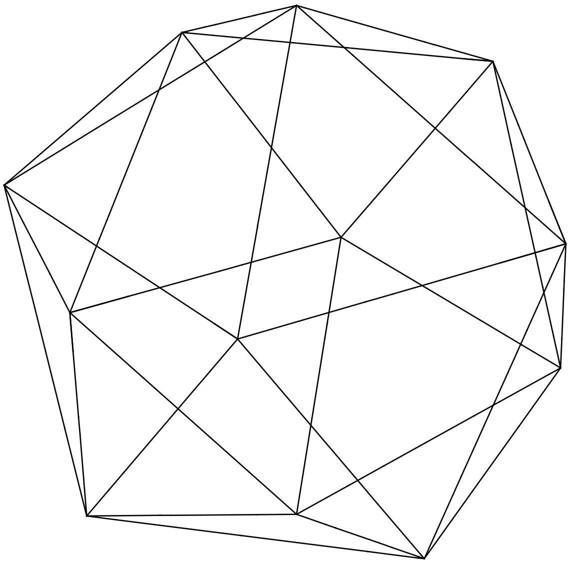
LECTURES ON CONVEX SETS

Niels Lauritzen



> *There is no trivial mathematics,*
> *there are only trivial mathematicians!*
> *A mathematician is trivial if he or she*
> *believes that there exists trivial mathematics.*
> D. Zeilberger

NIELS LAURITZEN
DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF AARHUS
DENMARK

EMAIL: `niels@imf.au.dk`
URL: `http://home.imf.au.dk/niels/`

March 2009
March 2010

Contents

Preface

The theory of convex sets is a vibrant and classical field of modern mathematics with rich applications in economics and optimization.

The material in these notes is introductory starting with a small chapter on linear inequalities and Fourier-Motzkin elimination. The aim is to show that linear inequalities can be solved and handled with very modest means. At the same time you get a real feeling for what is going on, instead of plowing through tons of formal definitions before encountering a triangle.

The more geometric aspects of convex sets are developed introducing notions such as extremal points and directions, convex cones and their duals, polyhedra and separation of convex sets by hyperplanes.

The emphasis is primarily on polyhedra. In a beginning course this seems to make a lot of sense: examples are readily available and concrete computations are feasible with the double description method ([7], [3]).

The main theoretical result is the theorem of Minkowski and Weyl on the structure of polyhedra. We stop short of introducing general faces of polyhedra.

I am grateful to Markus Kiderlen and Jesper Funch Thomsen for very useful comments on these notes.

Århus, March 2009.

I have incorporated many changes suggested by the first teaching from these notes. I am especially grateful to Jonas Andersen Seebach for his careful reading of the notes and for several valuable suggestions.

Århus, March 2010.

Notation

- $\mathbb{Z}$ denotes the set of integers $\ldots, -2, -1, 0, 1, 2, \ldots$ and $\mathbb{R}$ the set of real numbers.
- $\mathbb{R}^n$ denotes the set of all n-tuples (or vectors) $\{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in \mathbb{R}\}$ of real numbers. This is a vector space over $\mathbb{R}$ — you can add vectors and multiply them by a real number. The zero vector $(0, \ldots, 0) \in \mathbb{R}^n$ will be denoted 0.
- Let $u, v \in \mathbb{R}^n$. The inequality $u \leq v$ means that $\leq$ holds for every coordinate. For example $(1, 2, 3) \leq (1, 3, 4)$, since $1 \leq 1, 2 \leq 3$ and $3 \leq 4$. But $(1, 2, 3) \leq (1, 2, 2)$ is not true, since $3 \nleq 2$.
- When $x \in \mathbb{R}^n, b \in \mathbb{R}^m$ and A is an $m \times n$ matrix, the m inequalities, $Ax \leq b$, are called a system of linear inequalities. If $b = 0$ this system is called homogeneous.
- Let $u, v \in \mathbb{R}^n$. Viewing u and v as $n \times 1$ matrices, the matrix product $u^t\, v$ is nothing but the usual inner product of u and v. In this setting, $u^t\, u = |u|^2$, where $|u|$ is the usual length of u.

Chapter 1

Introduction

You probably agree that it is quite easy to solve the equation

$$2x = 4. \tag{1.1}$$

This is an example of a linear equation in one variable having the unique solution $x = 2$. Perhaps you will be surprised to learn that there is essentially no difference between solving a simple equation like (1.1) and the more complicated system

$$\begin{aligned} 2x + y + z &= 7 \\ x + 2y + z &= 8 \\ x + y + 2z &= 9 \end{aligned} \tag{1.2}$$

of linear equations in x, y and z. Using the first equation $2x + y + z = 7$ we solve for x and get

$$x = (7 - y - z)/2. \tag{1.3}$$

This may be substituted into the remaining two equations in (1.2) and we get the simpler system

$$\begin{aligned} 3y + z &= 9 \\ y + 3z &= 11 \end{aligned}$$

of linear equations in y and z. Again using the first equation in this system we get

$$y = (9 - z)/3 \tag{1.4}$$

to end up with the simple equation

$$8z = 24.$$

This is a simple equation of the type in (1.1) giving $z = 3$. Now $z = 3$ gives $y = 2$ using (1.4). Finally $y = 2$ and $z = 3$ gives $x = 1$ using (1.3). So solving a seemingly complicated system of linear equations like (1.2) is really no more difficult than solving the simple equation (1.1).

We wish to invent a similar method for solving systems of linear inequalities like

$$\begin{array}{rcrcl} x & & & \geq & 0 \\ x & + & 2y & \leq & 6 \\ x & + & y & \geq & 2 \\ x & - & y & \leq & 3 \\ & & y & \geq & 0 \end{array} \tag{1.5}$$

1.1 Linear inequalities

Let us start out again with the simplest case: linear inequalities in just one variable. Take as an example the system

$$\begin{array}{rcrcl} 2x & + & 1 & \leq & 7 \\ 3x & - & 2 & \leq & 4 \\ -x & + & 2 & \leq & 3 \\ x & & & \geq & 0 \end{array} \tag{1.6}$$

This can be rewritten to

$$\begin{array}{rcrcl} & & x & \leq & 3 \\ & & x & \leq & 2 \\ -1 & \leq & x & & \\ 0 & \leq & x & & \end{array}$$

This system of linear inequalities can be reduced to just two linear inequalities:

$$\begin{array}{rcrcl} & & x & \leq & \min(2,3) = 2 \\ \max(-1,0) = 0 & \leq & x & & \end{array}$$

or simply $0 \leq x \leq 2$. Here you see the real difference between linear equations and linear inequalities. When you reverse $=$ you get $=$, but when you reverse $\leq$ after multiplying by -1, you get $\geq$. This is why solving linear inequalities is more involved than solving linear equations.

1.1.1 Two variables

Let us move on to the more difficult system of linear inequalities given in (1.5). Perhaps the most intuitive way of approaching (1.5) is through a drawing. First we sketch the five lines (see Figure 1.1)

$$\begin{array}{rcrcl} x & & & = & 0 \\ x & + & 2y & = & 6 \\ x & + & y & = & 2\ . \\ x & - & y & = & 3 \\ & & y & = & 0 \end{array}$$

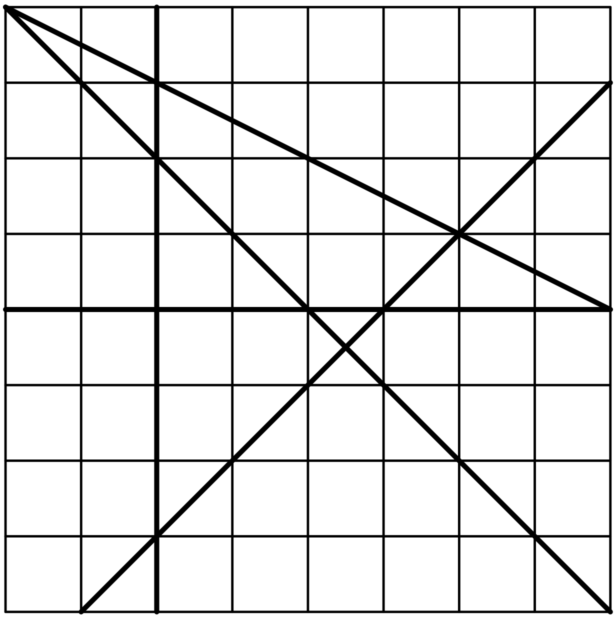


Figure 1.1: The lines in (1.5).

For each line we pick a point to decide which half plane to color e.g. we need to color below the line $x + 2y = 6$, since the corresponding inequality is $x + 2y \leq 6$ and (for example) $0 + 2 \cdot 0 < 6$. Figure 1.2 shows the intersection of these half planes (colorings).

We are aiming for a more effective way of handling the solutions of (1.5). Our sketching techniques are not of much use solving for example 17 linear inequalities with 5 unknowns.

We seek inspiration in the solution of the system (1.2) of linear equations and try to isolate or eliminate x. How should we do this? We rewrite (1.5) to

$$\begin{array}{rcccl} 0 & \leq & x & & \\ & & x & \leq & 6-2y \\ 2-y & \leq & x & & \\ & & x & \leq & 3+y \end{array}$$

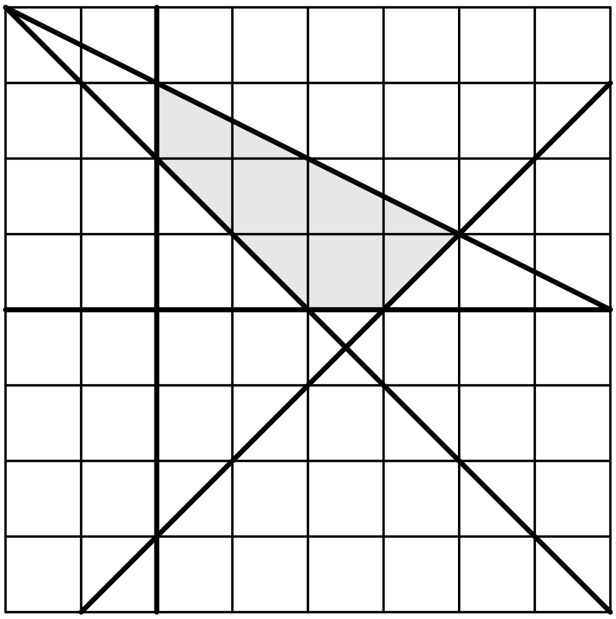


Figure 1.2: Sketch of solutions to (1.5).

along with the inequality $0 \leq y$ which does not involve x. Again, just like in one variable, this system can be reduced just two inequalities

$$\begin{aligned} x &\leq \min(6-2y, 3+y) \\ \max(0, 2-y) &\leq x \end{aligned} \tag{1.7}$$

carrying the inequality $0 \leq y$ along. Here is the trick. We can eliminate x from the two inequalities in (1.7) to get the system

$$\max(0, 2-y) \leq \min(6-2y, 3+y). \tag{1.8}$$

You can solve (1.7) in x and y if and only if you can solve (1.8) in y. If you think about it for a while you will realize that (1.8) is equivalent to the following four inequalities

$$\begin{aligned} 0 &\leq 6 - 2y \\ 0 &\leq 3 + y \\ 2 - y &\leq 6 - 2y \\ 2 - y &\leq 3 + y \end{aligned}$$

These inequalities can be solved just like we solved (1.6). We get

$$\begin{array}{rcccl} & & y & \leq & 3 \\ -3 & \leq & y & & \\ & & y & \leq & 4 \\ -\frac{1}{2} & \leq & y & & \\ 0 & \leq & y & & \end{array}$$

where the inequality $y \geq 0$ from before is attached. This system can be reduced to

$$0 \leq y \leq 3$$

Through a lot of labor we have proved that two numbers x and y solve the system (1.5) if and only if

$$\begin{array}{rcccl} 0 & \leq & y & \leq & 3 \\ \max(0, 2-y) & \leq & x & \leq & \min(6-2y, 3+y) \end{array}$$

If you phrase things a bit more geometrically, we have proved that the projection of the solutions to (1.5) on the y-axis is the interval $[0, 3]$. In other words, if x, y solve (1.5), then $y \in [0, 3]$ and if $y \in [0, 3]$, there exists $x \in \mathbb{R}$ such that x, y form a solution to (1.5). You should compare this very algebraic solution with the sketch in Figure 1.2.

Now we are ready to enter into the abstract setting. Keep the example of this section in mind.

1.2 Polyhedra

Let us introduce precise definitions. A linear inequality in n variables $x_1, \dots, x_n$ is an inequality of the form

$$a_1 x_1 + \cdots + a_n x_n \leq b,$$

where $a_1, \dots, a_n, b \in \mathbb{R}$.

DEFINITION 1.2.1

The set of solutions

$$P = \left\{ (x_1, \dots, x_n) \in \mathbb{R}^n \;\middle|\; \begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & \leq & b_1 \\ & & & & & \vdots & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & \leq & b_m \end{array} \right\}$$

to a system

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & \leq & b_1 \\ & & & & & \vdots & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & \leq & b_m \end{array}$$

of finitely many linear inequalities (here a_{ij} and b_i are real numbers) is called a polyhedron. A bounded polyhedron is called a polytope.

A polyhedron is an extremely important special case of a convex subset of $\mathbb{R}^n$. We will return to the definition of a convex subset in the next chapter.

The proof of the following important theorem may look intimidating at first. If this is so, then take a look at §1.1.1 once again. Do not get fooled by the slick presentation here. In its purest form the result goes back to a paper by Fourier[1] from 1826 (see [2]). It is also known as *Fourier-Motzkin elimination*, simply because you are eliminating the variable x_1 and because Motzkin[2] rediscovered it in his dissertation "Beiträge zur Theorie der linearen Ungleichungen" with Ostrowski[3] in Basel, 1933 (not knowing the classical paper by Fourier). The main result in the dissertation of Motzkin was published much later in [6].

THEOREM 1.2.2
Consider the projection $\pi : \mathbb{R}^n \to \mathbb{R}^{n-1}$ given by

$$\pi(x_1, \dots, x_n) = (x_2, \dots, x_n).$$

If $P \subset \mathbb{R}^n$ is a polyhedron, then $\pi(P) \subset \mathbb{R}^{n-1}$ is a polyhedron.

Proof. Suppose that P is the set of solutions to

$$\begin{array}{ccccccc} a_{11}x_1 & + & \cdots & + & a_{1n}x_n & \leq & b_1 \\ & & & & & \vdots & \\ a_{m1}x_1 & + & \cdots & + & a_{mn}x_n & \leq & b_m \end{array}$$

We divide the m inequalities into

$$\begin{aligned} G &= \{i \mid a_{i1} > 0\} \\ Z &= \{i \mid a_{i1} = 0\} \\ L &= \{i \mid a_{i1} < 0\} \end{aligned}$$

Inequality number i reduces to

$$x_1 \leq a'_{i2}x_2 + \cdots + a'_{in}x_n + b'_i,$$

---

[1] Jean Baptiste Joseph Fourier (1768 – 1830), French mathematician.

[2] Theodore Samuel Motzkin (1908 – 1970), American mathematician.

[3] Alexander Markowich Ostrowski (1893 – 1986), Russian mathematician

if $i \in G$ and to

$$a'_{j2}x_2 + \cdots + a'_{jn}x_n + b'_j \le x_1,$$

if $j \in L$, where $a'_{ik} = -a_{ik}/a_{i1}$ and $b'_i = b_i/a_{i1}$ for $k = 2, \ldots, n$. So the inequalities in L and G are equivalent to the two inequalities

$$\begin{aligned}\max\left(a'_{i2}x_2 + \cdots + a'_{in}x_n + b'_i \,\middle|\, i \in L\right) &\le x_1 \\ &\le \min\left(a'_{j2}x_2 + \cdots + a'_{jn}x_n + b'_j \,\middle|\, j \in G\right).\end{aligned}$$

Now $(x_2, \ldots, x_n) \in \pi(P)$ if and only if there exists x_1 such that $(x_1, \ldots, x_n) \in P$. This happens if and only if $(x_2, \ldots, x_n)$ satisfies the inequalities in Z and

$$\max\left(a'_{i2}x_2 + \cdots + a'_{in}x_n + b'_i \,\middle|\, i \in L\right) \le \min\left(a'_{j2}x_2 + \cdots + a'_{jn}x_n + b'_j \,\middle|\, j \in G\right)$$

This inequality expands to the system of $|L|\,|G|$ inequalities in $x_2, \ldots, x_n$ consisting of

$$a'_{i2}x_2 + \cdots + a'_{in}x_n + b'_i \le a'_{j2}x_2 + \cdots + a'_{jn}x_n + b'_j$$

or rather

$$(a'_{i2} - a'_{j2})x_2 + \cdots + (a'_{in} - a'_{jn})x_n \le b'_j - b'_i$$

where $i \in L$ and $j \in G$. Adding the inequalities in Z (where x_1 is not present) we see that $\pi(P)$ is the set of solutions to a system of $|L|\,|G| + |Z|$ linear inequalities i.e. $\pi(P)$ is a polyhedron. □

To get a feeling for Fourier-Motzkin elimination you should immediately immerse yourself in the exercises. Perhaps you will be surprised to see that Fourier-Motzkin elimination can be applied to optimize the production of vitamin pills.

1.3 Exercises

(1) Sketch the set of solutions to the system

$$\begin{array}{rcrcr} 2x & + & y & \geq & 2 \\ 3x & + & y & \leq & 9 \\ -x & + & 2y & \leq & 4 \\ & & y & \geq & 0 \end{array} \tag{1.9}$$

of linear inequalities. Carry out the elimination procedure for (1.9) as illustrated in §1.1.1.

(2) Let

$$P = \left\{ (x, y, z) \in \mathbb{R}^3 \;\middle|\; \begin{array}{rcrcrcr} -x & - & y & - & z & \leq & 0 \\ 3x & - & y & - & z & \leq & 1 \\ -x & + & 3y & - & z & \leq & 2 \\ -x & - & y & + & 3z & \leq & 3 \end{array} \right\}$$

and $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ be given by $\pi(x, y, z) = (y, z)$.

(i) Compute $\pi(P)$ as a polyhedron i.e. as the solutions to a set of linear inequalities in y and z.

(ii) Compute $\eta(P)$, where $\eta : \mathbb{R}^3 \to \mathbb{R}$ is given by $\eta(x, y, z) = x$.

(iii) How many integral points[4] does P contain?

(3) Find all solutions $x, y, z \in \mathbb{Z}$ to the linear inequalities

$$\begin{array}{rcrcrcr} -x & + & y & - & z & \leq & 0 \\ & - & y & + & z & \leq & 0 \\ & & & - & z & \leq & 0 \\ x & & & - & z & \leq & 1 \\ & & y & & & \leq & 1 \\ & & & & z & \leq & 1 \end{array}$$

by using Fourier-Motzkin elimination.

(4) Consider the polyhedron

$$P = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \;\middle|\; \begin{array}{rcrcr} x & + & 2y & \leq & 3 \\ 2x & + & y & \leq & 3 \\ x & & & \geq & 0 \\ & & y & \geq & 0 \end{array} \right\}.$$

---

[4] An integral point is simply a vector (x, y, z) with $x, y, z \in \mathbb{Z}$

Show that

$$m = \min\left\{ x+y \;\middle|\; \begin{pmatrix} x \\ y \end{pmatrix} \in P \right\} = \min\left\{ c \;\middle|\; x+y=c \quad \text{and} \quad \begin{pmatrix} x \\ y \end{pmatrix} \in P \right\}.$$

Use this observation and Fourier-Motzkin elimination to compute m by finding the smallest c such that

$$\begin{array}{rcrcl} x & + & y & = & c \\ x & + & 2y & \leq & 3 \\ 2x & + & y & \leq & 3 \\ x & & & \geq & 0 \\ & & y & \geq & 0 \end{array}$$

has a solution.

(5) Let $P \subseteq \mathbb{R}^n$ be a polyhedron and $c \in \mathbb{R}^n$. Define the polyhedron $P' \subseteq \mathbb{R}^{n+1}$ by

$$P' = \{ \begin{pmatrix} m \\ x \end{pmatrix} \mid c^t\, x = m,\; x \in P,\; m \in \mathbb{R} \},$$

where $c^t x$ is the (usual inner product) matrix product of c transposed (a $1 \times n$ matrix) with x (an $n \times 1$ matrix) giving a 1×1 matrix (also known as a real number!).

- (i) What does this have to do with Exercise 4?
- (ii) Show how projection onto the m-coordinate (and Fourier-Motzkin elimination) in P' can be used to solve the (linear programming) problem of finding $x \in P$, such that $c^t x$ is minimal (or proving that such an x does not exist).
- (iii) Let P denote the polyhedron from Exercise 2. You can see that

$$(0,0,0),\; (-1, \frac{1}{2}, \frac{1}{2}) \in P$$

have values 0 and -1 on their first coordinates, but what is the minimal first coordinate of a point in P?

(6) A vitamin pill P is produced using two ingredients M_1 and M_2. The pill needs to satisfy four constraints for the vital vitamins V_1 and V_2. It must contain at least 6 mg and at most 15 mg of V_1 and at least 5 mg and at most 12 mg of V_2. The ingredient M_1 contains 3 mg of V_1 and 2 mg of V_2 per gram. The ingredient M_2 contains 2 mg of V_1 and 3 mg of V_2 per gram:

	V_1	V_2
M_1	3	2
M_2	2	3

Let x denote the amount of M_1 and y the amount of M_2 (measured in grams) in the production of a vitamin pill. Write down a system of linear inequalities in x and y describing the constraints above.

We want a vitamin pill of minimal weight satisfying the constraints. How many grams of M_1 and M_2 should we mix? Describe how Fourier-Motzkin elimination can be used in solving this problem (hint: take another look at Exercise 5).

Chapter 2

Basics

The definition of a convex subset is quite elementary and profoundly important. It is surprising that such a simple definition can be so far reaching.

2.1 Convex subsets of $\mathbb{R}^n$

Consider two vectors $u, v \in \mathbb{R}^n$. The line through u and v is given parametrically as

$$f(\lambda) = u + \lambda(v - u) = (1 - \lambda)u + \lambda v,$$

where $\lambda \in \mathbb{R}$. Notice that $f(0) = u$ and $f(1) = v$. Let

$$[u, v] = \{f(\lambda) \mid \lambda \in [0, 1]\} = \{(1 - \lambda)u + \lambda v \mid \lambda \in [0, 1]\}$$

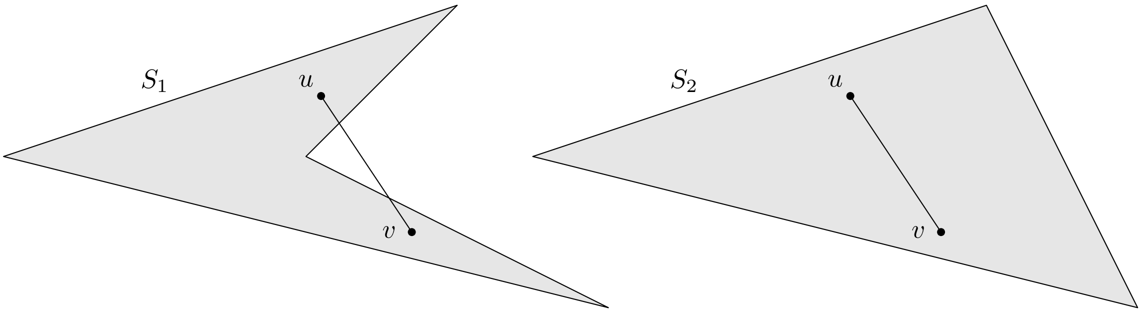
denote the line segment between u and v.

DEFINITION 2.1.1
A subset $S \subseteq \mathbb{R}^n$ is called convex if

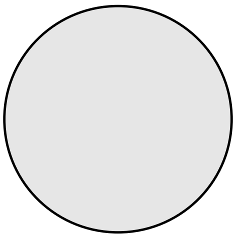
$$[u, v] \subset S$$

for every $u, v \in S$.

EXAMPLE 2.1.2
Two subsets $S_1, S_2 \subseteq \mathbb{R}^2$ are sketched below. Here S_2 is convex, but S_1 is not.



The triangle S_2 in Example 2.1.2 is a prominent member of the special class of polyhedral convex sets. Polyhedral means "intersection of finitely many half-spaces". A non-polyhedral convex set is for example a disc in the plane:



These non-polyhedral convex sets are usually much more complicated than their polyhedral cousins especially when you want to count the number of integral points[1] inside them. Counting the number $N(r)$ of integral points inside a circle of radius r is a classical and very difficult problem going back to Gauss[2]. Gauss studied the error term $E(r) = |N(r) - \pi r^2|$ and proved that $E(r) \leq 2\sqrt{2}\pi r$.

Counting integral points in polyhedral convex sets is difficult but theoretically much better understood. For example if P is a convex polygon in the plane with integral vertices, then the number of integral points inside P is given by the formula of Pick[3] from 1899:

$$|P \cap \mathbb{Z}^2| = \text{Area}(P) + \frac{1}{2}\,\text{B}(P) + 1,$$

where $\text{B}(P)$ is the number of integral points on the boundary of P. You can easily check this with a few examples. Consider for example the convex polygon P in Figure 2.1.

By subdivision into triangles it follows that $\text{Area}(P) = \frac{55}{2}$. Also, by an easy count we get $\text{B}(P) = 7$. Therefore the formula of Pick shows that

$$|P \cap \mathbb{Z}^2| = \frac{55}{2} + \frac{1}{2} \cdot 7 + 1 = 32.$$

The polygon contains 32 integral points. You should inspect Figure 2.1 to check that this is true!

---

[1] Point with coordinates in $\mathbb{Z}$.

[2] Carl Friedrich Gauss (1777–1855), German mathematician. Probably the greatest mathematician that ever lived.

[3] Georg Alexander Pick (1859–1942), Austrian mathematician.

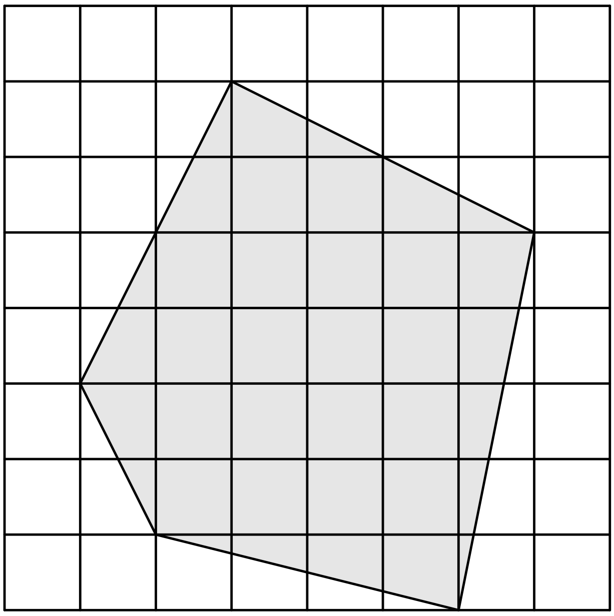
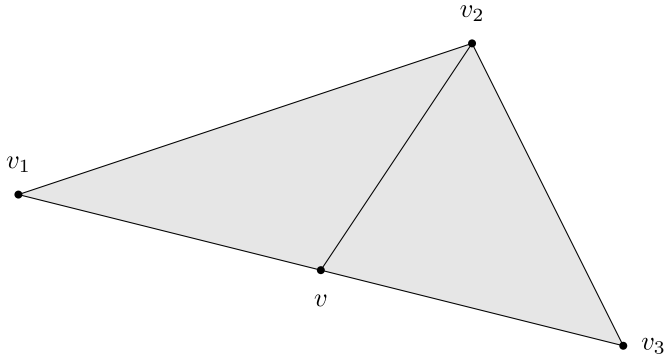


Figure 2.1: Counting integral points in a polygon: Pick's formula.

2.2 The convex hull

Take a look at the traingle T below.



We have marked the vertices v_1, v_2 and v_3. Notice that T coincides with the points on line segments from v_2 to v, where v is a point on $[v_1, v_3]$ i.e.

$$\begin{aligned} T &= \{(1-\lambda)v_2 + \lambda((1-\mu)v_1 + \mu v_3) \mid \lambda \in [0,1], \mu \in [0,1]\} \\ &= \{(1-\lambda)v_2 + \lambda(1-\mu)v_1 + \lambda\mu v_3 \mid \lambda \in [0,1], \mu \in [0,1]\} \end{aligned}$$

Clearly $(1-\lambda) \geq 0$, $\lambda(1-\mu) \geq 0$, $\lambda\mu \geq 0$ and

$$(1-\lambda) + \lambda(1-\mu) + \lambda\mu = 1.$$

It is not too hard to check that (see Exercise 5)

$$T = \{\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 \mid \lambda_1, \lambda_2, \lambda_3 \geq 0,\ \lambda_1 + \lambda_2 + \lambda_3 = 1\}.$$

With this example in mind we define the *convex hull* of a finite set of vectors.

DEFINITION 2.2.1

Let $v_1, \dots, v_m \in \mathbb{R}^n$. Then we let

$$\operatorname{conv}(\{v_1, \dots, v_m\}) := \{\lambda_1 v_1 + \cdots + \lambda_m v_m \mid \lambda_1, \dots, \lambda_m \geq 0,\ \lambda_1 + \cdots + \lambda_m = 1\}.$$

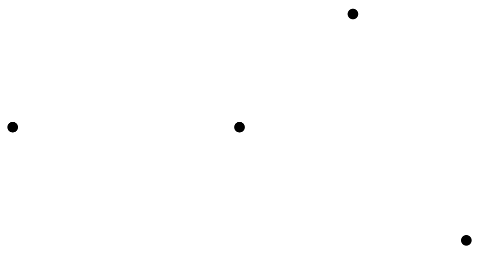
We will occasionally use the notation

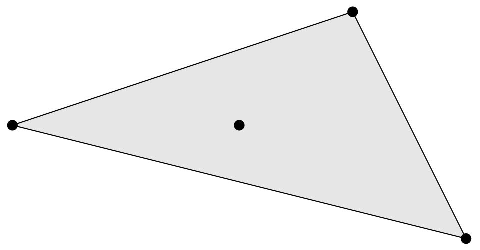
$$\operatorname{conv}(v_1, \dots, v_m)$$

for $\operatorname{conv}(\{v_1, \dots, v_m\})$.

EXAMPLE 2.2.2

To get a feeling for convex hulls, it is important to play around with (lots of) examples in the plane. Below you see a finite subset of points in the plane. To the right you have its convex hull.





This convex hull is a triangle. In suitable coordinates the four points are

$$\begin{pmatrix} 0 \\ 1 \end{pmatrix}, \quad \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \quad \begin{pmatrix} 2 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

You can check that the third point is contained in the convex hull of the other points, since

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \frac{3}{7} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \frac{2}{7} \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \frac{2}{7} \begin{pmatrix} 4 \\ 0 \end{pmatrix}.$$

You may wonder where the coefficients $\frac{3}{7}, \frac{2}{7}$ and $\frac{2}{7}$ came from. In fact, the coefficients λ_1, λ_2 and λ_3 in

$$\begin{pmatrix} 2 \\ 1 \end{pmatrix} = \lambda_1 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} 3 \\ 2 \end{pmatrix} + \lambda_3 \begin{pmatrix} 4 \\ 0 \end{pmatrix}$$

coming from Definition 2.2.1, must solve the system

$$\begin{array}{rcrcrcl} & & 3\lambda_2 & + & 4\lambda_3 & = & 2\\ \lambda_1 & + & 2\lambda_2 & & & = & 1\\ \lambda_1 & + & \lambda_2 & + & \lambda_3 & = & 1 \end{array}$$

of linear equations. This makes them very explicit and computable.

PROPOSITION 2.2.3
If $S \subseteq \mathbb{R}^n$ is a convex subset and $v_1, \dots, v_m \in S$, then

$$\operatorname{conv}(v_1, \dots, v_m) \subseteq S.$$

Proof. We must show that

$$\lambda_1 v_1 + \cdots + \lambda_{m-1} v_{m-1} + \lambda_m v_m \in S,$$

where $v_1, \dots, v_m \in S$, $\lambda_1, \dots, \lambda_m \geq 0$ and

$$\lambda_1 + \cdots + \lambda_{m-1} + \lambda_m = 1.$$

For $m = 2$ this is the definition of convexity. The general case is proved using induction on m. For this we may assume that $\lambda_m \neq 1$. Then the identity

$$\lambda_1 v_1 + \cdots + \lambda_{m-1} v_{m-1} + \lambda_m v_m =$$
$$(\lambda_1 + \cdots + \lambda_{m-1}) \left(\frac{\lambda_1}{(\lambda_1 + \cdots + \lambda_{m-1})} v_1 + \cdots + \frac{\lambda_{m-1}}{(\lambda_1 + \cdots + \lambda_{m-1})} v_{m-1} \right)$$
$$+ \lambda_m v_m$$

and the convexity of S proves the induction step. Notice that the induction step is the assumption that we already know $\operatorname{conv}(v_1, \dots, v_{m-1}) \subseteq S$ for $m - 1$ vectors $v_1, \dots, v_{m-1} \in S$. ☐

The next step is of course to prove that the convex hull is a convex subset in the sense of Definition 2.1.1. Before we do that let us introduce the more general concept of the convex hull of an arbitrary subset of $\mathbb{R}^n$.

DEFINITION 2.2.4
If $X \subseteq \mathbb{R}^n$, then

$$\operatorname{conv}(X) = \bigcup_{\substack{m \geq 1 \\ v_1, \dots, v_m \in X}} \operatorname{conv}(v_1, \dots, v_m).$$

PROPOSITION 2.2.5

(i) *The convex hull $\operatorname{conv}(v_1, \dots, v_m) \subseteq \mathbb{R}^n$ is a convex subset, where $v_1, \dots, v_m \in \mathbb{R}^n$.*

(ii) The convex hull conv(X) *of an arbitrary subset* $X \subseteq \mathbb{R}^n$ *is a convex subset containing* X.

Proof. Suppose that $u, v \in \operatorname{conv}(v_1, \dots, v_m)$ i.e.

$$\begin{aligned} u &= \lambda_1 v_1 + \cdots + \lambda_m v_m \\ v &= \mu_1 v_1 + \cdots + \mu_m v_m, \end{aligned}$$

where $\lambda_1, \dots, \lambda_m, \mu_1, \dots, \mu_m \geq 0$ and

$$\lambda_1 + \cdots + \lambda_m = \mu_1 + \cdots + \mu_m = 1.$$

Then (for $0 \leq \alpha \leq 1$)

$$\alpha u + (1 - \alpha) v = (\alpha \lambda_1 + (1 - \alpha)\mu_1) v_1 + \cdots + (\alpha \lambda_m + (1 - \alpha)\mu_m) v_m$$

where

$$\begin{aligned} &(\alpha \lambda_1 + (1 - \alpha)\mu_1) + \cdots + (\alpha \lambda_m + (1 - \alpha)\mu_m) = \\ &\alpha(\lambda_1 + \cdots + \lambda_m) + (1 - \alpha)(\mu_1 + \cdots + \mu_m) = \alpha + (1 - \alpha) = 1. \end{aligned}$$

This proves that $\operatorname{conv}(v_1, \dots, v_m)$ is a convex subset. Consider now $u, v \in \operatorname{conv}(X)$. By definition of $\operatorname{conv}(X)$,

$$\begin{aligned} u &\in \operatorname{conv}(u_1, \dots, u_r) \\ v &\in \operatorname{conv}(v_1, \dots, v_s) \end{aligned}$$

for $u_1, \dots, u_r, v_1, \dots, v_s \in X$. Therefore u and v both belong to the convex subset $\operatorname{conv}(u_1, \dots, u_r, v_1, \dots, v_s)$ and

$$\alpha u + (1 - \alpha) v \in \operatorname{conv}(u_1, \dots, u_r, v_1, \dots, v_s) \subseteq \operatorname{conv}(X)$$

where $0 \leq \alpha \leq 1$, proving that $\operatorname{conv}(X)$ is a convex subset. □

Intuitively, the convex hull of a subset X is the smallest convex subset containing X, but what exactly does it mean for a subset S to be the smallest convex subset containing X? A minimum requirement is that if C is a convex subset with $C \supseteq X$, then $C \supseteq S$: the smallest convex subset containing X is a subset of every convex subset containing X!

THEOREM 2.2.6

Let $X \subseteq \mathbb{R}^n$. *Then* conv(X) *is the smallest convex subset containing* X.

Proof. Let S be a convex subset containing X and let $v_1, \dots, v_m \in X$. Then $\operatorname{conv}(v_1, \dots, v_m) \subseteq S$ by Proposition 2.2.3. Therefore $\operatorname{conv}(X) \subseteq S$ proving that $\operatorname{conv}(X)$ is the smallest convex subset containing X. □

The smallest convex subset containing $X \subseteq \mathbb{R}^n$ is also the intersection of the convex subsets containing X. What do we mean by an aritrary intersection of subsets of $\mathbb{R}^n$?

The intersection of finitely many subsets $X_1, \dots, X_m$ of $\mathbb{R}^n$ is

$$X_1 \cap \cdots \cap X_m = \{x \in \mathbb{R}^n \mid x \in X_i, \text{ for every } i = 1, \dots, m\}$$

– the subset of elements common to every $X_1, \dots, X_m$. This concept makes perfectly sense for subsets X_i indexed by an arbitrary, not necessarily finite set I. The definition is practically the same:

$$\bigcap_{i \in I} X_i = \{x \in \mathbb{R}^n \mid x \in X_i, \text{ for every } i \in I\}.$$

Here $(X_i)_{i \in I}$ is called a family of subsets. In the above finite case,

$$\{X_1, \dots, X_m\} = (X_i)_{i \in I}$$

with $I = \{1, \dots, m\}$. With this out of the way we can state the following.

PROPOSITION 2.2.7
The intersection of a family of convex subsets of $\mathbb{R}^n$ *is a convex subset.*

The proof of this proposition is left to the reader as an exercise in the definition of the intersection of subsets. This leads us to the following "modern" formulation of conv(X):

PROPOSITION 2.2.8
The convex hull $\operatorname{conv}(X)$ *of a subset* $X \subseteq \mathbb{R}^n$ *equals the convex subset*

$$\bigcap_{C \in I_X} C,$$

where $I_X = \{C \subseteq \mathbb{R}^n \mid C \text{ convex subset and } X \subseteq C\}$.

This characterization of the convex hull is beautiful for its brevity but downright ugly for computational and pedagogical purposes.

2.3 Extremal points

DEFINITION 2.3.1
A point z *in a convex subset* $C \subseteq \mathbb{R}^n$ *is called extreme or an extremal point if*

$$z \in [x, y] \implies z = x \quad \text{or} \quad z = y$$

for every $x, y \in C$. *The set of extremal points in* C *is denoted* $\operatorname{ext}(C)$.

So an extremal point in a convex subset C is a point, which is not located in the interior of a line segment in C. This is the crystal clear mathematical definition of the intuitive notion of a vertex or a corner of set.

Perhaps the formal aspects are better illustrated in getting rid of superfluous vectors in a convex hull

$$X = \operatorname{conv}(v_1, \dots, v_N)$$

of finitely many vectors $v_1, \dots, v_N \in \mathbb{R}^n$. Here a vector v_j fails to be extremal if and only if it is contained in the convex hull of the other vectors (it is superfluous). It is quite instructive to carry out this proof (see Exercise 11).

Notice that only one of the points in triangle to the right in Example 2.2.2 fails to be extremal. Here the extremal points consists of the three corners (vertices) of the triangle.

2.4 The characteristic cone for a convex set

To every convex subset of $\mathbb{R}^n$ we have associated its characteristic cone[4] of *(infinite) directions*:

DEFINITION 2.4.1
A vector $d \in \mathbb{R}^n$ is called an (infinite) direction for a convex set $C \subseteq \mathbb{R}^n$ if

$$x + \lambda\, d \in C$$

for every $x \in C$ and every $\lambda \geq 0$. The set of (infinite) directions for C is called the characteristic cone for C and is denoted $\operatorname{ccone}(C)$.

Just as we have extremal points we have the analogous notion of *extremal directions.* An infinite direction d is extremal if $d = d_1 + d_2$ implies that $d = \lambda d_1$ or $d = \lambda d_2$ for some $\lambda > 0$.

Why do we use the term cone for the set of infinite directions? You can check that $d_1 + d_2 \in \operatorname{ccone}(C)$ if $d_1, d_2 \in \operatorname{ccone}(C)$ and $\lambda d \in \operatorname{ccone}(C)$ if $\lambda \geq 0$ and $d \in \operatorname{ccone}(C)$.

This leads to the next section, where we define this extremely important class of convex sets.

2.5 Convex cones

A mathematical theory is rarely interesting if it does not provide tools or algorithms to compute with the examples motivating it. A very basic question is:

---

[4] Some people use the term *recession cone* instead of characteristic cone.

how do we decide if a vector v is in the convex hull of given vectors $v_1, \dots, v_m$. We would like to use linear algebra i.e. the theory of solving systems of linear equations to answer this question.To do this we need to introduce convex cones.

DEFINITION 2.5.1
A (convex) cone in $\mathbb{R}^n$ *is a subset* $C \subseteq \mathbb{R}^n$ *with* $x + y \in C$ *and* $\lambda x \in C$ *for every* $x, y \in C$ *and* $\lambda \geq 0$.

It is easy to prove that a cone is a convex subset. Notice that any $d \in C$ is an infinite direction for C and $\text{ccone}(C) = C$. An *extremal ray* of a convex cone C is just another term for an extremal direction of C.

In analogy with the convex hull of finitely many points, we define the cone generated by $v_1, \dots, v_m \in \mathbb{R}^n$ as

DEFINITION 2.5.2

$$\text{cone}(v_1, \dots, v_m) := \{\lambda_1 v_1 + \cdots + \lambda_m v_m \mid \lambda_1, \dots, \lambda_m \geq 0\}.$$

Clearly $\text{cone}(v_1, \dots, v_m)$ is a cone. Such a cone is called *finitely generated*. There is an intimate relation between finitely generated cones and convex hulls. This is the content of the following lemma.

LEMMA 2.5.3

$$v \in \text{conv}(v_1, \dots, v_m) \iff \begin{pmatrix} v \\ 1 \end{pmatrix} \in \text{cone}\left(\begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} v_m \\ 1 \end{pmatrix} \right).$$

Proof. In Exercise 16 you are asked to prove this. If you look once more at Example 2.2.2 you will almost surely see the proof. ☐

EXAMPLE 2.5.4
A triangle T is the convex hull of 3 non-collinear points

$$(x_1, y_1), (x_2, y_2), (x_3, y_3)$$

in the plane. Lemma 2.5.3 says that a given point $(x, y) \in T$ if and only if

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \in \text{cone}\left(\begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}, \begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix}, \begin{pmatrix} x_3 \\ y_3 \\ 1 \end{pmatrix} \right). \tag{2.1}$$

You can solve this problem using linear algebra! Testing (2.1) amounts to solving the system

$$\begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{2.2}$$

of linear equations. So $(x, y) \in T$ if and only if the unique solution to (2.2) has $\lambda_1 \geq 0$, $\lambda_2 \geq 0$ and $\lambda_3 \geq 0$. Let us experiment with a few concrete numbers. You can plot the points

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix} \text{ and } \begin{pmatrix} 5 \\ 3 \end{pmatrix} \in \mathbb{R}^2$$

and realize that their convex hull, T, is a very thin triangle. From the drawing it can be difficult to decide if a given point is inside the triangle. Here the 3×3 matrix from (2.2) helps. In this case it is

$$A = \begin{pmatrix} 0 & 2 & 5 \\ 0 & 1 & 3 \\ 1 & 1 & 1 \end{pmatrix}.$$

A little computation shows that

$$A^{-1} = \begin{pmatrix} -2 & 3 & 1 \\ 3 & -5 & 0 \\ -1 & 2 & 0 \end{pmatrix}.$$

Let us check if $v_1 = (1, \frac{11}{20})^t \in T$ and $v_2 = (4, 2)^t \in T$. In the case of v_1, we see that

$$A^{-1} \begin{pmatrix} 1 \\ \frac{11}{20} \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{13}{20} \\ \frac{1}{4} \\ \frac{1}{10} \end{pmatrix}.$$

This is the solution $(\lambda_1, \lambda_2, \lambda_3)^t$ to the linear equations in (2.2). In this case, the solution satisfies $\lambda_1 \geq 0, \lambda_2 \geq 0$ and $\lambda_3 \geq 0$. Therefore $v_1 \in T$. As an added bonus you also see that

$$\begin{pmatrix} 1 \\ \frac{11}{20} \end{pmatrix} = \frac{13}{20} \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \frac{1}{4} \begin{pmatrix} 2 \\ 1 \end{pmatrix} + \frac{1}{10} \begin{pmatrix} 5 \\ 3 \end{pmatrix}.$$

For v_2 we get

$$A^{-1} \begin{pmatrix} 4 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 2 \\ 0 \end{pmatrix}.$$

Here the first coordinate of the solution to (2.2) is negative and therefore $v_2 \notin T$.

Why does (2.2) have a unique solution? This question leads to the concept of affine independence. How do we express precisely that three points in the plane are non-collinear?

2.6 Affine independence

DEFINITION 2.6.1

A set $\{v_1, \dots, v_m\} \subseteq \mathbb{R}^n$ *is called affinely independent if*

$$\left\{ \begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} v_m \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^{n+1}$$

is linearly independent.

As a first basic example notice that $-1, 1 \in \mathbb{R}$ are affinely independent (but certainly not linearly independent).

LEMMA 2.6.2

Let $v_1, \dots, v_m \in \mathbb{R}^n$*. Then the following conditions are equivalent.*

(i) $v_1, \dots, v_m$ *are affinely independent.*

(ii) If

$$\lambda_1 v_1 + \cdots + \lambda_m v_m = 0$$

and $\lambda_1 + \cdots + \lambda_m = 0$*, then* $\lambda_1 = \cdots = \lambda_m = 0$*.*

(iii)

$$v_2 - v_1, \dots, v_m - v_1$$

are linearly independent in $\mathbb{R}^n$*.*

Proof. Proving $(i) \implies (ii)$ is the definition of affine independence. For $(ii) \implies (iii)$, assume for $\mu_2, \dots, \mu_m \in \mathbb{R}$ that

$$\mu_2(v_2 - v_1) + \cdots + \mu_m(v_m - v_1) = 0.$$

Then

$$\lambda_1 v_1 + \cdots + \lambda_m = 0$$

with $\lambda_2 = \mu_2, \dots, \lambda_m = \mu_m$ and

$$\lambda_1 = -(\mu_2 + \cdots + \mu_m).$$

In particular, $\lambda_1 + \cdots + \lambda_m = 0$ and it follows that $\lambda_1 = \cdots = \lambda_m = 0$ and thereby $\mu_2 = \cdots = \mu_m = 0$. For $(iii) \implies (i)$ assume that

$$\lambda_1 \begin{pmatrix} v_1 \\ 1 \end{pmatrix} + \cdots + \lambda_m \begin{pmatrix} v_m \\ 1 \end{pmatrix} = 0.$$

Then

$$\begin{aligned} 0 = \lambda_1 v_1 + \cdots + \lambda_m v_m &= \lambda_1 v_1 + \cdots + \lambda_m v_m - (\lambda_1 + \cdots + \lambda_m) v_1 \\ &= \lambda_2 (v_2 - v_1) + \cdots + \lambda_m (v_m - v_1) \end{aligned}$$

By assumption this implies that $\lambda_2 = \cdots = \lambda_m = 0$ and thereby also $\lambda_1 = 0$.□

DEFINITION 2.6.3
The convex hull

$$\operatorname{conv}(v_1, \ldots, v_{m+1})$$

of $m + 1$ *affinely independent vectors is called an m-simplex.*

So a 0-simplex is a point, a 1-simplex is a line segment, a 2-simplex is a triangle, a 3 simplex is a tetrahedron, ... In a sense, simplices (= plural of simplex) are building blocks for all convex sets. In Figure 2.2 you see a picture of (the edges of) a tetrahedron.

2.7 Carathéodory's theorem

A finitely generated cone $\operatorname{cone}(v_1, \ldots, v_m)$ is called *simplicial* if $v_1, \ldots, v_m$ are linearly independent vectors. These cones are usually easy to manage.

Every finitely generated cone is the union of simplicial cones. This is the content of the following very important result essentially due to Carathéodory[5].

THEOREM 2.7.1 (Carathéodory)
Let $v_1, \ldots, v_m \in \mathbb{R}^n$. *If*

$$v \in \operatorname{cone}(v_1, \ldots, v_m)$$

then v belongs to the cone generated by a linearly independent subset of $\{v_1, \ldots, v_m\}$.

Proof. Suppose that

$$v = \lambda_1 v_1 + \cdots + \lambda_m v_m$$

with $\lambda_1, \ldots, \lambda_m > 0$ and $v_1, \ldots, v_m$ linearly dependent. The linear dependence means that there exists $\mu_1, \ldots, \mu_m \in \mathbb{R}$ not all zero such that

$$\mu_1 v_1 + \cdots + \mu_m v_m = 0. \tag{2.3}$$

---

[5] Constantin Carathéodory (1873–1950), Greek mathematician.

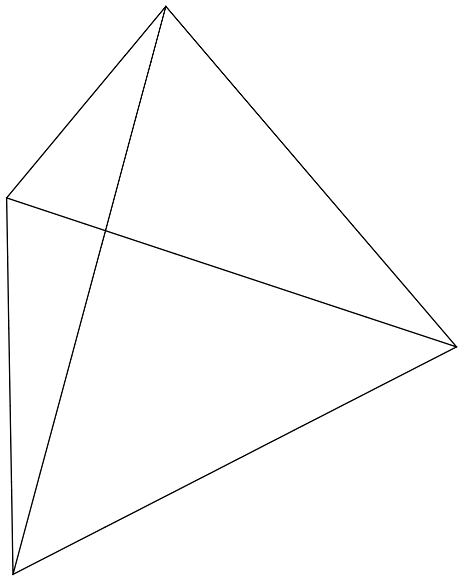


Figure 2.2: A 3-simplex (tetrahedron in $\mathbb{R}^3$) – the convex hull of four affinely independent points.

We may assume that at least one $\mu_i > 0$ multiplying (2.5) by -1 if necessary. But

$$\begin{aligned} v = v - \theta \cdot 0 &= v - \theta(\mu_1 v_1 + \cdots + \mu_m v_m) \\ &= (\lambda_1 - \theta\mu_1)v_1 + \cdots + (\lambda_m - \theta\mu_m)v_m. \end{aligned} \tag{2.4}$$

Let

$$\begin{aligned} \theta^* &= \max\{\theta \geq 0 \mid \lambda_i - \theta\mu_i \geq 0, \text{ for every } i = 1, \ldots, m\} \\ &= \min\left\{ \frac{\lambda_i}{\mu_i} \mid \mu_i > 0,\ i = 1, \ldots, m \right\}. \end{aligned}$$

When you insert θ^* into (2.4), you discover that v also lies in the subcone generated by a proper subset of $\{v_1, \ldots, v_m\}$. Now keep repeating this procedure until the proper subset consists of linearly independent vectors. Basically we are varying θ in (2.4) ensuring non-negative coefficients for $v_1, \ldots, v_m$ until "the first time" we reach a zero coefficient in front of some v_j. This (or these) v_j is (are) deleted from the generating set. Eventually we end up with a linearly independent subset of vectors from $\{v_1, \dots, v_m\}$. □

A special case of the following corollary is: if a point in the plane is in the convex hull of 17364732 points, then it is in the convex hull of at most 3 of these points. When you play around with points in the plane, this seems very obvious. But in higher dimensions you need a formal proof of the natural generalization of this!

COROLLARY 2.7.2

Let $v_1, \dots, v_m \in \mathbb{R}^n$. *If*

$$v \in \operatorname{conv}(v_1, \dots, v_m)$$

then v *belongs to the convex hull of an affinely independent subset of* $\{v_1, \dots, v_m\}$.

Proof. If $v \in \operatorname{conv}(v_1, \dots, v_m)$, then

$$\begin{pmatrix} v \\ 1 \end{pmatrix} \in \operatorname{cone}\Big(\begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} v_m \\ 1 \end{pmatrix} \Big)$$

by Lemma 2.5.3. Now use Theorem 2.7.1 to conclude that

$$\begin{pmatrix} v \\ 1 \end{pmatrix} \in \operatorname{cone}\Big(\begin{pmatrix} u_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} u_k \\ 1 \end{pmatrix} \Big),$$

where

$$\Big\{ \begin{pmatrix} u_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} u_k \\ 1 \end{pmatrix} \Big\} \subseteq \Big\{ \begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} v_m \\ 1 \end{pmatrix} \Big\}.$$

is a linearly independent subset. By Lemma 2.5.3 we get $u \in \operatorname{conv}(u_1, \dots, u_k)$. But by definition

$$\begin{pmatrix} u_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} u_k \\ 1 \end{pmatrix}$$

are linearly independent if and only if $u_1, \dots, u_k$ are affinely independent. □

EXAMPLE 2.7.3

The Carathéodory results may appear quite abstract until you study a few concrete examples. Here is a typical example. Consider the convex hull

$$C = \operatorname{conv}\Big(\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 \\ 0 \end{pmatrix} \Big).$$

The identity

$$\begin{pmatrix}2\\1\end{pmatrix} = \tfrac{1}{11}\begin{pmatrix}1\\0\end{pmatrix} + \tfrac{1}{11}\begin{pmatrix}0\\1\end{pmatrix} + \tfrac{5}{11}\begin{pmatrix}1\\2\end{pmatrix} + \tfrac{4}{11}\begin{pmatrix}4\\0\end{pmatrix}$$

shows that

$$v = \begin{pmatrix}2\\1\end{pmatrix} \in C.$$

Corollary 2.7.2 says however that v is in the convex hull of at most 3 of these 4 vectors (why?). Let us emulate the proof of Corollary 2.7.2 in a concrete computation. The first step is to find a linear dependence between the four vectors

$$\begin{pmatrix}1\\0\\1\end{pmatrix}, \quad \begin{pmatrix}0\\1\\1\end{pmatrix}, \quad \begin{pmatrix}1\\2\\1\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}4\\0\\1\end{pmatrix}$$

in $\mathbb{R}^3$. You should really compare this with the beginning of the proof of Corollary 2.7.2! Here is a linear dependence:

$$5\begin{pmatrix}1\\0\\1\end{pmatrix} - 6\begin{pmatrix}0\\1\\1\end{pmatrix} + 3\begin{pmatrix}1\\2\\1\end{pmatrix} - 2\begin{pmatrix}4\\0\\1\end{pmatrix} = \begin{pmatrix}0\\0\\0\end{pmatrix}. \tag{2.5}$$

Just to be sure your linear algebra is not founded on a too abstract setting, let us show how to find such a linear dependence. A linear dependence in the above case can be seen as a non-zero solution to the equations

$$\begin{array}{ccccccccc} \lambda_1 & & & + & \lambda_3 & + & 4\lambda_4 & = & 0 \\ & & \lambda_2 & + & 2\lambda_3 & & & = & 0 \\ \lambda_1 & + & \lambda_2 & + & \lambda_3 & + & \lambda_4 & = & 0 \end{array} \tag{2.6}$$

in $\lambda_1, \lambda_2, \lambda_3$ and λ_4. Such a non-zero solution always exists (cf. Appendix A). By subtracting the first equation from the third (a step in Gaussian elimination), you get

$$\begin{array}{ccccccccc} \lambda_1 & & & + & \lambda_3 & + & 4\lambda_4 & = & 0 \\ & & \lambda_2 & + & 2\lambda_3 & & & = & 0. \\ & & \lambda_2 & & & - & 3\lambda_4 & = & 0 \end{array}$$

From this system of equations you can spot the non-zero solution

$$\lambda_2 = -6, \quad \lambda_4 = -2, \quad \lambda_3 = 3, \quad \text{and} \quad \lambda_1 = 5,$$

which is the linear dependence in (2.5). In this particular case (2.4) becomes

$$\begin{pmatrix}2\\1\\1\end{pmatrix} = (\tfrac{1}{11} - \theta\,5)\begin{pmatrix}1\\0\\1\end{pmatrix} + (\tfrac{1}{11} + \theta\,6)\begin{pmatrix}0\\1\\1\end{pmatrix} + (\tfrac{5}{11} - \theta\,3)\begin{pmatrix}1\\2\\1\end{pmatrix} + (\tfrac{4}{11} + \theta\,2)\begin{pmatrix}4\\0\\1\end{pmatrix}.$$

How big is θ allowed to be, when

$$\begin{aligned}\tfrac{1}{11} - \theta\, 5 &\geq 0\\ \tfrac{5}{11} - \theta\, 3 &\geq 0\end{aligned}$$

must hold? Solving the inequalities for θ we get $\frac{1}{55} \geq \theta$ and $\frac{5}{33} \geq \theta$. To ensure that both inequalities are satisfied, we must have $\frac{1}{55} \geq \theta$. The magic appears when we pick $\theta = \frac{1}{55}$. Then

$$\begin{aligned}\begin{pmatrix}2\\1\\1\end{pmatrix} &= (\tfrac{1}{11} - \tfrac{1}{55}5)\begin{pmatrix}1\\0\\1\end{pmatrix} + (\tfrac{1}{11} + \tfrac{1}{55}6)\begin{pmatrix}0\\1\\1\end{pmatrix} + (\tfrac{5}{11} - \tfrac{1}{55}3)\begin{pmatrix}1\\2\\1\end{pmatrix} + (\tfrac{4}{11} + \tfrac{1}{55}2)\begin{pmatrix}4\\0\\1\end{pmatrix}\\ &= \tfrac{11}{55}\begin{pmatrix}0\\1\\1\end{pmatrix} + \tfrac{22}{55}\begin{pmatrix}1\\2\\1\end{pmatrix} + \tfrac{22}{55}\begin{pmatrix}4\\0\\1\end{pmatrix}.\end{aligned}$$

This implies

$$\begin{pmatrix}2\\1\end{pmatrix} = \tfrac{11}{55}\begin{pmatrix}0\\1\end{pmatrix} + \tfrac{22}{55}\begin{pmatrix}1\\2\end{pmatrix} + \tfrac{22}{55}\begin{pmatrix}4\\0\end{pmatrix}.$$

Therefore

$$\begin{pmatrix}2\\1\end{pmatrix} \in \operatorname{conv}\left(\begin{pmatrix}0\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}, \begin{pmatrix}4\\0\end{pmatrix}\right).$$

2.8 The polar cone

A *hyperplane* in $\mathbb{R}^n$ is given by

$$H = \{x \in \mathbb{R}^n \mid \alpha^t x = 0\}$$

for $\alpha \in \mathbb{R}^n \setminus \{0\}$. Such a hyperplane divides $\mathbb{R}^n$ into the two *half spaces*

$$\begin{aligned}&\{x \in \mathbb{R}^n \mid \alpha^t x \leq 0\}\\ &\{x \in \mathbb{R}^n \mid \alpha^t x \geq 0\}.\end{aligned}$$

DEFINITION 2.8.1
If $C \subseteq \mathbb{R}^n$ is a convex cone, we call

$$C^* = \{\alpha \in \mathbb{R}^n \mid \alpha^t x \leq 0, \textit{for every } x \in C\}$$

the polar cone of C.

The subset $C^* \subseteq \mathbb{R}^n$ is clearly a convex cone. One of the main results of these notes is that C^* is finitely generated if C is finitely generated. If C is finitely generated, then

$$C = \operatorname{cone}(v_1, \dots, v_m)$$

for suitable $v_1, \dots, v_m \in \mathbb{R}^n$. Therefore

$$C^* = \{\alpha \in \mathbb{R}^n \mid \alpha^t v_1 \leq 0, \dots, \alpha^t v_m \leq 0\}. \tag{2.7}$$

The notation in (2.7) seems to hide the basic nature of the polar cone. Let us unravel it. Suppose that

$$v_1 = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix}, \quad \dots \quad , v_m = \begin{pmatrix} a_{1m} \\ \vdots \\ a_{nm} \end{pmatrix}$$

and

$$\alpha = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Then (2.7) merely says that C^* is the set of solutions to the inequalities

$$\begin{aligned} a_{11}x_1 + \cdots + a_{n1}x_n &\leq 0 \\ &\vdots \\ a_{1m}x_1 + \cdots + a_{nm}x_n &\leq 0. \end{aligned} \tag{2.8}$$

The main result on finitely generated convex cones says that there always exists finitely many solutions $u_1, \dots, u_N$ from which any other solution to (4.9) can be constructed as

$$\lambda_1 u_1 + \cdots + \lambda_N u_N,$$

where $\lambda_1, \dots, \lambda_N \geq 0$. This is the statement that C^* is finitely generated in down to earth terms. Looking at it this way, I am sure you see that this is a non-trivial result. If not, try to prove it from scratch!

In the Figure 2.3 we have sketched a finitely generated cone C along with its polar cone C^*. If you look closer at the drawing, you will see that

$$C = \operatorname{cone}\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right) \quad \text{and} \quad C^* = \operatorname{cone}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right).$$

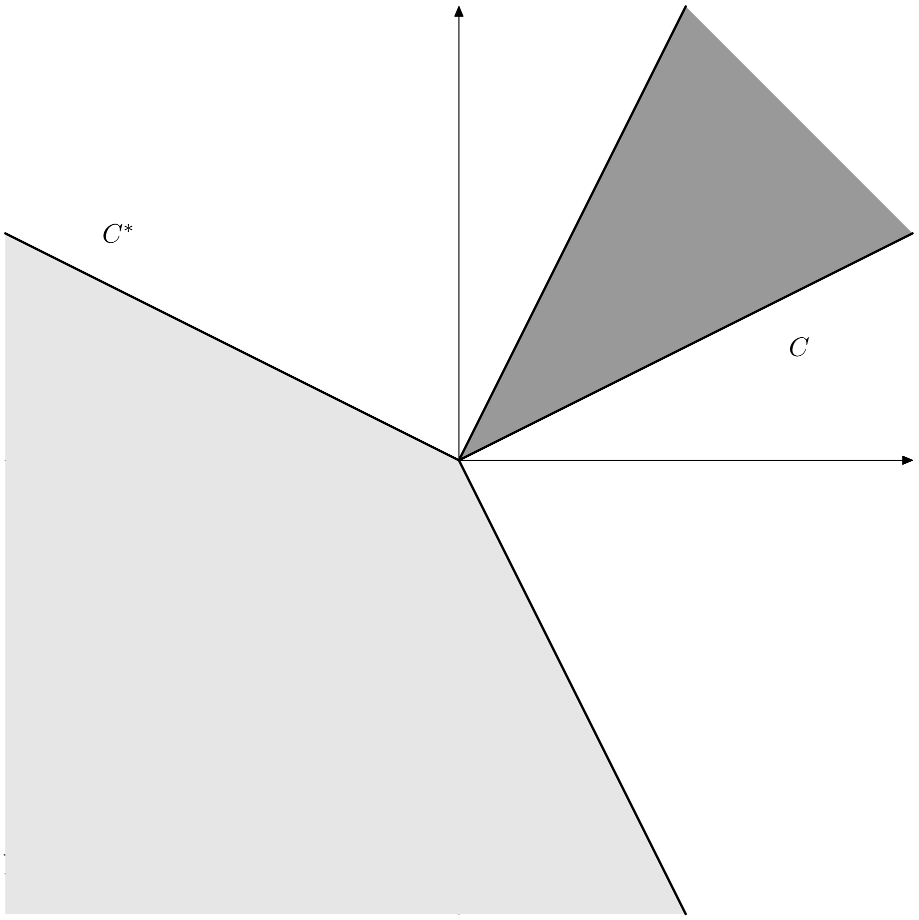


Figure 2.3: A finitely generated cone C and its polar cone C^*.

Notice also that C^* encodes the fact that C is the intersection of the two affine half planes

$$\left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \middle| \begin{pmatrix} 1 \\ -2 \end{pmatrix}^t \begin{pmatrix} x \\ y \end{pmatrix} \leq 0 \right\} \quad \text{and} \quad \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \middle| \begin{pmatrix} -2 \\ 1 \end{pmatrix}^t \begin{pmatrix} x \\ y \end{pmatrix} \leq 0 \right\}.$$

2.9 Exercises

(1) Prove that $A + z = \{x + z \mid x \in A\} \subseteq \mathbb{R}^n$ is a convex subset if $A \subseteq \mathbb{R}^n$ is a convex subset.

(2) Prove that $A \cap B \subseteq \mathbb{R}^n$ is a convex subset if $A, B \subseteq \mathbb{R}^n$ are convex subsets.

(3) Prove that

$$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

is a convex subset of $\mathbb{R}^2$.

(4) Draw the *half plane* $H = \{(x, y)^t \in \mathbb{R}^2 \mid a\,x + b\,y \leq c\} \subseteq \mathbb{R}^2$ for $a = b = c = 1$. Show, without drawing, that H is convex for every a, b, c. Prove in general that the *half space*

$$H = \{(x_1, \dots, x_n)^t \in \mathbb{R}^n \mid a_1 x_1 + \cdots + a_n x_n \leq c\} \subseteq \mathbb{R}^n$$

is convex, where $a_1, \dots, a_n, c \in \mathbb{R}$.

(5) Let $v_1, v_2, v_3 \in \mathbb{R}^n$. Show that

$$\begin{aligned}&\{(1 - \lambda)v_3 + \lambda((1 - \mu)v_1 + \mu v_2) \mid \lambda \in [0, 1], \mu \in [0, 1]\} = \\ &\{\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 \mid \lambda_1, \lambda_2, \lambda_3 \geq 0,\ \lambda_1 + \lambda_2 + \lambda_3 = 1\}.\end{aligned}$$

(6) Prove that $\operatorname{conv}(v_1, \dots, v_m)$ is a closed subset of $\mathbb{R}^n$.

(7) Draw the convex hull of

$$S = \{(0, 0), (1, 0), (1, 1)\} \subseteq \mathbb{R}^2.$$

Write $\operatorname{conv}(S)$ as the intersection of 3 half planes.

(8) Let $u_1, u_2, v_1, v_2 \in \mathbb{R}^n$. Show that

$$\operatorname{conv}(u_1, u_2) + \operatorname{conv}(v_1, v_2) = \operatorname{conv}(u_1 + v_1, u_1 + v_2, u_2 + v_1, u_2 + v_2).$$

Here the sum of two subsets A and B of $\mathbb{R}^n$ is $A + B = \{x + y \mid x \in A, y \in B\}$.

(9) Let $S \subseteq \mathbb{R}^n$ be a convex set and $v \in \mathbb{R}^n$. Show that

$$\operatorname{conv}(S, v) := \{(1 - \lambda)s + \lambda v \mid \lambda \in [0, 1], s \in S\}$$

(*the cone over* S) is a convex set.

(10) Prove Proposition 2.2.7.

(11) Let $X = \operatorname{conv}(v_1, \dots, v_N)$, where $v_1, \dots, v_N \in \mathbb{R}^n$.

  (i) Prove that if $z \in X$ is an extremal point, then $z \in \{v_1, \dots, v_N\}$.

  (ii) Prove that v_1 is not an extremal point in X if and only if

$$v_1 \in \operatorname{conv}(v_2, \dots, v_N).$$

This means that the extremal points in a convex hull like X consists of the "indispensable vectors" in spanning the convex hull.

(12) What are the extremal points of the subset

$$\{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 1\}$$

of $\mathbb{R}^2$? Can you prove it?

(13) What is the characteristic cone of a bounded convex subset?

(14) Can you give an example of an unbounded convex set C with $\operatorname{ccone}(C) = \{0\}$?

(15) Draw a few examples of convex subsets in the plane $\mathbb{R}^2$ along with their characteristic cones, extremal points and extremal directions.

(16) Prove Lemma 2.5.3

(17) Give an example of a cone that is not finitely generated.

(18) Prove that you can have no more than $m + 1$ affinely independent vectors in $\mathbb{R}^m$.

(19) The vector $v = \begin{pmatrix} \frac{7}{4} \\ \frac{19}{8} \end{pmatrix}$ is the convex combination

$$\tfrac{1}{8}\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \tfrac{1}{8}\begin{pmatrix} 1 \\ 2 \end{pmatrix} + \tfrac{1}{4}\begin{pmatrix} 2 \\ 2 \end{pmatrix} + \tfrac{1}{2}\begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

of four vectors in $\mathbb{R}^2$. Use the method outlined in Example 2.7.3 to answer the following questions.

  (i) Is v in the convex hull of three of the four vectors.

  (ii) Is v in the convex hull of two of the four vectors?

(20) Let

$$C = \operatorname{cone}\left(\begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right).$$

(i) Show that

$$C^* = \operatorname{cone}\left(\begin{pmatrix} 1 \\ -2 \end{pmatrix}, \begin{pmatrix} -2 \\ 1 \end{pmatrix} \right).$$

(ii) Suppose that

$$C = \operatorname{cone}\left(\begin{pmatrix} a \\ c \end{pmatrix}, \begin{pmatrix} b \\ d \end{pmatrix} \right),$$

where

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is an invertible matrix. How do you compute C^*?

Chapter 3

Separation

A (non-trivial) linear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by $f(v) = \alpha^t v$ for $\alpha \in \mathbb{R}^n \setminus \{0\}$. Given $\beta \in \mathbb{R}$ such a linear function defines an affine hyperplane by

$$H = \{v \in \mathbb{R}^n \mid f(v) = \beta\}.$$

We will usually omit "affine" in front of hyperplane. A hyperplane divides $\mathbb{R}^n$ into two (affine) half spaces given by

$$\begin{aligned} H_{\leq} &= \{v \in \mathbb{R}^n \mid f(v) \leq \beta\} \\ H_{\geq} &= \{v \in \mathbb{R}^n \mid f(v) \geq \beta\}. \end{aligned}$$

Two given subsets S_1 and S_2 of $\mathbb{R}^n$ are *separated* by H if $S_1 \subseteq H_{\leq}$ and $S_2 \subseteq H_{\geq}$. A separation of S_1 and S_2 given by a hyperplane H is called *proper* if $S_1 \nsubseteq H$ or $S_2 \nsubseteq H$ (the separation is not too interesting if $S_1 \cup S_2 \subseteq H$).

A separation of S_1 and S_2 given by a hyperplane H is called *strict* if $S_1 \subseteq H_{<}$ and $S_2 \subseteq H_{>}$, where

$$\begin{aligned} H_{<} &= \{v \in \mathbb{R}^n \mid f(v) < \beta\} \\ H_{>} &= \{v \in \mathbb{R}^n \mid f(v) > \beta\}. \end{aligned}$$

Finally a separation of S_1 and S_2 is called *strong* if there exists $\epsilon > 0$, such that the two convex subsets $S_1 + \epsilon B$ and $S_2 + \epsilon B$ are strictly separated. Here B denotes the unit ball $\{x \in \mathbb{R}^n \mid |x| \leq 1\}$ in $\mathbb{R}^n$.

Separation by half spaces shows the important result that closed[1] convex sets are solutions to systems of (perhaps infinitely many) linear inequalities. Geometrically this means that a closed convex subset is the intersection of the closed half spaces that contain it.

---

[1] At this point let me refer you to Appendix B for refreshing the necessary concepts from analysis like infimum, supremum, convergent sequences, closed subsets etc.

The most basic and probably most important separation result is strict separation of a closed convex set C from a point $x \notin C$. We need a small preliminary result about closed (and convex) sets.

LEMMA 3.0.1
Let $F \subseteq \mathbb{R}^n$ be a closed subset. Then there exists $x_0 \in F$ such that

$$|x_0| = \inf \left\{ |x| \,\middle|\, x \in F \right\}.$$

If F in addition is convex, then x_0 is unique.

Proof. Let

$$\beta = \inf \left\{ |x| \,\middle|\, x \in F \right\}.$$

We may assume from the beginning that F is bounded. Now construct a sequence (x_n) of points in F with the property that $|x_n| - \beta < 1/n$. Such a sequence exists by the definition of infimum. Since F is bounded, (x_n) has a convergent subsequence (x_{n_i}). Let x_0 be the limit of this convergent subsequence. Then $x_0 \in F$, since F is closed. Also, as $x \mapsto |x|$ is a continuous function from F to $\mathbb{R}$ we must have $|x_0| = \beta$. This proves the existence of x_0.

If F in addition is convex, then x_0 is unique: suppose that $y_0 \in F$ is another point with $|y_0| = |x_0|$. Consider

$$z = \frac{1}{2}(x_0 + y_0) = \frac{1}{2}x_0 + \frac{1}{2}y_0 \in F.$$

But in this case, $|z| = \frac{1}{2}|x_0 + y_0| \leq \frac{1}{2}|x_0| + \frac{1}{2}|y_0| = |x_0|$. Therefore $|x_0 + y_0| = |x_0| + |y_0|$. From the triangle inequality it follows that x_0 are y_0 collinear i.e. there exists $\lambda \in \mathbb{R}$ such that $x_0 = \lambda y_0$. Then $\lambda = \pm 1$. In both cases we have $x_0 = y_0$. $\square$

COROLLARY 3.0.2
Let $F \subseteq \mathbb{R}^n$ be a closed subset and $z \in \mathbb{R}^n$. Then there exists $x_0 \in F$ such that

$$|x_0 - z| = \inf \left\{ |x - z| \,\middle|\, x \in F \right\}.$$

If F in addition is convex, then x_0 is unique.

Proof. If F is closed (convex) then

$$F - z = \{x - z \mid z \in F\}$$

is also closed (convex). Now the results follow from applying Lemma 3.0.1 to $F - z$. $\square$

If $F \subseteq \mathbb{R}^n$ is a closed subset with the property that to each point of $\mathbb{R}^n$ there is a unique nearest point in F, then one may prove that F is convex! This result is due to Bunt (1934)[2] and Motzkin (1935).

The following is a geometric characterization of the unique vector of smallest length in a closed convex set.

LEMMA 3.0.3

Let $C \subseteq \mathbb{R}^n$ be a closed convex subset. Then $x_0 \in C$ is the unique vector of smallest length in C if and only if

$$(z - x_0)^t x_0 \geq 0 \tag{3.1}$$

for every $z \in C$.

Proof. Suppose that (3.1) holds. Then

$$|z|^2 = |z - x_0 + x_0|^2 = |z - x_0|^2 + |x_0|^2 + 2(z - x_0)^t x_0 \geq 0.$$

This shows that $|z| \geq |x_0|$ for $z \in C$ proving that x_0 has minimal length in C. Now assume that $|x_0| \leq |z|$ for every $z \in C$. Then

$$\begin{aligned} |(1-\lambda)x_0 + \lambda z)|^2 &= |x_0 + \lambda(z - x_0)|^2 \\ &= |x_0|^2 + \lambda^2 |z - x_0|^2 + 2\lambda(z - x_0)^t x_0. \end{aligned} \tag{3.2}$$

This identity forces the condition $(z - x_0)^t x_0 \geq 0$. If there exists $z \in C$ with $(z - x_0)^t x_0 < 0$, then we can use (3.2) to prove the existence of a small $\lambda > 0$, such that

$$|(1-\lambda)x_0 + \lambda z)|^2 - |x_0|^2 < 0$$

violating the minimality of $|x_0|$. ☐

3.1 Separation of a point from a closed convex set

COROLLARY 3.1.1

Let C be a closed convex subset of $\mathbb{R}^n$ and let

$$|x_0| = \inf \left\{ |x| \middle| x \in C \right\},$$

where $x_0 \in C$. If $0 \notin C$, then

$$x_0^t z > |x_0|^2/2$$

for every $z \in C$.

---

[2] L. N. H. Bunt, Dutch mathematician

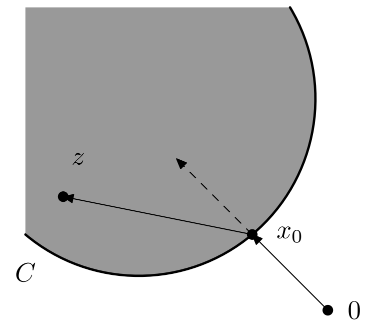


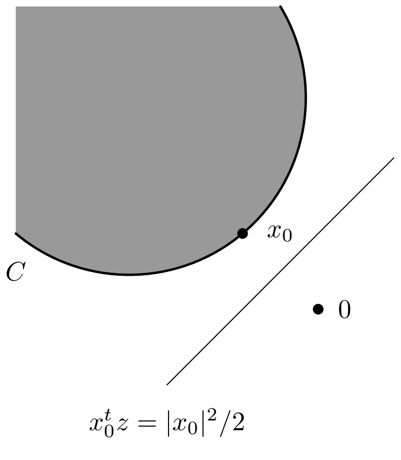
Figure 3.1: The geometric characterization of the vector x_0 of smallest length in a closed convex set C. The angle between x_0 and $z - x_0$ must be acute for every $z \in C$.

Proof. This is an immediate consequence of Lemma 3.0.3. ☐

Corollary 3.1.1 shows that $\{0\}$ is strongly separated from a closed convex set C if $0 \notin C$. The separating hyperplane

$$\{z \in \mathbb{R}^n \mid x_0^t z = \frac{1}{2}|x_0|^2\}.$$

is constructed from the unique vector x_0 of minimal length.



THEOREM 3.1.2

Let C be a closed convex subset of $\mathbb{R}^n$ with $v \notin C$ and let x_0 be the unique point in C closest to v. Then

$$(x_0 - v)^t (z - v) > \frac{1}{2}|x_0 - v|^2$$

for every $z \in C$: the hyperplane $H = \{x \in \mathbb{R}^n \mid \alpha^t x = \beta\}$ with $\alpha = x_0 - v$ and $\beta = (x_0 - v)^t v + |x_0 - v|^2/2$ separates $\{v\}$ strongly from C.

Proof. Let $C' = C - v = \{x - v \mid x \in C\}$. Then C' is closed and convex and $0 \notin C'$. The point closest to 0 in C' is $x_0 - v$. Now the result follows from Corollary 3.1.1 applied to C'. □

With this result we get one of the key properties of closed convex sets.

THEOREM 3.1.3
A closed convex set $C \subseteq \mathbb{R}^n$ is the intersection of the half spaces containing it.

Proof. We let J denote the set of all half spaces

$$H_{\leq} = \{x \in \mathbb{R}^n \mid \alpha^t\, x \leq \beta\}$$

with $C \subseteq H_{\leq}$. One inclusion is easy:

$$C \subseteq \bigcap_{H_{\leq} \in J} H_{\leq}.$$

In the degenerate case, where $C = \mathbb{R}^n$ we have $J = \emptyset$ and the above intersection is $\mathbb{R}^n$. Suppose that there exists

$$x \in \bigcap_{H_{\leq} \in J} H_{\leq} \setminus C. \tag{3.3}$$

Then Theorem 3.1.2 shows the existence of a hyperplane H with $C \subseteq H_{\leq}$ and $x \notin H_{\leq}$. This contradicts (3.3). □

This result tells an important story about closed convex sets. A closed half space $H_{\leq} \subseteq \mathbb{R}^n$ is the set of solutions to a linear inequality

$$a_1 x_1 + \cdots + a_n x_n \leq b.$$

Therefore a closed convex set really is the set of common solutions to a (possibly infinite) set of linear inequalities. If C happens to be a (closed) convex cone, we can say even more.

COROLLARY 3.1.4
Let $C \subseteq \mathbb{R}^n$ be a closed convex cone. Then

$$C = \bigcap_{\alpha \in C^*} \{x \in \mathbb{R}^n \mid \alpha^t\, x \leq 0\}.$$

Proof. If C is contained in a half space $\{x \in \mathbb{R}^n \mid \alpha^t\, x \leq \beta\}$ we must have $\beta \geq 0$, since $0 \in C$. We cannot have $\alpha^t x > 0$ for any $x \in C$, since this would imply that $\alpha^t(\lambda x) = \lambda(\alpha^t x) \to \infty$ for $\lambda \to \infty$. As $\lambda x \in C$ for $\lambda \geq 0$ this violates $\alpha^t x \leq \beta$ for every $x \in C$. Therefore $\alpha \in C^*$ and $\beta = 0$. Now the result follows from Theorem 3.1.3. □

3.2 Supporting hyperplanes

With some more attention to detail we can actually prove that any convex set $C \subseteq \mathbb{R}^n$ (not necessarily closed) is always contained on one side of an affine hyperplane "touching" C at its boundary. Of course here you have to assume that $C \neq \mathbb{R}^n$.

First we need to prove that the closure of a convex subset is also convex.

PROPOSITION 3.2.1
Let $S \subseteq \mathbb{R}^n$ be a convex subset. Then the closure, $\overline{S}$, of S is a convex subset.

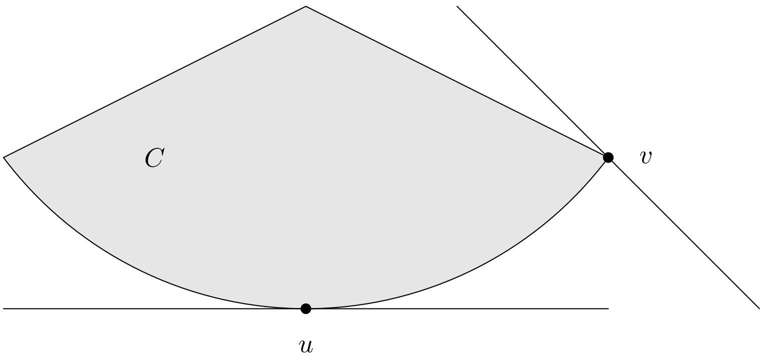
Proof. Consider $x, y \in \overline{S}$. We must prove that $z := (1-\lambda)x + \lambda y \in \overline{S}$ for $0 \leq \lambda \leq 1$. By definition of the closure $\overline{S}$ there exists convergent sequences (x_n) and (y_n) with $x_n, y_n \in S$ such that $x_n \to x$ and $y_n \to y$. Now form the sequence $((1-\lambda)x_n + \lambda y_n)$. Since S is convex this is a sequence of vectors in S. The convergence of (x_n) and (y_n) allows us to conclude that

$$(1-\lambda)x_n + \lambda y_n \to z.$$

Since z is the limit of a convergent sequence with vectors in S, we have shown that $z \in \overline{S}$. $\square$

DEFINITION 3.2.2
A supporting hyperplane for a convex set $C \subseteq \mathbb{R}^n$ at a boundary point $z \in \partial C$ is hyperplane H with $z \in H$ and $C \subseteq H_{\leq}$. Here $H_{\leq}$ is called a supporting half space for C at z.

Below you see an example of a convex subset $C \subseteq \mathbb{R}^2$ along with supporting hyperplanes at two of its boundary points u and v. Can you see any difference between u and v?



THEOREM 3.2.3
Let $C \subseteq \mathbb{R}^n$ be a convex set and $z \in \partial C$. Then there exists a supporting hyperplane for C at z.

Proof. The fact that $z \in \partial C$ means that there exists a sequence of points (z_n) with $z_n \notin \overline{C}$, such that $z_n \to z$ (z can be approximated outside of $\overline{C}$). Proposition 3.2.1 says that $\overline{C}$ is a convex subset. Therefore we may use Lemma 3.0.1 to conclude that $\overline{C}$ contains a unique point closest to any point. For each z_n we let $x_n \in \overline{C}$ denote the point closest to z_n and put

$$u_n = \frac{z_n - x_n}{|z_n - x_n|}.$$

Theorem 3.1.2 then shows that

$$u_n^t \,(x - z_n) > \frac{1}{2}\,|z_n - x_n| \tag{3.4}$$

for every $x \in \overline{C}$. Since (u_n) is a bounded sequence, it has a convergent subsequence. Let u be the limit of this convergent subsequence. Then (3.4) shows that $H = \{x \in \mathbb{R}^n \mid u^t\, x = u^t\, z\}$ is a supporting hyperplane for C at z, since $|z_n - x_n| \leq |z_n - z| \to 0$ as $n \to \infty$. ☐

3.3 Separation of disjoint convex sets

THEOREM 3.3.1
Let $C_1, C_2 \subseteq \mathbb{R}^n$ be disjoint ($C_1 \cap C_2 = \emptyset$) convex subsets. Then there exists a separating hyperplane

$$H = \{u \in \mathbb{R}^n \mid \alpha^t\, u = \beta\}$$

for C_1 and C_2. If C_1 is closed and bounded and C_2 is closed, then C_1 and C_2 can be strongly separated.

Proof. The trick is to observe that $C_1 - C_2 = \{x - y \mid x \in C_1, y \in C_2\}$ is a convex subset and $0 \notin C_1 - C_2$. There exists $\alpha \in \mathbb{R}^n \setminus \{0\}$ with

$$\alpha^t(x - y) \geq 0$$

or $\alpha \cdot x \geq \alpha \cdot y$ for every $x \in C_1, y \in C_2$. Here is why. If $0 \notin \overline{C_1 - C_2}$ this is a consequence of Corollary 3.1.1. If $0 \in \overline{C_1 - C_2}$ we get it from Theorem 3.2.3 with $z = 0$.

Therefore $\beta_1 \geq \beta_2$, where

$$\beta_1 = \inf\{\alpha^t\, x \mid x \in C_1\}$$
$$\beta_2 = \sup\{\alpha^t\, y \mid y \in C_2\}$$

and

$$H = \{u \in \mathbb{R}^n \mid \alpha^t u = \beta_1\}$$

is the desired hyperplane.

Assume now that C_1 is closed and bounded and C_2 is closed (still with empty intersection). Then we claim that 0 cannot be in the closure of $C_1 - C_2$. Suppose this was the case. Then there exists a sequence $(x_n - y_n)$ converging to 0 with $x_n \in C_1$ and $y_n \in C_2$. Since C_1 is bounded, (x_n) has a convergent subsequence (x_{n_j}) convering to say $x \in C_1$ (here we use that C_1 is closed). This implies that the sequence (y_{n_j}) is convergent with limit point say $y \in C_2$ (here we use the C_2 is closed). But then, $x - y = 0$ contradicting that $C_1 \cap C_2 = \emptyset$.

Now Corollary 3.1.1 gives the existence of a non-zero $\alpha \in \mathbb{R}^n$ and $\epsilon > 0$, such that

$$\alpha^t(x - y) > \epsilon$$

for every $x \in C_1, y \in C_2$ proving strong separation between C_1 and C_2, since $\beta_1 > \beta_2$ above. $\square$

The separation in the theorem does not have to be proper (example?). However, if $C_1^\circ \neq \emptyset$ or $C_2^\circ \neq \emptyset$ then the separation is proper (why?).

3.4 An application

We give the following application, which is a classical result [4] due to Gordan[3] dating back to 1873.

THEOREM 3.4.1

Let A be an $m \times n$ matrix. Then precisely one of the following two conditions holds.

(i) There exists $x \in \mathbb{R}^n$, such that

$$Ax < 0.$$

(ii) There exists a non-zero $y \geq 0$ in $\mathbb{R}^m$ such that

$$y^t A = 0$$

Proof. If $Ax < 0$, then we cannot have a non-zero $y \geq 0$ with $y^t A = 0$, since this would imply that

$$0 = (y^t A)x = y^t(Ax) < 0.$$

---

[3] Paul Gordan (1837–1912), German mathematician.

Now define the convex subsets

$$\begin{aligned} C_1 &= \{Ax \mid x \in \mathbb{R}^n\} \quad \text{and} \\ C_2 &= \{y \in \mathbb{R}^m \mid y < 0\} \end{aligned}$$

of $\mathbb{R}^m$. If $Ax < 0$ is unsolvable, then $C_1 \cap C_2 = \emptyset$ and Theorem 3.3.1 implies the existence of a separating hyperplane

$$L = \{x \in \mathbb{R}^n \mid \alpha^t x = \beta\}$$

such that

$$C_1 \subseteq \{x \in \mathbb{R}^n \mid \alpha^t x \geq \beta\} \tag{3.5}$$

$$C_2 \subseteq \{x \in \mathbb{R}^n \mid \alpha^t x \leq \beta\}. \tag{3.6}$$

These containments force strong restrictions on α and β: $\beta \leq 0$ by (3.5), since $0 \in C_1$ and $\beta \geq 0$ by (3.6) as $\alpha^t z \to 0$ for $z \to 0$ and $z \in C_2$. Therefore $\beta = 0$. Also from (3.6) we must have $\alpha \geq 0$ to ensure that $\alpha^t z \leq \beta$ holds for every z in the unbounded set C_2. We claim that the result follows by putting $y = \alpha$. Assume on the contrary that $\alpha^t A \neq 0$. Then the containment

$$C_1 = \{Ax \mid x \in \mathbb{R}^n\} \subseteq \{x \in \mathbb{R}^n \mid \alpha^t x \geq 0\}$$

breaks down, as we can find $x \in \mathbb{R}^n$ with $(\alpha^t A)x < 0$. ☐

3.5 Farkas' lemma

The lemma of Farkas[4] is an extremely important result in the theory of convex sets. Farkas published his result in 1901 (see [1]). The lemma itself may be viewed as the separation of a finitely generated cone

$$C = \operatorname{cone}(v_1, \dots, v_r) \tag{3.7}$$

from a point $v \notin C$. In the classical formulation this is phrased as solving linear equations with non-negative solutions. There is no need to use our powerful separation results in proving Farkas' lemma. It follows quite easily from Fourier-Motzkin elimination. Using separation in this case, only complicates matters and hides the "polyhedral" nature of the convex subset in (3.7).

---

[4] Gyula Farkas (1847–1930), Hungarian mathematician.

In teaching convex sets some years ago, I tried to convince the authors of a certain engineering textbook, that they really had to prove, that a finitely generated cone (like the one in (3.7)) is a closed subset of $\mathbb{R}^n$. After three or four email notes with murky responses, I gave up.

The key insight is the following little result, which is the finite analogue of Corollary 3.1.4.

LEMMA 3.5.1
A finitely generated cone

$$C = \operatorname{cone}(v_1, \dots, v_r) \subseteq \mathbb{R}^n$$

is a finite interesection of half spaces i.e. there exists an $m \times n$ *matrix* A*, such that*

$$C = \{v \in \mathbb{R}^n \mid Av \leq 0\}.$$

Proof. Let B denote the $n \times r$-matrix with $v_1, \dots, v_r$ as columns. Consider the polyhedron

$$\begin{aligned} P &= \{(x, y) \in \mathbb{R}^{r+n} \mid y = Bx,\ x \geq 0\} \\ &= \left\{ (x, y) \in \mathbb{R}^{r+n} \,\middle|\, \begin{array}{rl} y - Bx & \leq 0 \\ Bx - y & \leq 0 \\ -x & \leq 0 \end{array} \right\} \end{aligned}$$

and let $\pi : \mathbb{R}^{r+n} \to \mathbb{R}^n$ be the projection defined by $\pi(x, y) = y$. Notice that P is deliberately constructed so that $\pi(P) = C$.

Theorem 1.2.2 now implies that $C = \pi(P) = \{y \in \mathbb{R}^n \mid Ay \leq b\}$ for $b \in \mathbb{R}^m$ and A an $m \times n$ matrix. Since $0 \in C$ we must have $b \geq 0$. In fact $b = 0$ has to hold: a $y \in C$ with $Ay \not\leq 0$ means that a coordinate, $z = (Ay)_j > 0$. Since $\lambda x \in C$ for every $\lambda \geq 0$ this would imply $\lambda z = (A(\lambda x))_j$ is not bounded for $\lambda \to \infty$ and we could not have $Ax \leq b$ for every $x \in C$ (see also Exercise 9). ☐

A cone of the form $\{v \in \mathbb{R}^n \mid Av \leq 0\}$ is called *polyhedral* (because it is a polyhedron in the sense of Definition 1.2.1). Lemma 3.5.1 shows that a finitely generated cone is polyhedral. A polyhedral cone is also finitely generated. We shall have a lot more to say about this in the next chapter.

The following result represents the classical Farkas lemma in the language of matrices.

LEMMA 3.5.2 (Farkas)
Let A *be an* $m \times n$ *matrix and* $b \in \mathbb{R}^m$*. Then precisely one of the following two conditions is satisfied.*

(i) *The system*

$$Ax = b$$

of linear equations is solvable with $x \in \mathbb{R}^n$ *with* non-negative *entries.*

(ii) *There exists* $y \in \mathbb{R}^m$ *such that*

$$y^t A \geq 0 \quad \textit{and} \quad y^t b < 0.$$

Proof. The properties (i) and (ii) cannot be true at the same time. Suppose that they both hold. Then we get that

$$y^t b = y^t (Ax) = (y^t A)x \geq 0$$

since $y^t A \geq 0$ and $x \geq 0$. This contradicts $y^t b < 0$. The real surprise is the existence of y if $Ax = b$ cannot be solved with $x \geq 0$. Let $v_1, \dots, v_m$ denote the m columns in A. Then the key observation is that

$$x_1 v_1 + \cdots + x_m v_m = Ax = b$$

is solvable with $x = (x_1, \dots, x_m)^t \geq 0$ if and only if

$$b \in C = \text{cone}(v_1, \dots, v_m).$$

So if $Ax = b, x \geq 0$ is non-solvable we must have $b \notin C$. Lemma 3.5.1 shows that $b \notin C$ implies you can find $y \in \mathbb{R}^n$ with $y^t b < 0$ and $y^t A \geq 0$ simply by using the description of C as a finite intersection of half planes. ☐

3.6 Exercises

(1) Give an example of a non-proper separation of convex subsets.

(2) Let $F_1, F_2 \subseteq \mathbb{R}^2$ be closed subsets. Is

$$F_1 - F_2 = \{x - y \mid x \in F_1, y \in F_2\}$$

a closed subset of $\mathbb{R}^2$?

(3) Give an example of two disjoint closed convex subsets of $\mathbb{R}^2$ which are strictly but not strongly separated.

(4) Suppose that $C_1, C_2 \subseteq \mathbb{R}^n$ are closed convex subsets with $C_1 \cap C_2 = \emptyset$. Are C_1 and C_2 strictly separated?

(5) Let

$$B_1 = \{(x,y) \mid x^2 + y^2 \leq 1\}$$
$$B_2 = \{(x,y) \mid (x-2)^2 + y^2 \leq 1\}$$

(a) Show that B_1 and B_2 are closed convex subsets of $\mathbb{R}^2$.

(b) Find a hyperplane properly separating B_1 and B_2.

(c) Can you separate B_1 and B_2 strictly?

(d) Put $B_1' = B_1 \setminus \{(1,0)\}$ and $B_2' = B_2 \setminus \{(1,0)\}$. Show that B_1' and B_2' are convex subsets. Can you separate B_1' from B_2 strictly? What about B_1' and B_2'?

(6) Let $C = \{(x,y) \in \mathbb{R}^2 \mid (x-1)^2 + y^2 \leq 1\}$ and $v = (0,2)$. What is the point in C closest to v? Find the equation of a hyperplane separating $\{v\}$ from C strictly.

(7) Let S be the square with vertices $(0,0), (1,0), (0,1)$ and $(1,1)$ and $P = (2,0)$.

(i) Find the set of hyperplanes through $(1, \frac{1}{2})$, which separate S from P.

(ii) Find the set of hyperplanes through $(1,0)$, which separate S from P.

(iii) Find the set of hyperplanes through $(\frac{3}{2}, 1)$, which separate S from P.

(8) Let $C_1, C_2 \subseteq \mathbb{R}^n$ be convex subsets. Prove that

$$C_1 - C_2 := \{x - y \mid x \in C_1, y \in C_2\}$$

is a convex subset.

(9) Take another look at the proof of Theorem 1.2.2. Show that

$$\pi(P) = \{x \in \mathbb{R}^{n-1} \mid A'x \leq 0\}$$

if $P = \{x \in \mathbb{R}^n \mid Ax \leq 0\}$, where A and A' are matrices with n and $n-1$ columns respectively.

(10) Show using the lemma of Farkas that

(i)

$$\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 5 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}$$

is unsolvable with $x \geq 0, y \geq 0$ and $z \geq 0$.

(ii)

$$\begin{pmatrix} 1 & 2 & 3 & 1 \\ 3 & 1 & 5 & 1 \\ 1 & 2 & 1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ w \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$$

is unsolvable with $x \geq 0, y \geq 0, z \geq 0$ and $w \geq 0$.

(11) Recall Theorem 3.3.1, where we separated two disjoint convex subsets C_1 and C_2 of $\mathbb{R}^n$. Can this result be strengthened to the existence of $\alpha \in \mathbb{R}^n$ and $\beta \in \mathbb{R}$ such that

$$\begin{aligned} C_1 &\subseteq \{x \in \mathbb{R}^n \mid \alpha^t x < \beta\} \quad \text{and} \\ C_2 &\subseteq \{x \in \mathbb{R}^n \mid \alpha^t x \geq \beta\}? \end{aligned}$$

If not, give a counterexample (perhaps in $\mathbb{R}^2$).

(12) Find all the supporting hyperplanes of the triangle with vertices $(0,0), (0,2)$ and $(1,0)$.

Chapter 4

Polyhedra

You already know from linear algebra that the set of solutions L to a system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{n1}x_n &= 0 \\ &\vdots \\ a_{1m}x_1 + \cdots + a_{nm}x_n &= 0 \end{aligned} \tag{4.1}$$

of (homogeneous) linear equations can be generated from a set of basic solutions $v_1, \dots, v_r \in \mathbb{R}^n$, where $r \leq n$. This simply means that the set of solutions to (4.1) is

$$\{\lambda_1 v_1 + \cdots + \lambda_r v_r \mid \lambda_i \in \mathbb{R}\} = \text{cone}(\pm v_1, \dots, \pm v_r).$$

The basic properties are here that $u + v \in L$ and $\lambda u \in L$ if $u, v \in L$ and λ is any real number.

Things change dramatically when we replace $=$ with $\leq$ and look at the set U of solutions to a set of linear inequalities

$$\begin{aligned} a_{11}x_1 + \cdots + a_{n1}x_n &\leq 0 \\ &\vdots \\ a_{1m}x_1 + \cdots + a_{nm}x_n &\leq 0. \end{aligned} \tag{4.2}$$

Here we only have $u + v \in U$ and $\lambda u \in U$ for $u, v \in U$ and $\lambda \geq 0$ i.e. U is a convex cone. The condition $\lambda \geq 0$ makes all the difference. In a sense we are working with a generalization of linear algebra.

The main result in this chapter is that we still have a set of basic solutions $v_1, \dots, v_r$. However here the set of solutions to (4.2) is

$$\{\lambda_1 v_1 + \cdots + \lambda_r v_r \mid \lambda_i \in \mathbb{R} \text{ and } \lambda_i \geq 0\} = \text{cone}(v_1, \dots, v_r) \tag{4.3}$$

and r can be very big compared to n. In addition we have to change our notion of a basic solution (to an extremal generator).

Geometrically we are saying that an intersection of half spaces like (4.2) is generated by finitely many rays as in (4.3). This is a very intuitive and very powerful mathematical result. In some cases it is easy to see as in

$$\begin{array}{rrrcl} -x & & & \leq & 0 \\ & -y & & \leq & 0 \\ & & -z & \leq & 0 \end{array} \tag{4.4}$$

Here the set of solutions is

$$\operatorname{cone}\left(\begin{pmatrix}1\\0\\0\end{pmatrix}, \begin{pmatrix}0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\end{pmatrix} \right) \tag{4.5}$$

What if we add the inequality $x - y + z \leq 0$ to (4.4)? How does the set of solutions change in (4.5)? With this extra inequality

$$\begin{pmatrix}1\\0\\0\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}0\\0\\1\end{pmatrix}$$

are no longer (basic) solutions. The essence of our next result is to describe this change.

4.1 The double description method

The double description method is a very clever algorithm for solving (homogeneous) linear inequalities. It first appeared in [7] with later refinements in [3]. It gives a nice inductive proof of the classical theorem of Minkowski[1] and Weyl[2] on the structure of polyhedra (Theorem 4.5.1).

The first step of the algorithm is computing the solution set to one inequality in n variables like

$$3x + 4y + 5z \leq 0 \tag{4.6}$$

in the three variables x, y and z. Here the solution set is the set of vectors with $3x + 4y + 5z = 0$ along with the non-negative multiples of just one vector (x_0, y_0, z_0) with

$$3x_0 + 4y_0 + 5z_0 < 0.$$

---

[1] Hermann Minkowski (1864–1909), German mathematician.

[2] Hermann Weyl (1885–1955), German mathematician.

Concretely the solution set to (4.6) is

$$\operatorname{cone}\left(\begin{pmatrix}-4\\3\\0\end{pmatrix}, \begin{pmatrix}4\\-3\\0\end{pmatrix}, \begin{pmatrix}0\\5\\-4\end{pmatrix}, \begin{pmatrix}0\\-5\\4\end{pmatrix}, \begin{pmatrix}-1\\-1\\-1\end{pmatrix}\right). \tag{4.7}$$

The double description method is a systematic way of updating the solution set when we add further inequalities.

4.1.1 The key lemma

The lemma below is a bit technical, but its main idea and motivation are very simple: We have a description of the solutions to a system of m linear inequalities in basis form. How does this solution change if we add a linear inequality? If you get stuck in the proof, then take a look at the examples that follow. Things are really quite simple (but nevertheless clever).

LEMMA 4.1.1
Consider

$$C = \left\{ x \in \mathbb{R}^n \middle| \begin{array}{rcl} a_1^t x & \leq & 0 \\ & \vdots & \\ a_m^t x & \leq & 0 \end{array} \right\}$$

for $a_1, \dots, a_m \in \mathbb{R}^n$. *Suppose that*

$$C = \operatorname{cone}(v \mid v \in V)$$

for $V = \{v_1, \dots, v_N\}$. *For* $a \in \mathbb{R}^n$ *divide* V *into the disjoint subsets*

$$\begin{aligned} V^- &= \{v \in V \mid a^t v < 0\} \\ V^0 &= \{v \in V \mid a^t v = 0\} \\ V^+ &= \{v \in V \mid a^t v > 0\}. \end{aligned}$$

Then

$$C \cap \{x \in \mathbb{R}^n \mid a^t x \leq 0\} = \operatorname{cone}(w \mid w \in W),$$

where

$$W = V^- \cup V^0 \ \cup \ \{(a^t u)v - (a^t v)u \mid u \in V^+ \quad \text{and} \quad v \in V^-\}.$$

Proof. Let $C' = C \cap \{x \in \mathbb{R}^n \mid a^t x \leq 0\}$ and $C'' = \operatorname{cone}(w | w \in W)$. Then $C'' \subseteq C'$ as the generators $(a^t u)v - (a^t v)u$ are designed so that they belong to C' (check this!). We will prove that $C' \subseteq C''$. Suppose that $z \in C'$. Since $C' \subseteq C$ we have

$$z = \lambda_1 v_1 + \cdots + \lambda_N v_N, \tag{4.8}$$

where $\lambda_i \geq 0$. Let

$$J_z^- = \{v_i \mid \lambda_i > 0,\, v_i \in V^-\} \quad \text{and} \quad J_z^+ = \{v_i \mid \lambda_i > 0,\, v_i \in V^+\}.$$

We will prove that $z \in C''$ by induction on the number of elements in $J_z^- \cup J_z^+$. If $J_z^+ = \emptyset$, then z is in the cone generated by V^- and V^0 i.e. $z \in C''$. Suppose therefore that J_z^+ is non-empty and let $u \in J_z^+$. Since $a^t z \leq 0$, we must have $J_z^- \neq \emptyset$. Let $v \in J_z^-$ and consider

$$z' = z - \mu((a^t u)v - (a^t v)u)$$

for $\mu > 0$. By varying $\mu > 0$ suitably you can hit the sweet spot where $|J_{z'}^+ \cup J_{z'}^-| < |J_z^+ \cup J_z^-|$ by cancelling the coefficient in front of either u or v in (4.8). The induction assumption now gives that $z' \in C''$. Therefore $z \in C''$. $\square$

EXAMPLE 4.1.2

Let us apply Lemma 4.1.1 to the system

$$\begin{array}{rcrcrcc} -x & + & y & + & z & \leq & 0 \\ x & - & y & + & z & \leq & 0 \\ x & + & y & - & z & \leq & 0 \end{array} \tag{4.9}$$

of inequalities. The solution set of the first inequality is the cone $C = \operatorname{cone}(v \mid v \in V)$, where

$$V = \{(1,0,0), (0,1,-1), (0,-1,1), (1,1,0), (-1,-1,0)\}.$$

The second inequality amounts to applying the lemma with $a = (1,-1,1)$. Here

$$\begin{aligned} V^- &= \{(0,1,-1)\} \\ V^0 &= \{(1,1,0), (-1,-1,0)\} \\ V^+ &= \{(1,0,0), (0,-1,1)\} \end{aligned}$$

and in the notation of the lemma we get

$$\begin{aligned} W &= V^- \cup V^0 \cup \{(0,1,-1) + 2(1,0,0), 2(0,1,-1) + 2(0,-1,1)\} \\ &= V^- \cup V^0 \cup \{(2,1,-1)\} = \{(0,1,-1), (1,1,0), (-1,-1,0), (2,1,-1)\}. \end{aligned}$$

Finally putting $V = W$ and adding the third inequality amounts to applying the lemma with $a = (1,1,-1)$. Here

$$\begin{aligned} V^- &= \{(-1,-1,0))\} \\ V^0 &= \emptyset \\ V^+ &= \{(0,1,-1), (1,1,0), (2,1,-1)\} \end{aligned}$$

and

$$\begin{aligned} W = & V^- \cup V^0 \cup \\ & \{2(-1,-1,0)+2(0,1,-1), 2(-1,-1,0)+2(1,1,0), 4(-1,-1,0)+2(2,1,-1)\} \\ & = \{(-1,-1,0),(-2,0,-2),(0,-2,-2)\} \end{aligned}$$

Therefore the complete solution set to the inequalities in (4.9) is

$$\operatorname{cone}\Big((-1,-1,0),(-1,0,-1),(0,-1,-1)\Big).$$

Later you will see that this set of generators is minimal in the sense that the generators are socalled extremal rays.

EXAMPLE 4.1.3

Let us see what happens to the solutions to (4.4) when we add the extra inequality $x - y + z \le 0$. Here

$$a = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix} \quad \text{and} \quad V = \left\{ \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

Therefore

$$W = \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

EXAMPLE 4.1.4

Suppose on the other hand that the inequality $x - y - z \le 0$ was added to (4.4). Then

$$W = \left\{ \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} \right\}.$$

All of these generators are "basic" solutions. You cannot leave out a single of them. Let us add the inequality $x - 2y + z \le 0$, so that we wish to solve

$$\begin{array}{rrrcl} -x & & & \le & 0 \\ & -y & & \le & 0 \\ & & -z & \le & 0 \\ x & -y & -z & \le & 0 \\ x & -2y & +z & \le & 0 \end{array} \tag{4.10}$$

Here we apply Lemma 4.1.1 with $a = (1, -2, 1)^t$ and $V = W$ above and the new generators are (in transposed form)

$$\begin{aligned}
&= (0, 1, 0) \\
&= (1, 1, 0) \\
(0, 1, 0) + 2(0, 0, 1) &= (0, 1, 2) \\
2(0, 1, 0) + 2(1, 0, 1) &= (2, 2, 2) \\
(1, 1, 0) + (0, 0, 1) &= (1, 1, 1) \\
2(1, 1, 0) + (1, 0, 1) &= (3, 2, 1)
\end{aligned}$$

You can see that the generators $(1, 1, 1)$ and $(2, 2, 2)$ are superfluous, since

$$(1, 1, 1) = \frac{1}{3}(0, 1, 2) + \frac{1}{3}(3, 2, 1).$$

We would like to have a way of generating only the necessary "basic" or extremal solutions when adding a new inequality. This is the essence of the following section.

4.2 Extremal and adjacent rays

Recall that an extremal ray in a convex cone C is an element $v \in C$, such that $v = u_1 + u_2$ with $u_1, u_2 \in C$ implies $v = \lambda u_1$ or $v = \lambda u_2$ for some $\lambda \geq 0$. So the extremal rays are the ones necessary for generating the cone. You cannot leave any of them out. We need some general notation. Suppose that A is an $m \times d$ with the m rows $a_1, \dots, a_m \in \mathbb{R}^d$. For a given $x \in \mathbb{R}^d$ we let

$$I(x) = \{i \mid a_i^t\, x = 0\} \subseteq \{1, \dots, m\}.$$

For a given subset $J \subseteq \{1, \dots, m\}$ we let A_J denote the matrix with rows $(a_j \mid j \in J)$ and $Ax \leq 0$ denotes the collection $a_1^t\, x \leq 0, \dots, a_m^t\, x \leq 0$ of linear inequalities.

PROPOSITION 4.2.1
Let

$$C = \{x \in \mathbb{R}^d \mid Ax \leq 0\} \tag{4.11}$$

where A is an $m \times d$ matrix of full rank d. Then $v \in C$ is an extremal ray if and only if the rank of A_I is $d - 1$, where $I = I(v)$.

Proof. Suppose that the rank of A_I is $< d - 1$, where v is an extremal ray with $I = I(v)$. Then we may find a non-zero $x \in \mathbb{R}^d$ with $A_I x = 0$ and $v^t x = 0$. Consider

$$v = \frac{1}{2}(v - \epsilon x) + \frac{1}{2}(v + \epsilon x). \tag{4.12}$$

For small $\epsilon > 0$ you can check the inequalities in (4.11) and show that $v \pm \epsilon x \in C$. Since $v^t x = 0$, v cannot be a non-zero multiple of $v - \epsilon x$ or $v + \epsilon x$. Now the identity in (4.12) contradicts the assumption that v is an extremal ray. Therefore if v is an extremal ray it follows that the rank of A_I is $d-1$.

On the other hand if the rank of A_I is $d-1$, then there exists a non-zero vector $w \in \mathbb{R}^d$ with

$$\{x \in \mathbb{R}^d \mid A_I x = 0\} = \{\lambda w \mid \lambda \in \mathbb{R}\}.$$

If $v = u_1 + u_2$ for $u_1, u_2 \in C \setminus \{0\}$, then we must have $I(u_1) = I(u_2) = I$. Therefore v, u_1 and u_2 are all proportional to w. We must have $v = \lambda u_1$ or $v = \lambda u_2$ for some $\lambda > 0$ proving that v is an extremal ray. $\square$

DEFINITION 4.2.2
Two extremal rays u and v in

$$C = \{x \in \mathbb{R}^d \mid Ax \leq 0\}$$

are called adjacent if the rank of A_K is $d-2$, where $K = I(u) \cap I(v)$.

Geometrically this means that u and v span a common face of the cone. We will, however, not give the precise definition of a face in a cone.

The reason for introducing the concept of adjacent extremal rays is rather clear when you take the following extension of Lemma 4.1.1 into account.

LEMMA 4.2.3
Consider

$$C = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{rcl} a_1^t x & \leq & 0 \\ & \vdots & \\ a_m^t x & \leq & 0 \end{array} \right\}$$

for $a_1, \ldots, a_m \in \mathbb{R}^n$. Let V be the set of extremal rays in C and $a \in \mathbb{R}^n$. Then

$$\begin{aligned} W = &\{v \in V \mid a^t v < 0\} \bigcup \\ &\quad \{(a^t u)v - (a^t v)u \mid u, v \text{ adjacent in } V, \text{ with } a^t u > 0 \quad \text{and} \quad a^t v < 0\}. \end{aligned}$$

is the set of extremal rays in

$$C \cap \{x \in \mathbb{R}^n \mid a^t x \leq 0\}.$$

Proof. If you compare with Lemma 4.1.1 you will see that we only need to prove for extremal rays u and v of C that

$$w := (a^t u)v - (a^t v)u$$

is extremal if and only if u and v are adjacent, where $a^t u > 0$ and $a^t v < 0$. We assume that the rank of the matrix A consisting of the rows $a_1, \dots, a_m$ is n. Let A' denote the matrix with rows $a_1, \dots, a_m, a_{m+1} := a$. We let $I = I(u), J = I(v)$ and $K = I(w)$ with respect to the matrix A'. Since w is a positive linear combination of u and v we know that $a_i^t w = 0$ if and only if $a_i^t u = a_i^t v = 0$, where a_i is a row of A. Therefore

$$K = (I \cap J) \cup \{m+1\}. \tag{4.13}$$

If u and v are adjacent then $A_{I\cap J}$ has rank $n-2$. The added row a in A' satisfies $a^t w = 0$ and the vector a is not in the span of the rows in $A_{I\cap J}$. This shows that the rank of A'_K is $n-1$. Therefore w is extremal. Suppose on the other hand that w is extremal. This means that A'_K has rank $n-1$. By (4.13) this shows that the rank of $A_{I\cap J}$ has to be $n-2$ proving that u and v are adjacent. □

We will now revisit our previous example and weed out in the generators using Lemma 4.2.3.

EXAMPLE 4.2.4

$$\begin{array}{rrrcl} -x & & & \leq & 0 \\ & -y & & \leq & 0 \\ & & -z & \leq & 0 \\ x & -y & -z & \leq & 0 \end{array} \tag{4.14}$$

Here $a_1 = (-1, 0, 0), a_2 = (0, -1, 0), a_3 = (0, 0, -1), a_4 = (1, -1, -1)$. The extremal rays are

$$V = \{(0,1,0), (1,1,0), (0,0,1), (1,0,1)\}$$

We add the inequality $x - 2y + z \leq 0$ and form the matrix A' with the extra row $a_5 := a = (1, -2, 1)$. The extremal rays are divided into two groups

$$V = \{v \mid a^t v < 0\} \cup \{v \mid a^t v > 0\}$$

corresponding to

$$V = \{(0,1,0), (1,1,0)\} \cup \{(0,0,1), (1,0,1)\}.$$

You can check that

$$\begin{aligned} I((0,1,0)) &= \{1,3\} \\ I((1,1,0)) &= \{3,4\} \\ I((0,0,1)) &= \{1,2\} \\ I((1,0,1)) &= \{2,4\} \end{aligned}$$

From this you see that $(1,1,0)$ is not adjacent to $(0,0,1)$ and that $(0,1,0)$ is not adjacent to $(1,0,1)$. These two pairs correspond exactly to the superfluous rays encountered in Example 4.1.4. So Lemma 4.2.3 tells us that we only need to add the vectors

$$\begin{aligned}(0,1,0)+2(0,0,1)&=(0,1,2)\\ 2(1,1,0)+(1,0,1)&=(3,2,1)\end{aligned}$$

to $\{(0,1,0),(1,1,0)\}$ to get the new extremal rays.

4.3 Farkas: from generators to half spaces

Now suppose that $C = \operatorname{cone}(v_1,\dots,v_m) \subseteq \mathbb{R}^n$. Then we may write $C = \{x \mid Ax \le 0\}$ for a suitable matrix A. This situation is dual to what we have encountered finding the basic solutions to $Ax \le 0$. The key for the translation is Corollary 3.1.4, which says that

$$C = \bigcap_{a\in C^*} H_a. \tag{4.15}$$

The dual cone to C is given by

$$C^* = \{a \in \mathbb{R}^n \mid a^t v_1 \le 0, \dots, a^t v_m \le 0\},$$

which is in fact the solutions to a system of linear inequalities. By Lemma 4.1.1 (and Lemma 4.2.3) you know that these inequalities can be solved using the double description method and that

$$C^* = \operatorname{cone}(a_1,\dots,a_r)$$

for certain (extremal) rays $a_1,\dots,a_r$ in C^*. We claim that

$$C = H_{a_1}\cap\cdots\cap H_{a_r}$$

so that the intersection in (4.15) really is finite! Let us prove this. If

$$a = \lambda_1 a_1 + \cdots + \lambda_j a_j \in C^*$$

with $\lambda_i > 0$, then

$$H_{a_1}\cap\cdots\cap H_{a_j} \subseteq H_a,$$

as $a_1^t x \le 0, \dots, a_j^t x \le 0$ for every $x \in C$ implies $(\lambda_1 a_1^t + \cdots + \lambda_j a_j^t)x = a^t x \le 0$. This proves that

$$H_{a_1}\cap\cdots\cap H_{a_r} = \bigcap_{a\in C^*} H_a = C.$$

EXAMPLE 4.3.1

In Example 4.1.4 we showed that the solutions of

$$\begin{array}{rrrcl} -x & & & \leq & 0 \\ & -y & & \leq & 0 \\ & & -z & \leq & 0 \\ x & -y & -z & \leq & 0 \\ x & -2y & +z & \leq & 0 \end{array} \tag{4.16}$$

are generated by the extremal rays

$$V = \{(0,1,0),(1,1,0),(0,1,2),(3,2,1)\}.$$

Let us go back from the extremal rays above to inequalities. Surely five inequalities in (4.16) are too many for the four extremal rays. We should be able to find four inequalities doing the same job. Here is how. The dual cone to the cone generated by V is given by

$$\begin{array}{rrrcl} & y & & \leq & 0 \\ x & +y & & \leq & 0 \\ & y & +2z & \leq & 0 \\ 3x & +2y & +z & \leq & 0 \end{array}$$

We solve this system of inequalities using Lemma 4.1.1. The set of solutions to

$$\begin{array}{rrrcl} & y & & \leq & 0 \\ x & +y & & \leq & 0 \\ & y & +2z & \leq & 0 \end{array}$$

is generated by the extremal rays $V = \{(2,-2,1),(-1,0,0),(0,0,-1)\}$. We add the fourth inequality $3x + 2y + z \leq 0$ ($a = (3,2,1)$) and split V according to the sign of $a^t v$:

$$V = \{(2,-2,1)\} \cup \{(-1,0,0),(0,0,-1)\}.$$

This gives the new extremal rays

$$\begin{aligned} 3(-1,0,0) + 3(2,-2,1) &= (3,-6,3) \\ 3(0,0,-1) + (2,-2,1) &= (2,-2,-2) \end{aligned}$$

and the extremal rays are $\{(-1,0,0),(0,0,-1),(1,-2,1),(1,-1,-1)\}$ showing that the solutions of (4.16) really are the solutions of

$$\begin{array}{rrrcl} -x & & & \leq & 0 \\ & & -z & \leq & 0 \\ x & -2y & +z & \leq & 0 \\ x & -y & -z & \leq & 0. \end{array}$$

We needed four inequalities and not five! Of course we could have spotted this from the beginning noting that the inequality $-y \leq 0$ is a consequence of the inequalities $-x \leq 0$, $x - 2y + z \leq 0$ and $x - y - z \leq 0$.

4.4 Polyhedra: general linear inequalities

The set of solutions to a general system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &\leq b_1 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &\leq b_m \end{aligned} \tag{4.17}$$

of linear inequalities is a polyhedron. It may come as a surprise to you, but we have already done all the work for studying the structure of polyhedra. The magnificent trick is to adjoin an extra variable x_{n+1} and rewrite (4.17) into the homogeneous system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n - b_1x_{n+1} &\leq 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n - b_mx_{n+1} &\leq 0 \\ -x_{n+1} &\leq 0. \end{aligned} \tag{4.18}$$

The **key observation** is that

$$(x_1, \dots, x_n) \quad \text{solves (4.17)} \qquad \Longleftrightarrow \qquad (x_1, \dots, x_n, 1) \quad \text{solves (4.18).}$$

But (4.18) is a system of (homogeneous) linear inequalities as in (4.2) and we know that the solution set is

$$\operatorname{cone}\left(\begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \dots, \begin{pmatrix} u_r \\ 0 \end{pmatrix}, \begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} v_s \\ 1 \end{pmatrix}\right), \tag{4.19}$$

where $u_1, \dots, u_r, v_1, \dots, v_s \in \mathbb{R}^n$. Notice that we have divided the solutions of (4.18) into $x_{n+1} = 0$ and $x_{n+1} \neq 0$. In the latter case we may assume that $x_{n+1} = 1$ (why?). The solutions with $x_{n+1} = 0$ in (4.18) correspond to the solutions C of the homogeneous system

$$\begin{aligned} a_{11}x_1 + \cdots + a_{1n}x_n &\leq 0 \\ &\vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n &\leq 0. \end{aligned}$$

In particular the solution set to (4.17) is bounded if $C = \{0\}$.

4.5 The decomposition theorem for polyhedra

You cannot expect (4.17) to always have a solution. Consider the simple example

$$\begin{aligned} x &\leq 1 \\ -x &\leq -2 \end{aligned}$$

Adjoining the extra variable y this lifts to the system

$$\begin{aligned} x - y &\leq 0 \\ -x + 2y &\leq 0 \\ -y &\leq 0 \end{aligned}$$

Here $x = 0$ and $y = 0$ is the only solution and we have no solutions with $y = 1$ i.e. we may have $s = 0$ in (4.19). This happens if and only if (4.17) has no solutions.

We have now reached the main result of these notes: a complete characterization of polyhedra due to Minkowski in 1897 (see [5]) and Weyl in 1935 (see [8]). Minkowski showed that a polyhedron admits a description as a sum of a polyhedral cone and a polytope. Weyl proved the other implication: a sum of a polyhedral cone and a polytope is a polyhedron. You probably know by now that you can use the double description method and Farkas' lemma to reason about these problems.

THEOREM 4.5.1 (Minkowski, Weyl)
A non-empty subset $P \subseteq \mathbb{R}^n$ is a polyhedron if and only if it is the sum $P = C + Q$ of a polytope Q and a polyhedral cone C.

Proof. A polyhedron P is the set of solutions to a general system of linear inequalities as in (4.17). If P is non-empty, then $s \geq 1$ in (4.19). This shows that

$$P = \{\lambda_1 u_1 + \cdots + \lambda_r u_r + \mu_1 v_1 + \cdots + \mu_s v_s \mid \lambda_i \geq 0,\ \mu_j \geq 0, \text{ and } \mu_1 + \cdots + \mu_s = 1\}$$

or

$$P = \operatorname{cone}(u_1, \ldots, u_r) + \operatorname{conv}(v_1, \ldots, v_s). \tag{4.20}$$

On the other hand, if P is the sum of a cone and a polytope as in (4.20), then we define $\hat{P} \subseteq \mathbb{R}^{n+1}$ to be the cone generated by

$$\begin{pmatrix} u_1 \\ 0 \end{pmatrix}, \ldots, \begin{pmatrix} u_r \\ 0 \end{pmatrix}, \begin{pmatrix} v_1 \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} v_s \\ 1 \end{pmatrix}$$

If

$$\hat{P}^* = \operatorname{cone}\Big(\begin{pmatrix} \alpha_1 \\ \beta_1 \end{pmatrix}, \dots, \begin{pmatrix} \alpha_N \\ \beta_N \end{pmatrix} \Big),$$

where $\alpha_1, \dots, \alpha_n \in \mathbb{R}^n$ and $\beta_1, \dots, \beta_n \in \mathbb{R}$, we know from §4.3 that

$$\hat{P} = \Big\{ \begin{pmatrix} x \\ z \end{pmatrix} \in \mathbb{R}^{n+1} \,\Big|\, \alpha_1^t x + \beta_1 z \leq 0, \dots, \alpha_N^t x + \beta_N z \leq 0 \Big\}.$$

But an element $x \in \mathbb{R}^n$ belongs to P if and only if $(x, 1) \in \hat{P}$. Therefore

$$\begin{aligned} P &= \Big\{ x \in \mathbb{R}^n \,\Big|\, \alpha_1^t x + \beta_1 \leq 0, \dots, \alpha_N^t x + \beta_N \leq 0 \Big\} \\ &= \Big\{ x \in \mathbb{R}^n \,\Big|\, \alpha_1^t x \leq -\beta_1, \dots, \alpha_N^t x \leq -\beta_N \Big\} \end{aligned}$$

is a polyhedron. ☐

4.6 Extremal points in polyhedra

There is a natural connection between the extremal rays in a cone and the extremal points in a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. Let $a_1, \dots, a_m$ denote the rows in A and let $b = (b_1, \dots, b_m)^t$. With this notation we have

$$P = \{x \in \mathbb{R}^n \mid Ax \leq b\} = \left\{ x \in \mathbb{R}^n \middle| \begin{array}{c} a_1^t\, x \leq b_1 \\ \vdots \\ a_m^t\, x \leq b_m \end{array} \right\}.$$

For $z \in P$ we define the submatrix

$$A_z = \{a_i \mid a_i^t\, z = b_i\},$$

consisting of those rows where the inequalities are equalities (binding constraints) for z. The following result shows that a polyhedron only contains finitely many extremal points and gives a method for finding them.

PROPOSITION 4.6.1
$z \in P$ is an extremal point if and only if A_z has full rank n.

Proof. The proof is very similar to Proposition 4.2.1 and we only sketch the details. If $z \in P$ and the rank of A_z is $< n$, then we may find $u \neq 0$ with $A_z u = 0$. We can then choose $\epsilon > 0$ sufficiently small so that $z \pm \epsilon u \in P$ proving that

$$z = (1/2)(z + \epsilon u) + (1/2)(z - \epsilon u)$$

cannot be an extremal point. This shows that if z is an extremal point, then A_z must have full rank n. On the other hand if A_z has full rank n and $z = (1-\lambda)z_1 + \lambda z_2$ with $0 < \lambda < 1$ for $z_1, z_2 \in P$, then we have for a row a_i in A_z that

$$a_i^t z = (1-\lambda)\, a_i^t z_1 + \lambda\, a_i^t z_2.$$

As $a_i^t z_1 \leq a_i^t z = b_i$ and $a_i^t z_2 \leq b_i$ we must have $a_i^t z_1 = a_i^t z_2 = b_i = a_i^t z$. But then $A_z(z - z_1) = 0$ i.e. $z = z_1$. $\square$

EXAMPLE 4.6.2

Find the extremal points in

$$P = \left\{ \begin{pmatrix} x \\ y \end{pmatrix} \in \mathbb{R}^2 \,\middle|\, \begin{pmatrix} -1 & -1 \\ 2 & -1 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \leq \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \right\}.$$

We will find the "first" extremal point and leave the computation of the other extremal points to the reader. First we try and see if we can find $z \in P$ with

$$A_z = \{a_1, a_2\} = \begin{pmatrix} -1 & -1 \\ 2 & -1 \end{pmatrix}.$$

If $z = (x, y)^t$ this leads to solving

$$\begin{aligned} -x - y &= 0 \\ 2x - y &= 1, \end{aligned}$$

giving $(x, y) = (1/3, -1/3)$. Since $-1/3 + 2 \cdot (1/3) = 1/3 < 1$ we see that $z = (1/3, -1/3) \in P$. This shows that z is an extremal point in P. Notice that the rank of A_z is 2.

Before moving on to the next result, notice that the characteristic cone of a polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ is given by

$$\operatorname{ccone}(P) = \{x \in \mathbb{R}^n \mid Ax \leq 0\}.$$

DEFINITION 4.6.3

A subset

$$L = \{u + tv \mid t \in \mathbb{R}\}$$

with $u, v \in \mathbb{R}^n$ and $v \neq 0$ is called a line in $\mathbb{R}^n$.

THEOREM 4.6.4

Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\} \neq \emptyset$. The following conditions are equivalent *(i) P contains an extremal point.*

(ii) The characteristic cone ccone(P) *does not contain a line.*

(iii) P does not contain a line.

Proof. If $z \in P$ and ccone(P) contains a line $L = \{v + tu \mid t \in \mathbb{R}\}$, we must have $Au = 0$. Therefore

$$z = 1/2(z + u) + 1/2(z - u),$$

where $z \pm u \in P$ and none of the points in P are extremal. Suppose on the other hand that ccone(P) does not contain a line. Then P does not contain a line, since a line L as above inside P implies $Au = 0$ making it a line inside ccone(P).

Now assume that P does not contain a line and consider $z \in P$. If A_z has rank n then z is an extremal point. If not we can find a non-zero u with $A_z u = 0$. Since P does not contain a line we must have

$$z + \lambda u \notin P$$

for λ sufficiently big. Let

$$\lambda_0 = \sup\{\lambda \mid z + \lambda u \in P\}$$

and

$$z_1 = z + \lambda_0 u.$$

Then $z_1 \in P$ and the rank of A_{z_1} is strictly greater than the rank of A_z. If the rank of A_{z_1} is not n we continue the procedure. Eventually we will hit an extremal point. □

COROLLARY 4.6.5

Let $c \in \mathbb{R}^n$ and $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$. If

$$M = \sup\{c^t x \mid x \in P\} < \infty$$

and ext$(P) \neq \emptyset$*, then there exists $x_0 \in$* ext(P) *such that*

$$c^t x_0 = M.$$

Proof. The non-empty set

$$Q = \{x \in \mathbb{R}^n \mid c^t x = M\} \cap P$$

is a polyhedron not containing a line, since P does not contain a line. Therefore Q contains an extremal point z. But such an extremal point is also an extremal point in P. If this was not so, we could find a non-zero u with $A_z u = 0$. For $\epsilon > 0$ small we would then have $z \pm \epsilon u \in P$. But this implies that $c^t\, u = 0$ and $z \pm \epsilon u \in Q$. The well known identity

$$z = \frac{1}{2}(z + \epsilon u) + \frac{1}{2}(z - \epsilon u)$$

shows that z is not an extremal point in Q. This is a contradiction. □

Now we have the following refinement of Theorem 4.5.1 for polyhedra with extremal points.

THEOREM 4.6.6
Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ *be a polyhedron with* $\operatorname{ext}(P) \neq \emptyset$. *Then*

$$P = \operatorname{conv}(\operatorname{ext}(P)) + \operatorname{ccone}(P).$$

Proof. The definition of $\operatorname{ccone}(P)$ shows that

$$\operatorname{conv}(\operatorname{ext}(P)) + \operatorname{ccone}(P) \subseteq P.$$

You also get this from $\operatorname{ccone}(P) = \{x \in \mathbb{R}^n \mid Ax \leq 0\}$. If $P \neq \operatorname{conv}(\operatorname{ext}(P)) + \operatorname{ccone}(P)$, there exists $z \in P$ and $c \in \mathbb{R}^n$ such that $c^t\, z > c^t\, x$ for every $x \in \operatorname{conv}(\operatorname{ext}(P)) + \operatorname{ccone}(P)$. But this contradicts Corollary 4.6.5, which tells us that there exists an extremal point z_0 in P with

$$c^t\, z_0 = \sup\{c^t\, x \mid x \in P\}.$$

4.7 Exercises

(1) Verify that the set of solutions to (4.6) is as described in (4.7).

(2) Find the set of solutions to the system

$$\begin{array}{ccccc} x & + & & z & \leq 0 \\ & & y & + \; z & \leq 0 \\ & & & z & \leq 0 \end{array}$$

of (homogeneous) linear inequalities.

(3) Express the convex hull

$$\operatorname{conv}\left\{ \begin{pmatrix} 1/4 \\ 1/4 \\ -1/2 \end{pmatrix}, \begin{pmatrix} -1/2 \\ 1/4 \\ 1/4 \end{pmatrix}, \begin{pmatrix} 1/4 \\ -1/2 \\ 1/4 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right\} \subseteq \mathbb{R}^3$$

as a polyhedron (an intersection of halfspaces).

(4) Convert the inequalities

$$\begin{array}{rcrcr} x & + & y & \leq & 1 \\ -x & + & y & \leq & -1 \\ x & - & 2y & \leq & -2 \end{array}$$

to a set of 4 homogeneous inqualities by adjoining an extra variable z. Show that the original inequalities are unsolvable using this.

(5) Is the set of solutions to

$$\begin{array}{rcrcrcr} -x & + & 2y & - & z & \leq & 1 \\ -x & - & y & - & z & \leq & -2 \\ 2x & - & y & - & z & \leq & 1 \\ & - & y & + & z & \leq & 1 \\ -x & - & y & + & z & \leq & 0 \\ -x & + & y & & & \leq & 1 \end{array}$$

bounded in $\mathbb{R}^3$?

(6) Let P_1 and P_2 be polytopes in $\mathbb{R}^n$. Show that

$$P_1 + P_2 = \{u + v \mid u \in P_1, v \in P_2\}$$

is a polytope. Show that the sum of two polyhedra is a polyhedron.

(7) Give an example of a polyhedron with no extremal points.

(8) Let $P = \{x \in \mathbb{R}^n \mid Ax \leq b\}$ be a polyhedron, where A is an $m \times n$ matrix and $b \in \mathbb{R}^m$. How many extremal points can P at the most have?

(9) Show that a polyhedron is a polytope (bounded) if and only if it is the convex hull of its extremal points.

(10) Let K be a closed convex set in $\mathbb{R}^n$. Show that K contains a line if and only if $\text{ccone}(K)$ contains a line.

(11) Give an example showing that Theorem 4.6.6 is far from true if $\text{ext}(P) = \emptyset$.

(12) Let

$$P = \text{conv}(u_1, \ldots, u_r) + \text{cone}(v_1, \ldots, v_s)$$

be a polyhedron, where $u_1, \ldots, u_r, v_1, \ldots, v_s \in \mathbb{R}^n$. Show for $c \in \mathbb{R}^n$ that if $M = \sup\{c^t\, x \mid x \in P\} < \infty$, then

$$\sup_{x \in P} c^t\, x = \sup_{x \in K} c^t\, x,$$

where $K = \text{conv}(u_1, \ldots, u_r)$. Does there exist $x_0 \in P$ with $c^t\, x_0 = M$?

Appendix A

Linear (in)dependence

The concept of linear (in)dependence is often a stumbling block in introductory courses on linear algebra. When presented as a sterile definition in an abstract vector space it can be hard to grasp. I hope to show here that it is simply a fancy way of restating a quite obvious fact about solving linear equations.

A.1 Linear dependence and linear equations

You can view the equation

$$3x + 5y = 0$$

as one linear equation with two unknowns. Clearly $x = y = 0$ is a solution. But there is also a non-zero solution with $x = -5$ and $y = 3$. As one further example consider

$$\begin{aligned} 2x + y - z &= 0 \\ x + y + z &= 0 \end{aligned} \tag{A.1}$$

Here we have 3 unknowns and only 2 equations and $x = 2, y = -3$ and $z = 1$ is a non-zero solution.

These examples display a fundamental fact about linear equations. A system

$$\begin{aligned} a_{11}\, x_1 + \cdots + a_{1n}\, x_n &= 0 \\ a_{21}\, x_1 + \cdots + a_{2n}\, x_n &= 0 \\ &\vdots \\ a_{m1}\, x_1 + \cdots + a_{mn}\, x_n &= 0 \end{aligned}$$

of linear equations always has a non-zero solution if the number of unknowns n is greater than the number n of equations i.e. $n > m$.

In modern linear algebra this fact about linear equations is coined using the abstract term "linear dependence":

> *A set of vectors* $\{v_1, \dots, v_n\} \subset \mathbb{R}^m$ *is linearly dependent if* $n > m$. *This means that there exists* $\lambda_1, \dots, \lambda_n \in \mathbb{R}$ *not all* 0 *such that*
>
> $$\lambda_1 v_1 + \cdots + \lambda_n v_n = 0.$$

With this language you can restate the non-zero solution $x = 2, y = -3$ and $z = 1$ of (A.1) as the linear dependence

$$2 \cdot \begin{pmatrix} 2 \\ 1 \end{pmatrix} + (-3) \cdot \begin{pmatrix} 1 \\ 1 \end{pmatrix} + 1 \cdot \begin{pmatrix} -1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Let us give a simple induction proof of the fundamental fact on (homogeneous) systems of linear equations.

THEOREM A.1.1
The system

$$\begin{aligned} a_{11}\, x_1 + \cdots + a_{1n}\, x_n &= 0 \\ a_{21}\, x_1 + \cdots + a_{2n}\, x_n &= 0 \\ &\vdots \\ a_{m1}\, x_1 + \cdots + a_{mn}\, x_n &= 0 \end{aligned} \tag{A.2}$$

of linear equations always has a non-zero solution if $m < n$.

Proof. The induction is on m — the number of equations. For $m = 1$ we have 1 linear equation

$$a_1 x_1 + \cdots + a_n x_n = 0$$

with n variables where $n > m = 1$. If $a_i = 0$ for some $i = 1, \dots, n$ then clearly $x_i = 1$ and $x_j = 0$ for $j \neq i$ is a non-zero solution. Assume otherwise that $a_i \neq 0$ for every $i = 1, \dots, m$. In this case $x_1 = 1, x_2 = -a_1/a_2, x_3 = \cdots = x_n = 0$ is a non-zero solution.

If every $a_{i1} = 0$ for $i = 1, \dots, m$, then $x_1 = 1, x_2 = \cdots = x_n = 0$ is a non-zero solution in (A.2). Assume therefore that $a_{11} \neq 0$ and substitute

$$x_1 = \frac{1}{a_{11}}(-a_{12}x_2 - \cdots - a_{1n}x_n)$$

x_1 into the remaining $m-1$ equations. This gives the following system of $m-1$ equations in the $n-1$ variables $x_2, \dots, x_n$

$$\begin{aligned} (a_{22} - \frac{a_{21}}{a_{11}} a_{12}) x_2 + \cdots + (a_{2n} - \frac{a_{21}}{a_{11}} a_{1n}) x_n &= 0 \\ &\vdots \\ (a_{m2} - \frac{a_{m1}}{a_{11}} a_{12}) x_2 + \cdots + (a_{mn} - \frac{a_{m1}}{a_{11}} a_{1n}) x_n &= 0 \end{aligned} \tag{A.3}$$

Since $n-1 > m-1$, the induction assumption on m gives the existence of a non-zero solution $(a_2, \dots, a_n)$ to (A.3). Now

$$(\frac{1}{a_{11}}(-a_{12}a_2 - \cdots - a_{1n}a_n), a_2, \dots, a_n)$$

is a non-zero solution to our original system of equations. This can be checked quite explicitly (Exercise 4). □

A.2 The rank of a matrix

A system

$$\begin{aligned} a_{11}\, x_1 + \cdots + a_{1n}\, x_n &= 0 \\ a_{21}\, x_1 + \cdots + a_{2n}\, x_n &= 0 \\ &\vdots \\ a_{m1}\, x_1 + \cdots + a_{mn}\, x_n &= 0 \end{aligned} \tag{A.4}$$

of linear equations can be conveniently presented in the matrix form

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

where A is the $m \times n$ matrix

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

We need to attach a very important invariant to A called *the rank* of the matrix. In the context of systems of linear equations the rank is very easy to understand. Let us shorten our notation a bit and let

$$L_i(x) = a_{i1}x_1 + \cdots + a_{in}x_n.$$

Then the solutions to (A.4) are

$$S = \{x \in \mathbb{R}^n \mid L_1(x) = 0, \ldots, L_m(x) = 0\}.$$

Suppose that $m = 3$. If one of the equations say $L_3(x)$ is expressible by the other equations say as

$$L_3(x) = \lambda L_1(x) + \mu L_2(x),$$

with $\lambda, \mu \in \mathbb{R}$, then we don't need the equation L_3 in S. This is because

$$\begin{aligned} L_1(x) &= 0 \\ L_2(x) &= 0 \\ L_3(x) &= 0 \end{aligned} \qquad \Longleftrightarrow \qquad \begin{aligned} L_1(x) &= 0 \\ L_2(x) &= 0 \end{aligned}$$

Clearly you see that $L_3(x) = 0$ if $L_1(x) = 0$ and $L_2(x) = 0$. In this case you can throw L_3 out without changing S. Informally the rank of the matrix A is the minimal number of equations you end up with after throwing excess equations away. It is not too hard to make this into a very well defined concept. We will only use the following very important consequence.

THEOREM A.2.1

Let A be an $m \times n$ matrix of rank $< n$. Then there exists a non-zero vector $u \in \mathbb{R}^n$ with $Au = 0$.

Proof. You may view $S = \{x \in \mathbb{R}^n \mid Ax = 0\}$ as the set of solutions to

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

By our informal definition of rank we know that

$$S = \{x \in \mathbb{R}^n \mid A_I x = 0\},$$

where A_I is an $m' \times n$- matrix consisting of a subset of the rows in A with $m' =$ the rank of A. Now the result follows by applying Theorem A.1.1. ☐

A.3 Exercises

(1) Find $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ not all 0 with

$$\lambda_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 3 \\ 4 \end{pmatrix} + \lambda_3 \begin{pmatrix} 5 \\ 6 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

(2) Show that a non-zero solution (x, y, z) to (A.1) must have $x \neq 0, y \neq 0$ and $z \neq 0$. Is it possible to find $\lambda_1, \lambda_2, \lambda_3$ in Exercise 1, where one of λ_1, λ_2 or λ_3 is 0?

(3) Can you find a non-zero solution to

$$\begin{aligned} x + y + z &= 0 \\ x - y + z &= 0, \end{aligned}$$

where

(i) $x = 0$?

(ii) $y = 0$?

(iii) $z = 0$?

(iv) What can you say in general about a system

$$\begin{aligned} ax + by + cz &= 0 \\ a'x + b'y + c'z &= 0 \end{aligned}$$

of linear equations in x, y and z, where a non-zero solution always has $x \neq 0, y \neq 0$ and $z \neq 0$?

(4) Check carefully that

$$(\frac{1}{a_{11}}(-a_{12}a_2 - \cdots - a_{1n}a_n), a_2, \ldots, a_n)$$

really is a non-zero solution to (A.2) in the proof of Theorem A.1.1.

(5) Compute the rank of the matrix

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 14 & 19 & 24 \\ 6 & 9 & 12 \end{pmatrix}.$$

Appendix B

Analysis

In this appendix we give a very brief overview of the basic concepts of introductory mathematical analysis. Focus is directed at building things from scratch with applications to convex sets. We have not formally constructed the real numbers.

B.1 Measuring distances

The limit concept is a cornerstone in mathematical analysis. We need a formal way of stating that two vectors are far apart or close together.

Inspired by the Pythagorean formula for the length of the hypotenuse in a triangle with a right angle, we define the length $|x|$ of a vector $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ as

$$|x| = \sqrt{x_1^2 + \cdots + x_n^2}.$$

Our first result about the length is the following lemma called the inequality of Cauchy-Schwarz. It was discovered by Cauchy[1] in 1821 and rediscovered by Schwarz[2] in 1888.

LEMMA B.1.1

For $x = (x_1, \ldots, x_n)^t \in \mathbb{R}^n$ and $y = (y_1, \ldots, y_n)^t \in \mathbb{R}^n$ the inequality

$$(x^t y)^2 = (x_1 y_1 + \cdots + x_n y_n)^2 \leq (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = |x|^2 |y|^2$$

holds. If

$$(x^t y)^2 = (x_1 y_1 + \cdots + x_n y_n)^2 = (x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = |x|^2 |y|^2,$$

then x and y are proportional i.e. $x = \lambda y$ for some $\lambda \in \mathbb{R}$.

---

[1] Augustin Louis Cauchy (1789–1857), French mathematician

[2] Hermann Amandus Schwarz (1843–1921), German mathematician

Proof. For $n = 2$ you can explicitly verify that

$$(x_1^2 + x_2^2)(y_1^2 + y_2^2) - (x_1y_1 + x_2y_2)^2 = (x_1y_2 - y_1x_2)^2. \tag{B.1}$$

This proves that inequality for $n = 2$. If equality holds, we must have

$$x_1y_2 - y_1x_2 = \begin{vmatrix} x_1 & y_1 \\ x_2 & y_2 \end{vmatrix} = 0.$$

This implies as you can check that there exists $\lambda \in \mathbb{R}$ such that $x_1 = \lambda y_1$ and $x_2 = \lambda y_2$.

The formula in (B.1) generalizes for $n > 2$ by induction (Exercise 1) to

$$\begin{aligned}&(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) - (x_1y_1 + \cdots + x_ny_n)^2 = \\ &(x_1y_2 - y_1x_2)^2 + \cdots + (x_{n-1}y_n - y_{n-1}x_n)^2,\end{aligned} \tag{B.2}$$

where the last sum is over the squares of the 2×2 minors in the matrix

$$A = \begin{pmatrix} x_1 & x_2 & \cdots & x_{n-1} & x_n \\ y_1 & y_2 & \cdots & y_{n-1} & y_n \end{pmatrix}.$$

The formula in (B.2) proves the inequality. If

$$(x_1^2 + \cdots + x_n^2)(y_1^2 + \cdots + y_n^2) = (x_1y_1 + \cdots + x_ny_n)^2,$$

then (B.2) shows that all the 2×2-minors in A vanish. The existence of λ giving proportionality between x and y is deduced as for $n = 2$. ☐

If you know about the vector (cross) product $u \times v$ of two vectors $u, v \in \mathbb{R}^3$ you will see that the method of the above proof comes from the formula

$$|u|^2|v|^2 = |u^t v|^2 + |u \times v|^2.$$

One of the truly fundamental properties of the length of a vector is the triangle inequality (also inspired by the one in 2 dimensions).

THEOREM B.1.2

For two vectors $x, y \in \mathbb{R}^n$ the inequality

$$|x + y| \leq |x| + |y|$$

holds.

Proof. Lemma B.1.1 shows that

$$\begin{aligned}|x+y|^2 &= (x+y)^t(x+y)\\ &= |x|^2+|y|^2+2x^ty\\ &\le |x|^2+|y|^2+2|x||y|\\ &= (|x|+|y|)^2\end{aligned}$$

proving the inequality. □

We need to define how far vectors are apart.

DEFINITION B.1.3

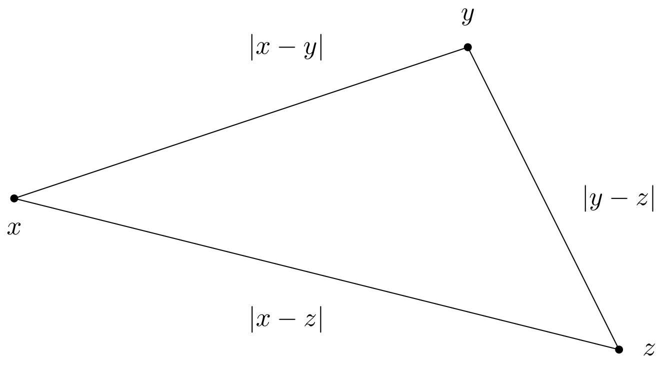
The distance between x and y in $\mathbb{R}^n$ is defined as

$$|x-y|.$$

From Theorem B.1.2 you get formally

$$|x-z| = |x-y+y-z| \le |x-y| + |y-z|$$

for $x, y, z \in \mathbb{R}^n$. This is the triangle inequality for distance saying that the shorter way is always along the diagonal instead of the other two sides in a triangle:



B.2 Sequences

Limits appear in connection with (infinite) sequences of vectors in $\mathbb{R}^n$. We need to formalize this.

DEFINITION B.2.1

A sequence in $\mathbb{R}^n$ is a function $f : \{1, 2, \ldots\} \to \mathbb{R}^n$. A subsequence f_I of f is f restricted to an infinite subset $I \subseteq \{1, 2, \ldots\}$.

A sequence f is usually denoted by an infinite tuple $(x_n) = (x_1, x_2, \dots)$, where $x_n = f(n)$. A subsequence of (x_n) is denoted (x_{n_i}), where $I = \{n_1, n_2, \dots\}$ and $n_1 < n_2 < \cdots$. A subsequence f_I is in itself a sequence, since it is given by picking out an infinite subset I of $\{1, 2, \dots\}$ and then letting $f_I(j) = f(n_j) = x_{n_j}$.

This definition is quite formal. Once you get to work with it, you will discover that it is easy to handle. In practice sequences are often listed as

$$1,\ 2,\ 3,\ 4,\ 5,\ 6,\ \dots \tag{B.3}$$

$$2,\ 4,\ 6,\ 8,\ 10,\ \dots \tag{B.4}$$

$$2,\ 6,\ 4,\ 8,\ 10,\ \dots \tag{B.5}$$

$$1,\ \frac{1}{2},\ \frac{1}{3},\ \frac{1}{4},\ \dots \tag{B.6}$$

Formally these sequences are given in the table below

	x_1	x_2	x_3	x_4	x_5	x_6	$\cdots$
(B.3)	1	2	3	4	5	6	$\cdots$
(B.4)	2	4	6	8	10	12	$\cdots$
(B.5)	2	6	4	8	10	12	$\cdots$
(B.6)	1	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{5}$	$\frac{1}{6}$	$\cdots$

The sequence (z_n) in (B.4) is a subsequence of the sequence in (x_n) (B.3). You can see this by noticing that $z_n = x_{2n}$ and checking with the definition of a subsequence. Why is the sequence in (B.5) not a subsequence of (x_n)?

DEFINITION B.2.2

A sequence (x_n) of real numbers is called increasing *if $x_1 \leq x_2 \leq \cdots$ and* decreasing *if $x_1 \geq x_2 \geq \cdots$.*

The sequences (B.3) and (B.4) are increasing. The sequence (B.6) is decreasing, whereas (B.5) is neither increasing nor decreasing.

You probably agree that the following lemma is very intuitive.

LEMMA B.2.3

Let T be an infinite subset of $\{1, 2, \dots\}$ and F a finite subset. Then $T \setminus F$ is infinite.

However, infinity should be treated with the greatest respect in this setting. It sometimes leads to really surprising statements such as the following.

LEMMA B.2.4

A sequence (x_n) of real numbers always contains an increasing or a decreasing subsequence.

Proof. We will prove that if (x_n) does not contain an increasing subsequence, then it must contain a decreasing subsequence (x_{n_i}), with

$$x_{n_1} > x_{n_2} > x_{n_3} > \cdots$$

The key observation is that if (x_n) does not contain an ascending subsequence, then there exists N_0 such that $X_N > x_n$ for every $n > N$. If this was not so, (x_n) would contain an increasing subsequence. You can try this out yourself!

The first element in our subsequence will be X_N. Now we pick $N_1 > N$ such that $x_n < X_{N_1}$ for $n > N_1$. We let the second element in our subsequence be x_{N_1} and so on. We use nothing but Lemma B.2.3 in this process. If the process should come to a halt after a finite number of steps $x_{n_1}, x_{n_2}, \ldots, x_{n_k}$, then the sequence (x_j) with $j \geq n_k$ must contain an increasing subsequence, which is also an increasing subsequence of (x_n). This is a contradiction. ☐

DEFINITION B.2.5
A sequence (x_n) converges to x (this is written $x_n \to x$) if

$$\forall \epsilon > 0 \, \exists N \in \mathbb{N} \, \forall n \geq N : |x - x_n| \leq \epsilon.$$

Such a sequence is called convergent.

This is a very formal (but necessary!) way of expressing that ...

the bigger n gets, the closer x_n is to x.

You can see that (B.3) and (B.4) are not convergent, whereas (B.6) converges to 0. To practice the formal definition of convergence you should (Exercise 4) prove the following proposition.

PROPOSITION B.2.6
Let (x_n) and (y_n) be sequences in $\mathbb{R}^n$. Then the following hold.

(i) If $x_n \to x$ and $x_n \to x'$, then $x = x'$.

(ii) If $x_n \to x$ and $y_n \to y$, then

$$x_n + y_n \to x + y \qquad \text{and} \qquad x_n y_n \to xy.$$

Before moving on with the more interesting aspects of convergent sequences we need to recall the very soul of the real numbers.

B.2.1 Supremum and infimum

A subset $S \subseteq \mathbb{R}$ is *bounded from above* if there exists $U \in \mathbb{R}$ such that $x \leq U$ for every $x \in S$. Similarly S is *bounded from below* if there exists $L \in \mathbb{R}$ such that $L \leq x$ for every $x \in S$.

THEOREM B.2.7

Let $S \subseteq \mathbb{R}$ be a subset bounded from above. Then there exists a number (supremum) $\sup(S) \in \mathbb{R}$, such that

(i) $x \leq \sup(S)$ *for every* $x \in S$.

(ii) For every $\epsilon > 0$, there exists $x \in S$ such that

$$x > \sup(S) - \epsilon.$$

Similarly we have for a bounded below subset S that there exists a number (infimum) $\inf(S)$ such that

(i) $x \geq \inf(S)$ *for every* $x \in S$.

(ii) For every $\epsilon > 0$, there exists $x \in S$ such that

$$x < \inf(S) + \epsilon.$$

Let $S = \{x_n \mid n = 1, 2, \dots\}$, where (x_n) is a sequence. Then (x_n) is bounded from above from if S is bounded from above, and similarly bounded from below if S is bounded from below.

LEMMA B.2.8

Let (x_n) be a sequence of real numbers. Then (x_n) is convergent if

(i) (x_n) is increasing and bounded from above.

(ii) (x_n) is decreasing and bounded from below.

Proof. In the increasing case $\sup\{x_n \mid n = 1, 2, \dots\}$ is the limit. In the decreasing case $\inf\{x_n \mid n = 1, 2, \dots\}$ is the limit. ☐

B.3 Bounded sequences

A sequence of real numbers is called bounded if it is both bounded from above and below.

COROLLARY B.3.1
A bounded sequence of real numbers has a convergent subsequence.

Proof. This is a consequence of Lemma B.2.4 and Lemma B.2.8. ☐

We want to generalize this result to $\mathbb{R}^m$ for $m > 1$. Surprisingly this is not so hard once we use the puzzling properties of infinite sets. First we need to define bounded subsets here.

A subset $S \subseteq \mathbb{R}^m$ is called *bounded* if there exists $R > 0$ such that $|x| \leq R$ for every $x \in S$. This is a very natural definition. You want your set S to be contained in vectors of length bounded by R.

THEOREM B.3.2
A bounded sequence (x_n) in $\mathbb{R}^m$ has a convergent subsequence.

Proof. Let the sequence be given by

$$x_n = (x_{1n}, \ldots, x_{mn}) \in \mathbb{R}^m.$$

The m sequences of coordinates $(x_{1n}), \ldots, (x_{mn})$ are all bounded sequences of real numbers. So the first one (x_{1n}) has a convergent subsequence (x_{1n_i}). Nothing is lost in replacing (x_n) with its subsequence (x_{n_i}). Once we do this we know that the first coordinate converges! Move on to the sequence given by the second coordinate and repeat the procedure. Eventually we end with a convergent subsequence of the original sequence. ☐

B.4 Closed subsets

DEFINITION B.4.1
A subset $F \subseteq \mathbb{R}^n$ is closed if for any convergent sequence (x_n) with

(i) $(x_n) \subseteq F$

(ii) $x_n \to x$

we have $x \in F$.

Clearly $\mathbb{R}^n$ is closed. Also an arbitrary intersection of closed sets is closed.

DEFINITION B.4.2
The closure of a subset $S \subseteq \mathbb{R}^n$ is defined as

$$\overline{S} = \{x \in \mathbb{R}^n \mid x_n \to x, \text{ where } (x_n) \subseteq S \text{ is a convergent sequence}\}.$$

The points of $\overline{S}$ are simply the points you can reach with convergent sequences from S. Therefore the following result must be true.

PROPOSITION B.4.3
Let $S \subseteq \mathbb{R}^n$. Then $\overline{S}$ is closed.

Proof. Consider a convergent sequence $(y_m) \subseteq \overline{S}$ with $y_m \to y$. We wish to prove that $y \in \overline{S}$. By definition there exists for each y_m a convergent sequence $(x_{m,n}) \subseteq S$ with

$$x_{m,n} \to y_m.$$

For each m we pick $x_m := x_{m,n}$ for n big enough such that $|y_m - x_m| < 1/m$. We claim that $x_m \to y$. This follows from the inequality

$$|y - x_m| = |y - y_m + y_m - x_m| \leq |y - y_m| + |y_m - x_m|,$$

using that $y_m \to y$ and $|y_m - x_m|$ being small for $m \gg 0$. □

The following proposition comes in very handy.

PROPOSITION B.4.4
Let $F_1, \dots, F_m \subseteq \mathbb{R}^n$ be finitely many closed subsets. Then

$$F := F_1 \cup \cdots \cup F_m \subseteq \mathbb{R}^n$$

is a closed subset.

Proof. Let $(x_n) \subseteq F$ denote a convergent sequence with $x_n \to x$. We must prove that $x \in F$. Again distributing infinitely many elements in finitely many boxes implies that one box must contain infinitely many elements. Here this means that at least one of the sets

$$\mathbb{N}_i = \{n \in \mathbb{N} \mid x_n \in F_i\}, \ \ i = 1, \dots, m$$

must be infinite. If $\mathbb{N}_k$ inifinite then $\{x_j \mid j \in \mathbb{N}_k\}$ is a convergent (why?) subsequence of (x_n) with elements in F_k. But F_k is closed so that $x \in F_k \subseteq F$.□

B.5 The interior and boundary of a set

The interior S° of a subset $S \subseteq \mathbb{R}^n$ consists of the elements which are not limits of sequences of elements outside S. The boundary ∂S consists of the points which can be approximated both from the inside and outside. This is formalized in the following definition.

DEFINITION B.5.1
Let $S \subseteq \mathbb{R}^n$. Then the interior S° of S is

$$\mathbb{R}^n \setminus \overline{\mathbb{R}^n \setminus S}.$$

The boundary ∂S is

$$\overline{S} \cap \overline{\mathbb{R}^n \setminus S}.$$

These terse definitions call for some words. The interior S° is the set of points in S, which are not limits of sequences coming from the complement of S. Informally they are the points which cannot be approximated arbitrarily well with points outside S.

The boundary are precisely those points that are limits of sequences of points from S and sequences from the the complement of S. Informally they are those point that can be both approximated arbitrarily well from inside and outside S.

B.6 Continuous functions

DEFINITION B.6.1
A function

$$f : S \to \mathbb{R}^n,$$

where $S \subseteq \mathbb{R}^m$ is called continuous if $f(x_n) \to f(x)$ for every convergent sequence $(x_n) \subseteq S$ with $x_n \to x \in S$.

We would like the length function to be continuous. This is the content of the following proposition.

PROPOSITION B.6.2
The length function $f(x) = |x|$ is a continuous function from $\mathbb{R}^n$ to $\mathbb{R}$.

Proof. You can deduce from the triangle inequality that

$$||x| - |y|| \leq |x - y|$$

for every $x, y \in \mathbb{R}^n$. This shows that

$$|f(x) - f(x_n)| \leq |x - x_n|$$

proving that $f(x_n) \to f(x)$ if $x_n \to x$. Therefore $f(x) = |x|$ is a continuous function. □

The following result is very useful for proving that certain subsets are closed.

LEMMA B.6.3

If $f : \mathbb{R}^m \to \mathbb{R}^n$ is continuous then

$$f^{-1}(F) = \{x \in \mathbb{R}^n \mid f(x) \in F\} \subseteq \mathbb{R}^m$$

is a closed subset, where F is a closed subset of $\mathbb{R}^n$.

Proof. If $(x_n) \subseteq f^{-1}(F)$ with $x_n \to x$, then $f(x_n) \to f(x)$ by the continuity of f. As F is closed we must have $f(x) \in F$. Therefore $x \in f^{-1}(F)$. ☐

B.7 The main theorem

A closed and bounded subset $C \subseteq \mathbb{R}^n$ is called *compact*. Even though the following theorem is a bit dressed up, its applications are many and quite down to Earth. Don't fool yourself by the simplicity of the proof. The proof is only simple because we have the right definitions.

THEOREM B.7.1

Let $f : C \to \mathbb{R}^n$ be a continuous function, where $C \subseteq \mathbb{R}^m$ is compact. Then $f(C) = \{f(x) \mid x \in C\}$ is compact in $\mathbb{R}^n$.

Proof. Suppose that $f(C)$ is not bounded. Then we may find a sequence $(x_n) \subseteq C$ such that $|f(x_n)| \geq n$. However, by Theorem B.3.2 we know that (x_n) has a convergent subsequence (x_{n_i}) with $x_{n_i} \to x$. Since C is closed we must have $x \in C$. The continuity of f gives $f(x_{n_i}) \to f(x)$. This contradicts our assumption that $|f(x_{n_i})| \geq n_i$ — after all, $|f(x)|$ is finite.

Proving that $f(C)$ is closed is almost the same idea: suppose that $f(x_n) \to y$. Then again (x_n) must have a convergent subsequence (x_{n_i}) with $x_{n_i} \to x \in C$. Therefore $f(x_{n_i}) \to f(x)$ and $y = f(x)$, showing that $f(C)$ is closed. ☐

One of the useful consequences of this result is the following.

COROLLARY B.7.2

Let $f : C \to \mathbb{R}$ be a continuous function, where $C \subseteq \mathbb{R}^n$ is a compact set. Then $f(C)$ is bounded and there exists $x, y \in C$ with

$$\begin{aligned} f(x) &= \inf\{f(x) \mid x \in C\} \\ f(y) &= \sup\{f(x) \mid x \in C\}. \end{aligned}$$

In more boiled down terms, this corollary says that a real continuous function on a compact set assumes its minimum and its maximum. As an example let

$$f(x, y) = x^{18}y^{113} + 3x\cos(x) + e^x \sin(y).$$

A special case of the corollary is that there exists points (x_0, y_0) and (x_1, y_1) in B, where

$$B = \{(x, y) \mid x^2 + y^2 \leq 117\}$$

such that

$$\begin{aligned} f(x_0, y_0) &\leq f(x, y) \\ f(x_1, y_1) &\geq f(x, y) \end{aligned} \tag{B.7}$$

for every $(x, y) \in B$. You may say that this is clear arguing that if (x_0, y_0) does not satisfy (B.7), there must exist $(x, y) \in B$ with $f(x, y) < f(x_0, y_0)$. Then put $(x_0, y_0) := (x, y)$ and keep going until (B.7) is satisfied. This argument is intuitive. I guess that all we have done is to write it down precisely in the language coming from centuries of mathematical distillation.

B.8 Exercises

(1) Use induction to prove the formula in (B.2).

(2) (i) Show that
$$2ab \leq a^2 + b^2$$
for $a, b \in \mathbb{R}$.

(ii) Let $x, y \in \mathbb{R}^n \setminus \{0\}$, where $x = (x_1, \dots, x_n)^t$ and $y = (y_1, \dots, y_n)^t$. Prove that
$$2\,\frac{x_i}{|x|}\,\frac{y_i}{|y|} \leq \frac{x_i^2}{|x|^2} + \frac{y_i^2}{|y|^2}$$
for $i = 1, \dots, n$.

(iii) Deduce the Cauchy-Schwarz inequality from (2ii).

(3) Show formally that $1, 2, 3, \dots$ does not have a convergent subsequence. Can you have a convergent subsequence of a non-convergent sequence?

(4) Prove Proposition B.2.6.

(5) Let S be a subset of the rational numbers $\mathbb{Q}$, which is bounded from above. Of course this subset always has a supremum in $\mathbb{R}$. Can you give an example of such an S, where $\sup(S) \notin \mathbb{Q}$.

(6) Let $S = \mathbb{R} \setminus \{0, 1\}$. Prove that S is not closed. What is $\overline{S}$?

(7) Let $S_1 = \{x \in \mathbb{R} \mid 0 \leq x \leq 1\}$. What is S_1° and ∂S_1?

Let $S_2 = \{(x, y) \in \mathbb{R}^2 \mid 0 \leq x \leq 1, y = 0\}$. What is S_2° and ∂S_2?

(8) Let $S \subseteq \mathbb{R}^n$. Show that $S^\circ \subseteq S$ and $S \cup \partial S = \overline{S}$. Is ∂S contained in S?

Let $U = \mathbb{R}^n \setminus F$, where $F \subseteq \mathbb{R}^n$ is a closed set. Show that $U^\circ = U$ and $\partial U \cap U = \emptyset$.

(9) Show that
$$||x| - |y|| \leq |x - y|$$
for every $x, y \in \mathbb{R}^n$.

(10) Give an example of a subset $S \subseteq \mathbb{R}$ and a continuous function $f : S \to \mathbb{R}$, such that $f(S)$ is not bounded.

Appendix C

Polyhedra in standard form

A polyhedron defined by $P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ is said to be in *standard form*. Here A is an $m \times n$-matrix and b an m-vector. Notice that

$$P = \left\{ x \in \mathbb{R}^n \middle| \begin{array}{rcr} Ax & \leq & b \\ -Ax & \leq & -b \\ -x & \leq & 0 \end{array} \right\}. \tag{C.1}$$

A polyhedron P in standard form does not contain a line (why?) and therefore always has an extremal point if it is non-empty. Also

$$\operatorname{ccone}(P) = \{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}.$$

C.1 Extremal points

The determination of extremal points of polyhedra in standard form is a direct (though somewhat laborious) translation of Proposition 4.6.1.

THEOREM C.1.1

Let A be an $m \times n$ matrix af rank m. There is a one to one correspondence between extremal points in

$$P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

and m linearly independent columns B in A with $B^{-1}b \geq 0$. The extremal point corresponding to B is the vector with zero entries except at the coordinates corresponding the columns of B. Here the entries are $B^{-1}b$.

Proof. Write P as in (C.1):

$$\left\{ x \in \mathbb{R}^n \middle| \quad A'x \leq b \quad \right\},$$

where

$$A' = \begin{pmatrix} A \\ -A \\ -I \end{pmatrix},$$

is an $(2m+n) \times n$ matrix with I the $n \times n$-identity matrix. For any $z \in P$, A'_z always contains the first $2m$ rows giving us rank m by assumption. If $z \in P$ is an extremal point then A'_z has rank n. So an extremal point corresponds to adding $n-m$ of the rows of $-I$ in order to obtain rank n. Adding these $n-m$ rows of the $n \times n$ identity matrix amounts to setting the corresponding variables $= 0$. The m remaining (linearly independent!) columns of A give the desired $m \times m$ matrix B with $B^{-1}b \geq 0$. □

We will illustrate this principle in the following example.

EXAMPLE C.1.2
Suppose that

$$P = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}, x, y, z \geq 0 \right\}.$$

According to Theorem C.1.1, the possible extremal points in P are given by

$$\begin{pmatrix} 1 & 2 \\ 4 & 5 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \end{pmatrix}, \quad \begin{pmatrix} 1 & 3 \\ 4 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 1/6 \end{pmatrix},$$
$$\begin{pmatrix} 2 & 3 \\ 5 & 6 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ -1/3 \end{pmatrix}.$$

From this you see that P only has the two extremal points

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 1/3 \\ 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} x_2 \\ y_2 \\ z_2 \end{pmatrix} = \begin{pmatrix} 1/2 \\ 0 \\ 1/6 \end{pmatrix}.$$

If you consider the polyhedron

$$P = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \middle| \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, x, y, z \geq 0 \right\},$$

then the possible extremal points are given by

$$\begin{pmatrix}1 & 2\\4 & 5\end{pmatrix}^{-1}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}-1\\1\end{pmatrix},\quad \begin{pmatrix}1 & 3\\4 & 6\end{pmatrix}^{-1}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}-1/2\\1/2\end{pmatrix},$$
$$\begin{pmatrix}2 & 3\\5 & 6\end{pmatrix}^{-1}\begin{pmatrix}1\\1\end{pmatrix}=\begin{pmatrix}-1\\1\end{pmatrix}.$$

What does this imply about Q?

C.2 Extremal directions

The next step is to compute the extremal infinite directions for a polyhedron in standard form.

THEOREM C.2.1
Let A be an $m \times n$ matrix of rank m. Then the extremal infinite directions for

$$P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$$

are in correspondence with (B, a_j), where B is a subset of m linearly independent columns in A, a_j a column not in B with $B^{-1}a_j \leq 0$. The extremal direction corresponding to (B, a_j) is the vector with entry 1 on the j-th coordinate, $-B^{-1}a_j$ on the coordinates corresponding to B and zero elsewhere.

Proof. Again let

$$P = \left\{x \in \mathbb{R}^n \middle| \quad A'x \leq b \quad \right\},$$

where

$$A' = \begin{pmatrix} A \\ -A \\ -I \end{pmatrix},$$

is an $(2m+n) \times n$ matrix with I the $n \times n$-identity matrix. The extremal directions in P are the extremal rays in the characteristic cone

$$\operatorname{ccone}(P) = \left\{x \in \mathbb{R}^n \middle| \quad A'x \leq 0 \quad \right\},$$

Extremal rays correspond to $z \in \operatorname{ccone}(P)$ where the rank of A'_z is $n-1$. As before the first $2m$ rows of A' are always in A'_z and add up to a matrix if rank m. So we must add an extra $n-m-1$ rows of $-I$. This corresponds to picking out $m+1$ columns B' of A such that the matrix B' has rank m. Therefore $B' = (B, a_j)$, where B is a matrix of m linearly independent columns and $a_j \notin B$. This pair defines an extremal ray if and only if $B'v = 0$ for a non-zero $v \geq 0$. This is equivalent with $B^{-1}a_j \leq 0$. $\square$

EXAMPLE C.2.2

The following example comes from the 2004 exam. Consider the polyhedron

$$P = \left\{ \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbb{R}^3 \;\middle|\; \begin{array}{ccccccc} x & + & 2y & - & 3z & = & 1 \\ 3x & + & y & - & 2z & = & 1 \\ x & & & & & \geq & 0 \\ & & y & & & \geq & 0 \\ & & & & z & \geq & 0 \end{array} \right\}$$

in standard form. Compute its infinite extremal directions and extremal points.

The relevant matrix above is

$$A = \begin{pmatrix} 1 & 2 & -3 \\ 3 & 1 & -2 \end{pmatrix}$$

The procedure according to Theorem C.2.1 is to pick out invertible 2 × 2 submatrices B of A and check if $B^{-1}a \leq 0$ with a the remaining column in A. Here are the computations:

$$\begin{pmatrix} 1 & 2 \\ 3 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -3 \\ -2 \end{pmatrix} = \begin{pmatrix} -\frac{1}{5} \\ -\frac{7}{5} \end{pmatrix}$$

$$\begin{pmatrix} 1 & -3 \\ 3 & -2 \end{pmatrix}^{-1} \begin{pmatrix} 2 \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{7} \\ -\frac{5}{7} \end{pmatrix}$$

$$\begin{pmatrix} 2 & -3 \\ 1 & -2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 3 \end{pmatrix} = \begin{pmatrix} -7 \\ -5 \end{pmatrix}.$$

Therefore the extremal rays are

$$\begin{pmatrix} \frac{1}{5} \\ \frac{7}{5} \\ 1 \end{pmatrix}, \quad \begin{pmatrix} \frac{1}{7} \\ 1 \\ \frac{5}{7} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 \\ 7 \\ 5 \end{pmatrix}.$$

Using Theorem C.1.1 you can check that the only extremal point of P is

$$\begin{pmatrix} \frac{1}{5} \\ \frac{2}{5} \\ 0 \end{pmatrix}.$$

This shows that

$$P = \left\{ \begin{pmatrix} \frac{1}{5} \\ \frac{2}{5} \\ 0 \end{pmatrix} + \lambda_1 \begin{pmatrix} \frac{1}{5} \\ \frac{7}{5} \\ 1 \end{pmatrix} + \lambda_2 \begin{pmatrix} \frac{1}{7} \\ 1 \\ \frac{5}{7} \end{pmatrix} + \lambda_3 \begin{pmatrix} 1 \\ 7 \\ 5 \end{pmatrix} \;\middle|\; \lambda_1, \lambda_2, \lambda_3 \geq 0 \right\}.$$

C.3 Exercises

1. Let P be a non-empty polyhedron in standard form. Prove

    (a) P does not contain a line.

    (b) P has an extremal point.

    (c)
    $$\operatorname{ccone}(P) = \{x \in \mathbb{R}^n \mid Ax = 0, x \geq 0\}.$$

2. Let

$$P = \left\{ \begin{pmatrix} x \\ y \\ z \\ u \\ v \end{pmatrix} \in \mathbb{R}^5 \;\middle|\; \begin{array}{rrrrrcl} x & -y & -2z & -u & & = & 1 \\ -x & +2y & +4z & -2u & +v & = & 1 \\ x & & & & & \geq & 0 \\ & y & & & & \geq & 0 \\ & & z & & & \geq & 0 \\ & & & u & & \geq & 0 \\ & & & & v & \geq & 0 \end{array} \right\}$$

Compute the extremal points and directions of P. Write P as the sum of a convex hull and a finitely generated convex cone.

Bibliography

[1] Julius Farkas, *Theorie der einfachen Ungleichungen*, J. Reine Angew. Math. **124** (1901), 1–27.

[2] Joseph Fourier, *Solution d'une question perticulière du calcul des inégalités*, Nouveau Bulletin des Sciences par la Société philomatique de Paris (1826), 317–319.

[3] Komei Fukuda and Alain Prodon, *Double description method revisited*, Combinatorics and computer science (Brest, 1995), Lecture Notes in Comput. Sci., vol. 1120, Springer, Berlin, 1996, pp. 91–111.

[4] Paul Gordan, *Ueber die Auflösung linearer Gleichungen mit reellen Coefficienten*, Math. Ann. **6** (1873), no. 1, 23–28.

[5] Hermann Minkowski, *Allgemeine Lehrsätze über die konvexen Polyeder*, Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen, mathematisch-physikalische Klasse (1897), 198–219.

[6] T. S. Motzkin, *Two consequences of the transposition theorem on linear inequalities*, Econometrica **19** (1951), 184–185.

[7] T. S. Motzkin, H. Raiffa, G. L. Thompson, and R. M. Thrall, *The double description method*, Contributions to the theory of games, vol. 2, Annals of Mathematics Studies, no. 28, Princeton University Press, Princeton, N. J., 1953, pp. 51–73.

[8] Hermann Weyl, *Elementare Theorie der konvexen Polyeder*, Commentarii Mathematici Helvetici **7** (1935), 290–306.

Index