

A mini-introduction to convexity

Geir Dahl*

January 20, 2014

1 Introduction

Convexity, or convex analysis, is an area of mathematics where one studies questions related to two basic objects, namely *convex sets* and *convex functions*. Triangles, rectangles and “certain” polygons are examples of convex sets in the plane, and the quadratic function $f(x) = ax^2 + bx + c$ is convex provided that $a \geq 0$. Actually, the points in the plane on or above the graph of this quadratic function is another example of a convex set. But one may also consider convex sets in \mathbb{R}^n for any n , and convex functions of several variables.

Convexity is the mathematical core of optimization, and it plays an important role in many other mathematical areas as statistics, approximation theory, differential equations and mathematical economics.

This short note is meant for a short (probably too short) introduction to some concepts and results in convexity. The focus is on convexity in connection with linear optimization. *This notes are meant for two (or three) lectures in the course MAT-INF3100 Linear Optimization where the main project is to study linear programming, but where some knowledge to convexity is useful.* Due to this limited scope of these notes we do not discuss convex functions, except for a few remarks a couple of places.

Example 1. (*Optimization and convex functions*) A basic optimization problem is to minimize a real-valued function f of n variables, say $f(x)$ where $x = (x_1, \dots, x_n) \in A$ and A is the domain of f . (Such problems arise in all sorts of applications: economics, statistics (estimation, regression, curve fitting), approximation problems, scheduling and planning problems, image

*University of Oslo, Dept. of Mathematics (geird@math.uio.no)

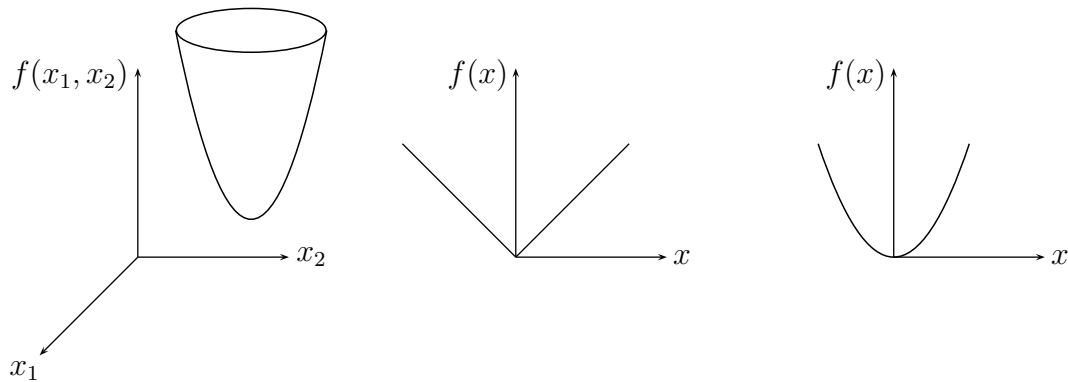


Figure 1: Some convex functions

analysis, medical imaging, engineering applications etc.). A *global minimum* of f is a point x^* with

$$f(x^*) \leq f(x) \quad \text{for all } x \in A$$

where A is the domain of f . Often it is hard to find a global minimum so one settles with a *local minimum* point which satisfies $f(x^*) \leq f(x)$ for all x in A that are sufficiently close to x^* . There are several optimization algorithms that are able to locate a local minimum of f .

Unfortunately, the function value in a local minimum may be much larger than the global minimum value. This raises the question: Are there functions where a local minimum point is also a global minimum? The main answer to this question is:

If f is a convex function, and the domain A is a convex set, then a local minimum point is also a global minimum point !

Thus, one can find the global minimum of convex functions whereas this may be hard (or even impossible) in other situations. Some convex functions are illustrated in Figure 1.

In linear optimization (= linear programming) we minimize a linear function f subject to linear constraints; more precisely these constraints are linear inequalities and linear equations. The feasible set in this case (the set of points satisfying the constraints) is always a convex set, in fact it is a special convex set called a *polyhedron*.

Example 2. (*Convex set*) Loosely speaking a convex set in \mathbb{R}^2 (or \mathbb{R}^n) is a set “with no holes”. More accurately, a convex set C has the following property: whenever we choose two points in the set, say $x, y \in C$, then all

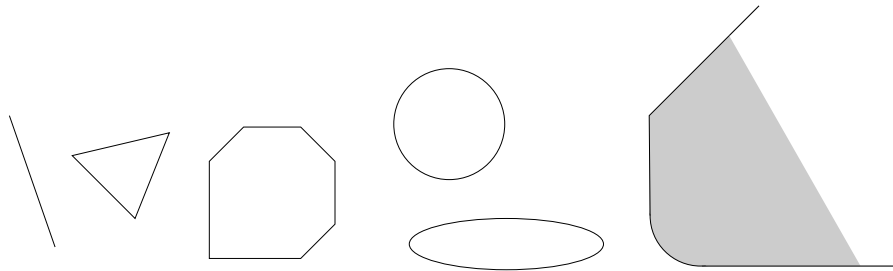


Figure 2: Some convex sets in the plane.

points on the line segment between x and y also lie in C . Some examples of convex sets in the plane are: a sphere (ball), an ellipsoid, a point, a line, a line segment, a rectangle, a triangle, see Fig. 2. But, for instance, a set with a finite number p of points is only convex when $p = 1$. The union of two disjoint (closed) triangles is also nonconvex.

Example 3. (*Approximation*) A basic approximation problem, with several applications, may be presented in the following way: given some closed set $S \subset \mathbb{R}^n$ and a vector $a \notin S$, find a *nearest point* defined as a point (vector) $x \in S$ which is as close to a as possible among elements in S . Let us measure the distance between vectors using the Euclidean norm, so $\|x - y\| = (\sum_{j=1}^n (x_j - y_j)^2)^{1/2}$ for $x, y \in \mathbb{R}^n$. One can show that there is always at least one nearest point provided that S is nonempty and closed (contains its boundary). *Now, if S is a convex set, then there is a unique nearest point.*

2 The definitions

You will now see three basic definitions:

1. A set $C \subseteq \mathbb{R}^n$ is called *convex* if

$$(1 - \lambda)x_1 + \lambda x_2 \in C \text{ whenever } x_1, x_2 \in C \text{ and } 0 \leq \lambda \leq 1.$$

Geometrically, this means that C contains the line segment between each pair of points in C .

2. Let $C \subseteq \mathbb{R}^n$ be a convex set and consider a real-valued function f defined on C . The function f is called *convex* if the inequality

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

holds for every $x, y \in C$ and every $0 \leq \lambda \leq 1$.

3. A *polyhedron* $P \subseteq \mathbb{R}^n$ is defined as the solution set of a system of linear inequalities. Thus, P has the form

$$P = \{x \in \mathbb{R}^n : Ax \leq b\} \quad (2)$$

where A is a real $m \times n$ matrix, $b \in \mathbb{R}^m$ and where the vector inequality means it holds for every component.

Here are some important comments to these definitions:

- In part 1 of the definition above the point $(1 - \lambda)x_1 + \lambda x_2$ is called a *convex combination* of x_1 and x_2 . So the definition of a convex set says that it is closed under taking convex combination of pairs of points. Actually, one can show that when C is convex it also contains every convex combination of any (finite) set of its points. A convex combination of points x_1, x_2, \dots, x_m is a linear combination

$$\sum_{j=1}^m \lambda_j x_j$$

where the coefficients $\lambda_1, \lambda_2, \dots, \lambda_m$ are nonnegative and sum to 1.

- In the definition of a convex function we actually use that the domain C is a convex set: this assures that the point $(1 - \lambda)x + \lambda y$ lies in C so the defining inequality for f makes sense.
- In the definition of a polyhedron we consider systems of *linear inequalities*. Since a linear equation $a^T x = \alpha$ may be written as two linear inequalities, namely $a^T x \leq \alpha$ and $-a^T x \leq -\alpha$, one may also say that a polyhedron is the solution set of a system of linear equations and inequalities.

Proposition 1. *Every polyhedron is a convex set.*

Proof. Consider a polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ and let $x_1, x_2 \in P$ and $0 \leq \lambda \leq 1$. Then

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

which shows that $(1 - \lambda)x_1 + \lambda x_2 \in P$ and the convexity of P follows. \square

3 Linear optimization and convexity

Recall that a *linear programming* problem may be written as

$$\begin{aligned} & \text{maximize} && c_1x_1 & + & \dots & + c_nx_n \\ & \text{subject to} && a_{11}x_1 & + & \dots & + a_{1n}x_n & \leq b_1; \\ & && & & \vdots & & \\ & && a_{m1}x_1 & + & \dots & + a_{mn}x_n & \leq b_m; \\ & && x_1, \dots, x_n & \geq 0. \end{aligned} \tag{3}$$

or more compactly in matrix form

$$\begin{aligned} & \text{maximize} && c^T x \\ & \text{subject to} && Ax \leq b \\ & && x \geq O. \end{aligned} \tag{4}$$

Here $A = [a_{i,j}]$ is the $m \times n$ coefficient matrix with (i, j) th element being $a_{i,j}$, $b \in \mathbb{R}^m$ (column vector) and O denotes a zero vector (here of dimension n). Again vector inequalities should be interpreted componentwise. Thus, the LP feasible set

$$\{x \in \mathbb{R}^n : Ax \leq b, x \geq O\}$$

is a polyhedron and therefore a convex set. (Actually, LP may be defined as minimizing or maximizing a linear function over a polyhedron). As a consequence we have that if x_1 and x_2 are two feasible points, then every convex combination of these points is also feasible. But what can be said about the set of optimal solutions?

Proposition 2. *In an LP problem with finite optimal value the set P^* of optimal solutions is a convex set, actually P^* is a polyhedron.*

Proof. Let v^* denote the optimal value. Then

$$P^* = \{x \in \mathbb{R}^n : Ax \leq b, x \geq O, c^T x = v^*\}$$

which is a polyhedron. □

So, if you have different optimal solutions of an LP problem every convex combination of these will also be optimal.

An attempt to illustrate the geometry of linear programming is given in Fig. 3 (where the feasible region is the solution set of five linear inequalities).

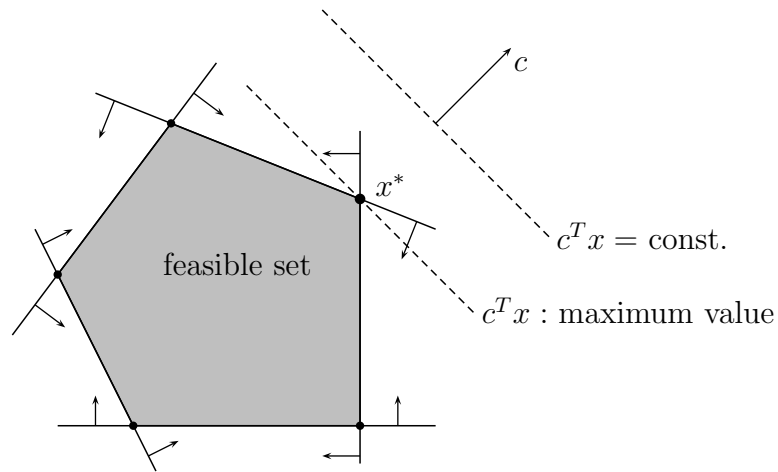


Figure 3: Linear programming

4 The convex hull

Given a (possibly nonconvex) set S it is natural to ask for the smallest convex set containing S . This question is what we consider in this section.

Let $S \subseteq \mathbb{R}^n$ be any set. Define the *convex hull* of S , denoted by $\text{conv}(S)$ as the set of all convex combinations of points in S (see Fig. 4). The convex hull of two points x_1 and x_2 is the line segment between the two points. An important fact is that $\text{conv}(S)$ is a convex set, whatever the set S might be. Thus, taking the convex hull becomes a way of producing new convex sets.

The following proposition tells us that the convex hull of a set S is the smallest convex set containing S . Recall that the intersection of an arbitrary family of sets consists of the points that lie in all of these sets.

Proposition 3. *Let $S \subseteq \mathbb{R}^n$. Then $\text{conv}(S)$ is equal to the intersection of all convex sets containing S . Thus, $\text{conv}(S)$ is the smallest convex set containing S .*

Proof. It is an exercise to show that $\text{conv}(S)$ is convex. Moreover, $S \subseteq \text{conv}(S)$; just look at a convex combination of one point! Therefore $W \subseteq \text{conv}(S)$ where W is defined as the intersection of all convex sets containing S . Now, consider a convex set C containing S . Then C must contain all convex combinations of points in S . But then $\text{conv}(S) \subseteq C$ and we conclude that W (the intersection of such sets C) must contain $\text{conv}(S)$. This completes the proof. \square

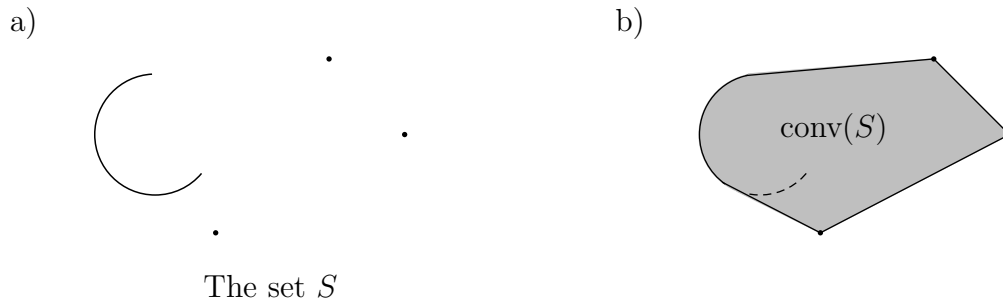
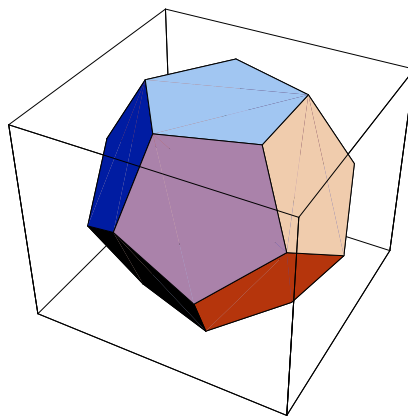


Figure 4: Convex hull

Note that if S is convex, then $\text{conv}(S) = S$. The proof is left as an exercise.

We have seen that by taking the convex hull we produce a convex set whatever set we might start with. If we start with a finite set, a very interesting class of convex sets arise. A set $P \subset \mathbb{R}^n$ is called a *polytope* if it is the convex hull of a finite set of points in \mathbb{R}^n . In Figure 3 the shaded area is a polytope (of dimension 2). An example of a three-dimensional polytope is the *dodecahedron* shown next



The convex set in Fig. 4 b) is not a polytope.

Every polytope is clearly bounded, i.e., it lies in some, suitably large, ball. A central result in the theory of polytopes is that: *a set is a polytope if and only if it is a bounded polyhedron*. We return to this in Section 6.

Polytopes have been studied a lot during the history of mathematics. Today polytope theory is still a fascinating subject with a lot of activity. One of the reasons is its relation to linear programming, because in LP problems the feasible set P is a polyhedron. If, moreover, P is bounded, then it is actually a polytope (see above). This means that there is a finite set S of points which “span” P in the sense that P consists of all convex combinations of the points in S .

Example 4. (*LP and polytopes*) Consider a polytope

$$P = \text{conv}(\{x_1, x_2, \dots, x_t\}).$$

We want to solve the optimization problem

$$\max\{c^T x : x \in P\} \tag{5}$$

where $c \in \mathbb{R}^n$. As mentioned above, this problem is an LP problem, but we do not worry too much about this now. The interesting thing is the combination of a linear objective function and the fact that the feasible set is a convex hull of finitely many points. To see this, consider an arbitrary feasible point $x \in P$. Then x may be written as a convex combination of the points x_1, x_2, \dots, x_t , say $x = \sum_{j=1}^t \lambda_j x_j$ for some $\lambda_j \geq 0$, $j = 1, \dots, t$ where $\sum_j \lambda_j = 1$. Define now $v^* = \max_j c^T x_j$. We then calculate

$$c^T x = c^T \sum_j \lambda_j x_j = \sum_{j=1}^t \lambda_j c^T x_j \leq \sum_{j=1}^t \lambda_j v^* = v^* \sum_{j=1}^t \lambda_j = v^*.$$

Thus, v^* is an upper bound for the optimal value in the optimization problem (5). We also see that this bound is attained whenever λ_j is positive only for those indices j satisfying $c^T x_j = v^*$. Let J be the set of such indices. We conclude that the optimal solutions of the problem (5) is the set

$$\text{conv}(\{x_j : j \in J\})$$

which is another polytope (contained in P). The procedure just described may be useful computationally if the number t of points defining P is not too large. In some cases, t is too large, and then we may still be able to solve the problem (5) by different methods, typically linear programming related methods.

5 Consistency and Farkas' lemma

Farkas' lemma is a theoretical results concerned with the consistency of a linear system of inequalities. It gives sufficient and necessary conditions for this system to be consistent, i.e., have at least one solution.

Here is Farkas' lemma and we prove it using LP duality theory. Other proofs also exists, for instance using separation theorems in convexity.

Lemma 4. *Let $A \in \mathbb{R}^{m,n}$ and $b \in \mathbb{R}^m$. Then the linear system $Ax \leq b$ has at least one solution x if and only if*

$$y^T b \geq 0 \text{ for every } y \in \mathbb{R}^m \text{ satisfying } y^T A = 0 \text{ and } y \geq 0.$$

Proof. Consider the pair of dual LP problems

$$(P) \max c^T x \text{ subject to } Ax \leq b$$

and

$$(D) \min y^T b \text{ subject to } y^T A = c^T, y \geq 0.$$

Let now $c = 0$. Then (D) is feasible ($y = 0$ is feasible) so by LP theory (D) either has an optimal solution or it is unbounded. If (D) is unbounded weak LP duality implies that (P) cannot have any feasible solution, i.e., $Ax \leq b$ is not consistent. On the other hand, if (D) has an optimal solution, then $y^T b \geq 0$ for every $y \in \mathbb{R}^m$ satisfying $y^T A = 0$, $y \geq 0$. (For: if $y^T b < 0$ for some y with $y^T A = 0$ and $y \geq 0$, then (D) would be unbounded; just consider $y' = \lambda y$ and increase λ). This proves the result. \square

The geometrical content of Farkas' lemma may be explained using a different version of this lemma: $Ax = b$ has a nonnegative solution if and only if $y^T b \geq 0$ for all y with $y^T A \geq 0$. Consider the convex cone C generated by the columns a^1, a^2, \dots, a^n of the matrix A ; this is the set of all linear combinations of the columns using *nonnegative* coefficients only. Then $b \in C$ means that there is a nonnegative solution x to the linear system $Ax = b$. Now, if $b \notin C$ Farkas' lemma says that there is a vector y with $y^T a^j < 0$ for $j \leq n$ and $y^T b > 0$ which geometrically means that there is a hyperplane $\{x \in \mathbb{R}^n : y^T x = 0\}$ that separates b from a^1, a^2, \dots, a^n . See the illustration in Fig. 5.

6 Main theorem for polyhedra

We now present one of the most important theorems concerning polyhedra and polytopes. Actually, the theorem explains how these two objects are related. The result in its general form was proved in 1936 by T.S. Motzkin, and earlier more specialized versions are due to G. Farkas, H. Minkowski and H. Weyl.

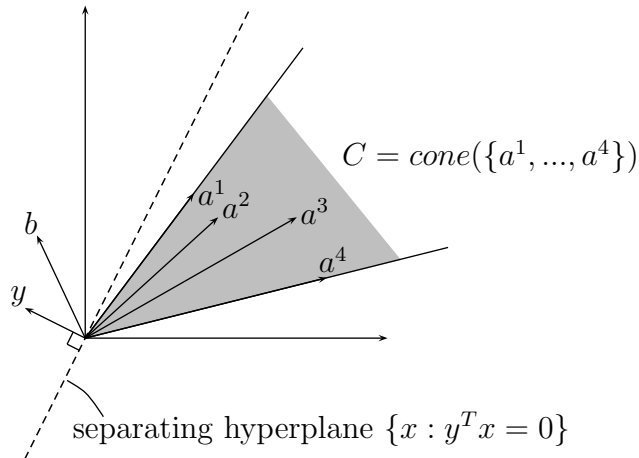


Figure 5: Geometry of Farkas' lemma

Theorem 5. *Let $P \subseteq \mathbb{R}^n$ be a nonempty polyhedron. Then there are vectors $v_1, v_2, \dots, v_s \in \mathbb{R}^n$ and $w_1, w_2, \dots, w_t \in \mathbb{R}^n$ such that P consists precisely of the vectors x that may be written*

$$x = \sum_{i=1}^s \lambda_i v_i + \sum_{j=1}^t \mu_j w_j \quad (6)$$

where all λ_i and μ_j are nonnegative and $\sum_{i=1}^s \lambda_i = 1$.

Conversely, let $v_1, v_2, \dots, v_s \in \mathbb{R}^n$ and $w_1, w_2, \dots, w_t \in \mathbb{R}^n$. Then the set Q of vectors x of the form (6) is a polyhedron, so there is a matrix $A \in \mathbb{R}^{m,n}$ and a vector $b \in \mathbb{R}^m$ such that

$$Q = \{x \in \mathbb{R}^n : Ax \leq b\}.$$

We omit the proof here. The theorem above is an *existence result*; it shows the existence of certain vectors and matrices, but it does not tell us how to find these. Still, this can indeed be done using, for instance, the Fourier-Motzkin method which we explain in the next section. The core of the main theorem above is that polyhedra may be represented in two ways

1. as the solution set of a linear system $Ax \leq b$: this may be considered an *exterior description* as the intersection of halfspaces.
2. an *interior description* via convex combinations and conical combinations of certain vectors.

An immediate consequence of the main theorem is the following result.

Corollary 6. *A set P is a polytope if and only if it is a bounded polyhedron.*

Theorem 5 has a strong connection to linear programming theory. Consider an LP problem

$$\max c^T x \text{ subject to } Ax \leq b$$

and the corresponding feasible polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ which we assume is nonempty. Let $v_1, v_2, \dots, v_s \in \mathbb{R}^n$ and $w_1, w_2, \dots, w_t \in \mathbb{R}^n$ be the vectors as described in the first part of Theorem 5. So every $x \in P$ may be written as

$$x = \sum_{i=1}^s \lambda_i v_i + \sum_{j=1}^t \mu_j w_j$$

for suitable nonnegative scalars where $\sum_{i=1}^s \lambda_i = 1$. There are now two possibilities:

- $c^t w_j > 0$ for some $j \leq t$. Then the LP is unbounded; we can get an arbitrary large value of the objective function by moving along the ray $\{\mu w_j : \mu \geq 0\}$.
- $c^t w_j \leq 0$ for all $j \leq t$. Then the LP has an optimal solution and, actually, one of the points v_1, v_2, \dots, v_s is optimal (as we may let $\mu_j = 0$ for all j); recall here the argument given in Example 4.

Moreover, the points v_1, v_2, \dots, v_s in Theorem 5 may be chosen to be *extreme points* of the polyhedron P . An *extreme point* of P is defined as a point which cannot be written as a convex combination of other points in P (actually, it suffices to check if it can be written as the midpoint between *two* points in P). Thus: $x \in P$ is an extreme point if and only if

$$x = \frac{1}{2}x^1 + \frac{1}{2}x^2 \text{ for } x^1, x^2 \in P \text{ implies that } x^1 = x^2 = x.$$

Proposition 7. *Let A be an $m \times n$ matrix with rank m , and consider the polyhedron $P = \{x \in \mathbb{R}^n : Ax = b, x \geq O\}$. Let $x \in P$. Then x is a basic solution (in LP sense) if and only if x is an extreme point of P .*

Proof. If x is a basic feasible solution, then (after possibly reordering variables; see LP lectures) we may write $x = \begin{bmatrix} x_B \\ x_N \end{bmatrix}$ where $x_B = A_B^{-1}b$, $x_N = O$ and $A = \begin{bmatrix} A_B & A_N \end{bmatrix}$; here the $m \times m$ submatrix A_B is invertible (nonsingular). (This is possible as A has full rowrank, so there must exist m linearly independent columns in A .) Assume that $x = (1/2)x^1 + (1/2)x^2$ for some $x^1, x^2 \in P$. This immediately implies that $x_N^1 = x_N^2 = O$ (as for each $j \in N$ $x_j = 0$ and $x_j^1, x_j^2 \geq 0$). Moreover, $Ax^1 = b$ so $A_Bx_B^1 + A_Nx_N^1 = b$ and therefore $x_B^1 = A_B^{-1}b = x_B$ (as $x_N^1 = O$). Similarly, $x_B^2 = A_B^{-1}b = x_B$, so $x^1 = x^2 = x$, and it follows that x is an extreme point.

Conversely, assume that x is an extreme point of P and consider the indices corresponding to positive components: $J = \{j \leq n : x_j > 0\}$.

Claim: the columns in A corresponding to J are linearly independent.

Proof of Claim: If J is empty, there is nothing to prove here. Otherwise, let a_j denote the j th column in A . From $Ax = b$ we obtain $\sum_{j \in J} x_j a_j = b$. Assume $\sum_{j \in J} \lambda_j a_j = O$. Multiply this equation by a number ϵ and add to the previous vector equation; this gives $\sum_{j \in J} (x_j + \epsilon \lambda_j) a_j = b$. In this equation each x_j is positive, so we can find a suitably small $\epsilon_0 > 0$ (small compared to the λ_j 's), so that $x_j + \epsilon \lambda_j > 0$ for all $\epsilon \in [-\epsilon_0, \epsilon_0]$. Define x^1 and x^2 by $x_j^1 = x_j + \epsilon_0 \lambda_j$ ($j \in J$), $x_j^1 = 0$ ($j \notin J$), and $x_j^2 = x_j - \epsilon_0 \lambda_j$ ($j \in J$), $x_j^2 = 0$ ($j \notin J$). Then $x^1, x^2 \in P$ (as $x^1 \geq O$ and $Ax^1 = Ax + \epsilon_0 \sum_{j \in J} \lambda_j a_j = Ax = b$; similar for x^2). Moreover, $x = (1/2)x^1 + (1/2)x^2$ and since x is an extreme point, we get $x^1 = x^2$ and therefore $\lambda_j = 0$ for each $j \in J$. This proves that the vectors a_j , $j \in J$ are linearly independent.

Finally, we may extend the columns a_j , $j \in J$ into a basis for \mathbb{R}^m by adding $m - |J|$ other columns from A ; this is possible since the columnrank of A is m (the “extension theorem” in basic linear algebra). So these m columns form a basis A_B in A (a nonsingular submatrix) and since $Ax = b$ (because $x \in P$) and $x_N = O$ we get $A_Bx_B = b$, so $x_B = A_B^{-1}b$ and x is a basic solution. \square

Thus, the two possibilities discussed in the bullets on the previous page correspond to what is known as the *fundamental theorem of linear programming*.

7 Finding extreme points

An interesting, and sometimes important, task is to determine all (or some) of the extreme points of a given polyhedron P . From the previous section we now have two techniques for finding the extreme points:

1. *Use the definition of extreme point.* Look at a general point x in the polyhedron P . Try to write it as the midpoint between two other points in P ; if so, x is not an extreme point. Modify x suitably so that “it is harder to write it as a midpoint”; this will happen when you force more inequalities to be active at your point. Eventually, if you succeed, you will find all extreme points in this way. Note that this technique is a *combinatorial* discussion/argument which could be very hard to perform in practice (sometimes impossible!). But the idea is to exploit the structure of the given inequalities/equations somehow. This technique may be used for a general polyhedron, e.g., in the form $P = \{x \in \mathbb{R}^n : Ax \leq b\}$.
2. *Use Proposition 7.* This method may be applied when the polyhedron has the form $P = \{x \in \mathbb{R}^n : Ax = b, x \geq O\}$ where the $m \times n$ matrix A has rank m . (Remark: there is a similar technique for $Ax \leq b$, but we do not go into this here; enough is enough!) The technique is “simply” to determine *all* bases of A . So, again there will be a combinatorial discussion. In principle you may consider all choices of m columns selected from the n columns (i.e., choosing a subset B of size m from $\{1, 2, \dots, n\}$), and for each choice you decide if the corresponding submatrix A_B is invertible by solving $A_B x_B = b$ to see if the solution is unique. If so, and if, in addition, $x_B \geq O$, then you have found an extreme point $x = (x_B, x_N) = (A_B^{-1}b, O)$.

But a warning here: finding all vertices can be a very difficult mathematical challenge. Let us look at some examples where these methods do the job.

Example 5. (i) Let a_1, a_2, \dots, a_n and b be given positive numbers and consider the polyhedron

$$P = \{x \in \mathbb{R}^n : a_1x_1 + a_2x_2 + \dots + a_nx_n = b, x \geq O\}.$$

Then P has the form so we can apply Technique 2 above. The matrix A is the $1 \times n$ matrix

$$A = [a_1 \quad a_2 \quad \dots \quad a_n]$$

which has rank $m = 1$. So a basis in A is simply an entry in A , a 1×1 submatrix, and it is invertible as each entry is nonzero. So if $B = \{j\}$, we solve $A_B x_B = b$ and get $x_B = b/a_j$ which is positive. The corresponding extreme point is

$$(0, \dots, 0, b/a_j, 0, \dots, 0)$$

where the nonzero is in position j . Thus, P has n extreme points (and they are the intersections between the coordinate axes and the hyperplane given by $a^T x = b$).

(ii) Let

$$P = \{x \in \mathbb{R}^4 : 2x_1 + 3x_2 = 12, x_1 + x_2 + x_3 + x_4 = 1, x \geq O\}$$

Again we may apply Technique 2 for

$$A = \begin{bmatrix} 2 & 3 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix}$$

which has rank 2. The following selection of index set B correspond to the bases in A :

$$\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}.$$

Here $\{3, 4\}$ does not give a basis, as the first row is the zero vector (so the submatrix is singular). Note that some of the index sets may lead to the same basis (for example, $\{2, 3\}$ and $\{2, 4\}$ do so), but these may give different points x (as the B -part are in different components). Then, for each basis, we may compute the corresponding basic solution, and the *feasible* (i.e., nonnegative) basic solutions are then the extreme points. We leave this computation to the reader!

(iii) Let

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

where $a_j > 0$ ($j \leq n$), $b > 0$ and e is the all ones vector. Let us find all extreme points using Technique 1 above.

Consider first an x satisfying $\sum_{j=1}^n a_j x_j < b$ and $0 < x_j < 1$ for some $j \leq n$. But then $x = (1/2)x' + (1/2)x''$ where $x' = x + \epsilon e_j$, $x'' = x - \epsilon e_j$ and $x', x'' \in P$ for suitable small $\epsilon > 0$ (and where e_j is the j th unit vector). Thus we see that if x is an extreme point satisfying $\sum_{j=1}^n a_j x_j < b$, then $x_j \in \{0, 1\}$ for each j ! This gives candidate extreme points that are those $(0, 1)$ -vectors that satisfy $\sum_{j=1}^n a_j x_j < b$.

Next, if an extreme point satisfies $\sum_{j=1}^n a_j x_j = b$, then a slight extension of the same argument as above, shows that $x_j \in \{0, 1\}$ for all *except at most one* j . Actually, if there were two components strictly between 0 and 1, we could increase one and decrease the other suitably, and violate the extreme point property. Moreover, if one such component x_j is strictly between 0 and 1, then x_j is determined by the equation $\sum_{j=1}^n a_j x_j = b$.

This gives all candidate extreme points, and it is not difficult to show that all these points are indeed extreme point of P . Again, we leave it to the reader to write down all the extreme points based on this discussion.

8 Fourier-Motzkin elimination and projection

Fourier-Motzkin elimination is a computational method which may be seen as a generalization of Gaussian elimination.

Fourier-Motzkin elimination is used for finding one or all solutions of a linear system of inequalities, say $Ax \leq b$ where $A \in \mathbb{R}^{m,n}$ and $b \in \mathbb{R}^m$. Moreover, the same method can be used to find the projection of a polyhedron into a subspace. The idea is to eliminate one variable at the time and rewrite the system accordingly.

To explain the method we assume that we want to eliminate the variables in the order x_1, x_2, \dots, x_n (although any order will do). The system $Ax \leq b$ may be split into three subsystems

$$\begin{aligned} a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n &\leq b_i & (i \in I^+) \\ 0 \cdot x_1 + a_{i2}x_2 + \dots + a_{in}x_n &\leq b_i & (i \in I^0) \\ a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n &\leq b_i & (i \in I^-) \end{aligned} \quad (7)$$

where $I^+ = \{i : a_{i1} > 0\}$, $I^0 = \{i : a_{i1} = 0\}$ and $I^- = \{i : a_{i1} < 0\}$. This system is clearly equivalent to

$$\begin{aligned} a'_{k2}x_2 + \dots + a'_{kn}x_n - b'_k &\leq x_1 \leq b'_i - a'_{i2}x_2 - \dots - a'_{in}x_n & (i \in I^+, k \in I^-) \\ a_{i2}x_2 + \dots + a_{in}x_n &\leq b_i & (i \in I^0) \end{aligned} \quad (8)$$

where $b'_i = b_i/|a_{i1}|$ and $a'_{ij} = a_{ij}/|a_{i1}|$ for each $i \in I^+ \cup I^-$.

It follows that x_1, x_2, \dots, x_n is a solution of the original system (7) if and only if x_2, x_3, \dots, x_n satisfy

$$\begin{aligned} a'_{k2}x_2 + \dots + a'_{kn}x_n - b'_k &\leq b'_i - a'_{i2}x_2 - \dots - a'_{in}x_n & (i \in I^+, k \in I^-) \\ a_{i2}x_2 + \dots + a_{in}x_n &\leq b_i & (i \in I^0) \end{aligned} \quad (9)$$

and x_1 satisfies

$$\max_{k \in I^-} (a'_{k2}x_2 + \dots + a'_{kn}x_n - b'_k) \leq x_1 \leq \min_{i \in I^+} (b'_i - a'_{i2}x_2 - \dots - a'_{in}x_n). \quad (10)$$

Note that, when values on x_2, x_3, \dots, x_n have been selected, the constraint in (10) says that x_1 lies in a certain interval (determined by x_2, x_3, \dots, x_n).

Thus, we have eliminated x_1 and may proceed to solve the system (9) which involves x_2, x_3, \dots, x_n . We here eliminate x_2 in similar way, and proceed until we obtain a linear system which only involves x_n , say $l \leq x_n \leq u$. Then a general solution to $Ax \leq b$ is obtained by choosing x_n in the interval $[l, u]$, and next choosing x_{n-1} in an interval which depends on x_n , then choose x_{n-2} in an interval depending on x_n and x_{n-1} etc.

We may summarize our findings in the following theorem.

Theorem 8. *The Fourier-Motzkin elimination method is a finite algorithm that finds a general solution to a given linear system $Ax \leq b$. If there is no solution, the method determines this fact by finding an implied and inconsistent inequality ($0 \leq -1$).*

Moreover, the method finds the projection P' of the given polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ into the space of a subset of the variables, and shows that P' is also polyhedron by finding a linear inequality description of P' .

Proof. This follows from our description above using induction. □

We note that the system we obtain after the elimination of a variable may contain many more inequalities than the original one. However, some of these new inequalities may be redundant, and this is checked for in computer implementations of the algorithm.

We conclude with a small illustration of Fourier-Motzkin elimination method.

Example 6. (*Fourier-Motzkin*) Consider the system

$$1 \leq x_1 \leq 2, 1 \leq x_2 \leq 4, x_1 - x_2 \geq -1.$$

We eliminate x_1 and get

$$\max\{1, x_2 - 1\} \leq x_1 \leq 2$$

and the new system

$$1 \leq 2, x_2 - 1 \leq 2, 1 \leq x_2 \leq 4.$$

The last system has a two redundant inequalities and is equivalent to

$$1 \leq x_2 \leq 3.$$

Thus, the general solution is:

$$1 \leq x_2 \leq 3, \max\{1, x_2 - 1\} \leq x_1 \leq 2.$$

Moreover, if P is the polyhedron consisting of the solutions to the original system, then the projection P into the x_2 -space is the interval $[1, 3]$. These facts may be checked by a suitable drawing in the plane.

9 Exercises

1. Show that the unit ball $B(O, 1) := \{x \in \mathbb{R}^n : \|x\| \leq 1\}$ is convex. Here $\|\cdot\|$ denotes the Euclidean norm. Hint: use the triangle inequality. Is the ball $B(a, r) := \{x \in \mathbb{R}^n : \|x - a\| \leq r\}$ convex for all $a \in \mathbb{R}^n$ and $r > 0$?
2. Prove that every linear subspace of \mathbb{R}^n is a convex set.
3. Is the union of two convex sets again convex?
4. Show that if $C_1, \dots, C_t \subseteq \mathbb{R}^n$ are all convex sets, then $C_1 \cap \dots \cap C_t$ is convex. In fact, a similar result for the intersection of any family of convex sets. Explain this.
5. Is the unit ball $B = \{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$ a polyhedron?
6. Consider the unit ball $B_\infty = \{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$. Here $\|x\|_\infty = \max_j |x_j|$ is the max norm of x . Show that B_∞ is a polyhedron. Illustrate when $n = 2$.
7. Consider the unit ball $B_1 = \{x \in \mathbb{R}^n : \|x\|_1 \leq 1\}$. Here $\|x\|_1 = \sum_{j=1}^n |x_j|$ is the absolute norm of x . Show that B_1 is a polyhedron. Illustrate when $n = 2$.
8. Explain how you can write the LP problem $\max \{c^T x : Ax \leq b\}$ in the form $\max \{c_1^T x_1 : A_1 x_1 = b_1, x_1 \geq 0\}$.
9. Consider the linear system $0 \leq x_i \leq 1$ for $i = 1, \dots, n$ and let P denote the solution set. Explain how to solve a linear programming problem

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

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

10. Show that $\text{conv}(S)$ is convex for all $S \subseteq \mathbb{R}^n$. (Hint: look at two convex combinations $\sum_j \lambda_j x_j$ and $\sum_j \mu_j y_j$, and note that both these points may be written as a convex combination of the same set of vectors.)
11. Give an example of two distinct sets S and T having the same convex hull. It makes sense to look for a smallest possible subset S_0 of a set S such that $S = \text{conv}(S_0)$. We study this question later.
12. Prove that if $S \subseteq T$, then $\text{conv}(S) \subseteq \text{conv}(T)$.

13. If S is convex, then $\text{conv}(S) = S$. Show this!
14. Let $S = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$, this is the unit circle in \mathbb{R}^2 . Determine $\text{conv}(S)$.
15. Let $S = \{(0, 0), (1, 0), (0, 1)\}$. Show that $\text{conv}(S) = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1\}$.
16. Let S consist of the points $(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1)$ and $(1, 1, 1)$. Show that $\text{conv}(S) = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : 0 \leq x_i \leq 1 \text{ for } i = 1, 2, 3\}$. Also determine $\text{conv}(S \setminus \{(1, 1, 1)\})$ as the solution set of a system of linear inequalities. Illustrate all these cases geometrically.
17. Use the version of Farkas' lemma given in Lemma 4 to prove the following version: $Ax = b$ has a nonnegative solution if and only if $y^T b \geq 0$ for all y with $y^T A \geq 0$.
18. Compute explicitly all the extreme points of the polyhedra discussed in Example 5.
19. Find all extreme points of the n -dimensional rectangle $P = \{x \in \mathbb{R}^n : a_i \leq x_i \leq b_i \text{ (} i \leq n)\}$ where $a_i \leq b_i \text{ (} i \leq n)$ are given real numbers.
20. Find all extreme points of the polyhedron

$$x_1 \geq 0, x_2 \geq 0, x_1 + x_2 \leq 1, x_2 \leq 1/2. \quad (11)$$

21. Find some (or even all!) extreme points of the polyhedron
- $$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_1 + x_2 + x_3 \leq 4, x_1 + x_2 \geq 1, x_3 \leq 2. \quad (12)$$
22. Use Fourier-Motzkin elimination to find all solutions to the linear system in (11). Illustrate the solution set geometrically.
23. Use Fourier-Motzkin elimination to find all solutions to the linear system in (12). Illustrate the solution set geometrically.
24. Consider Fourier-Motzkin elimination in the case $n = 2$ (two variables), and $m = 4$ (4 inequalities). Consider a general (or, if you prefer, choose a specific such linear system). Eliminate x_1 and try to interpret the inequalities in the new system geometrically.

10 Further reading

If you are interested in reading more about convexity, there are many topics to choose from, for instance

- Convex functions
- Projection onto convex sets
- Caratheodory's theorem
- Separation theorems
- Polyhedral theory
- Polytopes and graphs
- Polyhedra and combinatorial optimization
-

In the references you will find several suggested books for further reading.

Welcome to the world of convexity!

References

- [1] V. Chvatal. *Linear programming*. W.H. Freeman and Company, 1983.
- [2] G. Dahl. *An introduction to convexity*. Report 279. Dept. of Informatics, University of Oslo, 2001.
- [3] G. Dahl. Combinatorial properties of Fourier-Motzkin elimination. *Electronic Journal of Linear Algebra*, 16 (2007), 334–346.
- [4] M. Grötschel and M.W. Padberg. Polyhedral theory. In E.L. Lawler, J.K. Lenstra, A.H.G. Rinnooy Kan, and D.B. Shmoys, editors, *The traveling salesman problem*, chapter 8, pages 251–361. Wiley, 1985.
- [5] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms I*. Springer, 1993.
- [6] W.R. Pulleyblank. Polyhedral combinatorics. In Nemhauser et al., editor, *Optimization*, volume 1 of *Handbooks in Operations Research and Management Science*, chapter 5, pages 371–446. North-Holland, 1989.

- [7] R.T. Rockafellar. *Convex analysis*. Princeton, 1970.
- [8] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 44 of *Encyclopedia of Mathematics and its applications*. Cambridge University Press, Cambridge, 1993.
- [9] E. Torgersen. *Comparison of Statistical Experiments*, volume 36 of *Encyclopedia of Mathematics and its applications*. Cambridge University Press, Cambridge, 1992.
- [10] R. Webster. *Convexity*. Oxford University Press, Oxford, 1994.
- [11] G. Ziegler. *Lectures on polytopes*. Springer, 1995.