacity of the minimum cut, which is $\sum_{e \in K} c(e) = |K|$.

Exercise 12: The upper and lower bounds should be defined as follows:

- For $e = (u_i, v_j)$, we set $l(e) = \lfloor a_{ij} \rfloor$, $r(e) = \lceil a_{ij} \rceil$.
- For $e = (s, u_i)$, we set $l(e) = \lfloor r_i \rfloor$, $r(e) = \lceil r_i \rceil$.
- For $e = (v_j, t)$, we set $l(e) = \lfloor s_j \rfloor$, $r(e) = \lceil s_j \rceil$.

Hoffman's circulation theorem ensures that, if a circulation exists in thsi graph, an integral circulation also exists. Such an integral circulation represents a solution to the matrix rounding problem.

Exercise 13: See the proof of Exercise 4.28.

Exercise 14: That every permutation matrix is a vertex follows directly from the previous exercise. Integral matrices in Ω_n must have exactly one 1 in each row and column, the rest zeros (in order for a column/row summing to one). But this is equivalent to being a permutation matrix.