Lecture 2 - Introduction to Polytopes

Optimization and Approximation - ENS M1

Nicolas Bousquet

1 *Reminder of Linear Algebra definitions*

Let $x_1, \ldots, x_m$ be points in $\mathbb{R}^n$ and $\lambda_1, \ldots, \lambda_m$ be real numbers. Then $x = \sum_{i=1}^m \lambda_i x_i$ is said to be a:

- *Linear combination* (of $x_1, \ldots, x_m$) if the λ_i are arbitrary scalars.
- *Conic combination* if $\lambda_i \geq 0$ for every i.
- *Convex combination* if $\sum_{i=1}^m \lambda_i = 1$ and $\lambda_i \geq 0$ for every i.

In the following, λ will still denote a scalar (since we consider in real spaces, λ is a real number). The *linear space spanned* by $X = \{x_1, \ldots, x_m\}$ (also called the span of X), denoted by $Span(X)$, is the set of points x of $\mathbb{R}^n$ which can be expressed as linear combinations of $x_1, \ldots, x_m$. Given a set X of $\mathbb{R}^n$, the span of X is the smallest vectorial space containing the set X. In the following we will consider a little bit further the other types of combinations.

A set $x_1, \ldots, x_m$ of vectors are *linearly independent* if $\sum_{i=1}^m \lambda_i x_i = 0$ implies that for every $i \leq m$, $\lambda_i = 0$. The *dimension* of the space spanned by $x_1, \ldots, x_m$ is the cardinality of a maximum subfamily of $x_1, \ldots, x_m$ which is linearly independent.

The points $x_0, \ldots, x_\ell$ of an affine space are said to be *affinely independent* if the vectors $x_1 - x_0, \ldots, x_\ell - x_0$ are linearly independent. In other words, if we consider the space to be "centered" on x_0 then the vectors corresponding to the other points in the vectorial space are independent.

2 *Convex sets and functions*

In the following we denote by $||\cdot||$ a norm on $\mathbb{R}^n$ (we refer the reader to a linear algebra book for the formal definition of a norm). The *distance* between x and y, denoted by $d(x, y)$, is $||x - y||$. We denote by $B(x, \epsilon)$ the *ball of center x and of radius ϵ* which is the set of points are distance at most ϵ from x.

2.1 Convex sets.

A set X is *convex* if for any pair x, y of points of X and for every $0 \leq \lambda \leq 1$ the point $z = \lambda x + (1 - \lambda)y$ is in X. Differently, a set X is convex if the segment between every pair of points of X is contained in the set X.

Lemma 1. *A set X is convex if and only if any convex combination of a finite number of points of X is in X.*

Let X be a set of points. The *convex hull* of X is the set of points which are convex combinations of a finite number of points of X. The convex hull of X is denoted by $Conv(X)$. Let us first show that the convex hull of X is convex.

Lemma 2. *Let X be a set of points of $\mathbb{R}^n$. The set $Conv(X)$ is convex.*

Lemma 3. *The intersection of any collection (not necessarily finite) of convex sets is convex.*
For any $X \subseteq \mathbb{R}^n$, the set $Conv(X)$ is the intersection of all convex sets that contain X.

In particular, it means that the convex hull of X is the smallest convex set containing X. An *extreme point* of a convex set X is any point $x \in X$ which is not a convex combination of other points in X. (we will see a bit later several other ways to define extremal points).

Exercise 1. *Give an example of convex set which is not the convex hull of its extreme points.*

2.2 Convex functions.

A *(global) maximizer* for an optimization problem Π is a feasible point $x \in F$ such that for all $y \in F$, we have $||f(x)|| \geq ||f(y)||$ (notations are chosen accordingly to standard notations given at lecture 1). A *local maximizer* is a feasible point x for which there exists $\epsilon > 0$ such that for every y in $B(x, \epsilon) \cap F$, we have $||f(x)|| \geq ||f(y)||$.

Similarly *(global) minimizer* for an optimization problem Π is a feasible point $x \in F$ such that for all $y \in F$, we have $||f(x)|| \leq ||f(y)||$. A *local maximizer* is a feasible point x for which there exists $\epsilon > 0$ such that for every y in $B(x, \epsilon) \cap F$, we have $||f(x)|| \leq ||f(y)||$.

The *epigraph* of a function $f : F \longrightarrow \mathbb{R}$ is

$$Epi(f) = \{(x, y) \in F \times \mathbb{R} \text{ such that } y \geq f(x)\}.$$

Equivalently, the epigraph is the set of points which lie above the function f. Note that the epigraph is a set of a set of points of a $(d+1)$-dimensional space if F lies in $\mathbb{R}^d$. Let F be a convex set. A function $f : F \longrightarrow \mathbb{R}$ is *convex* if the epigraph of f is convex. Equivalently, a function f is convex if for every $x, y \in F$ and for every $0 \leq \lambda \leq 1$ we have:

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

A function f is *concave* if $(-f)$ is convex. Equivalently

$$f(\lambda x + (1-\lambda)y) \geq \lambda f(x) + (1-\lambda)f(y).$$

Exercise 2. 1. *Show that the sum of convex functions is convex.*

2. *Does it hold for the product of convex functions? For the multiplication by a scalar?*

3. *Let $c \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$. Show that $f(x) = c^T x + \alpha$ is convex and concave.*

Sketch of solution. **(1)** Let f, g be two convex functions. We have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$ and $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$. So we also have: $(f+g)(\lambda x + (1-\lambda)y) \leq \lambda(f+g)(x) + (1-\lambda)(f+g)(y)$.
(2) For the multiplication by a non-negative scalar yes. What happens if we multiply by a negative scalar? The functions $f(x) = x$ and $g(x) = x^2$ are convex (why), what about their product?
(3) It is a linear function so it is both convex and concave. (prove it !) □

Theorem 4. *Let D be a convex set. Let $f : D \longrightarrow \mathbb{R}$ be a convex function. If x is a local minimizer of f, then x is also a global minimizer of f.*

Proof. Assume by contradiction that there exists a local minimizer x that is not a global minimizer. And let x^* be such that $f(x) > f(x^*)$. By assumption, there exists $\epsilon > 0$ such that, for every $y \in D \cap B(x, \epsilon)$, we have $||f(y)|| \geq ||f(x)||$ since x is a local minimizer. The set $S = \{\lambda x + (1-\lambda)x^*, \lambda \leq 1\}$ is a segment. Since D is convex, the segment S is in D. Moreover, there exists λ_0 such that if $\lambda \leq \lambda_0$, we have (why?)

$$||\lambda x + (1-\lambda)x^* - x|| \leq \epsilon$$

So in particular, if $\lambda \leq \lambda_0$, we have $(\lambda x + (1-\lambda)x^*) \in B(x, \epsilon)$. Since f is convex we have:

$$f(\lambda x + (1-\lambda)x^*) \leq \lambda f(x) + (1-\lambda)f(x^*) < f(x)$$

a contradiction with the local minimality of x. □

For the same reason we have:

Corollary 5. *Let D be a convex set. Let $f : D \longrightarrow \mathbb{R}$ be a concave function. If x is a local maximizer of f, then x is also a global maximizer. minimizer of f.*

Exercise 3. *Adapt the proof of Theorem 4 for proving the Corollary 5.*

So, given an optimization problem with a linear objective function and a convex feasible region F, we just have to look for local maximizers in order to find global maximizers (when the objective function is concave). The simplex algorithm, for instancce, looks for local maximizers. Indeed it is usually much more simpler to find a local maximizer (which can often be done greedily) instead of a global maximizer (which implies a much deeper understanding of the underlying structure of the problem): a local property is always simpler to catch than a global property.

3 Polyhedra and polytopes

Definition 1. *A* hyperplane *of* $\mathbb{R}^n$ *is a set of the form* $\{x \text{ such that } a^T x = b\}$ *for some* $a, b \in \mathbb{R}^n$ *where* a *is a non-zero vector.*
A half-space *of* $\mathbb{R}^n$ *is a set of the form* $\{x \text{ such that } a^T x \leq b\}$ *for some* $a, b \in \mathbb{R}^n$ *where* a *is a non-zero vector.*

In the following, we will often use the word *constraint* in place of half-space when we consider the sets as feasible regions of a linear program. A hyperplane is a set satisfying one linear equation (recall that one linear equation creates one "constraint" and then the dimension of the affine space reaching this equality is $(n-1)$). A half-space is the part of the space below or above a hyperplane. Note that we have seen before that constraints of a LP define half-spaces. The vector a is called a *vector normal* to the hyperplane. Indeed if we consider the "canonical" basis of $\mathbb{R}^n$ with the classical scalar product on it, the line defined by the vector a is the set of vectors of the space normal to all the vectors of the hyperplane $a^T x = 0$ (recall that the classical scalar product of x and y is $x^T y$).

Definition 2. *A* polyhedron *is an intersection of finitely many half-spaces.*

Computationally, we need this finite number of halfspaces. Indeed, if there is a finite number of constraints, we can easily represent the halfspaces. If the number of constraints becomes infinite, we have to be clever when we represent the problem...

Observation 6. *A hyperplane is a polyhedron.*

Proof. Let H be a hyperplane. Then $H = \{x / a^T x = b\}$ for some pair a, b where a is non empty. So $H = \{x / a^T x \geq b\} \cap \{x / a^T x \leq b\}$ □

It means that we can "squeeze" polyhedron of $\mathbb{R}^n$ in spaces of smaller dimension (topologically, the interior of a polyhedron can be empty).

Lemma 7.
- *A hyperplane is a convex set.*
- *A polyhedron is a convex set.*

So the feasible region of a LP, which is a polyhedron, is a convex set. Moreover the objective function of a LP is a linear function which is both a convex and a concave function. **So if we find a local minimizer or a global minimizer of the objective function, we know that it is a global minimizer or a global minimizer according to Theorem 4 and Corollary 5.**

A *polytope* is a set which is the convex hull of finitely many points. *i.e.* it is a set of the form $conv(X)$ where X is finite. The extreme points of a polytope are sometimes called *vertices*. We denote by $V(P)$ the set of vertices of P.

Exercise 4. *Prove that if* P *is a polytope, then* $P = Conv(V(P))$.

Note that "finiteness" of X is important. Otherwise we could get "smooth" regions which cannot be input to a computer with a finite number of linear constraints, e.g. a closed circle is the convex hull of its extreme points (the points on the boundary). Note that it in particular ensures that every convex set is not a polyhedron.

A set X is *bounded* if there exists a constant M such that $||x|| \leq M$ for every $x \in X$.

Theorem 8 (Weyl-Minkowski). *P is a polytope if and only if it is a bounded polyhedron.*

Even if it seems visually obvious, the proof of Theorem 8 requires careful work to prove rigorously. In particular, it says that every polytope is indeed the solution space of some system of inequalities $Ax \leq b$ (we already mentioned it in Lecture 1). Hence we can view a polytope as having two descriptions: one given by its vertices, and one given by the system of constraints $Ax \leq b$. In the following we will juggle with both concepts.

4 Faces and facets

4.1 Faces and vertices

Let P be a polyhedron. Let c be a nonzero vector such that $\delta = \max_{x \in P} c^T x$ is finite. We then call

$$H = \{x \text{ such that } c^T x = \delta\}$$

a *supporting hyperplane of* P. Roughly speaking, a supporting hyperplane is just the hyperplane normal to the vector c that is maximized in the direction of c (in such a way it nevertheless intersects P). A *face* of a polyhedron P is any set of the form $P \cap H$ where H is a supporting hyperplane. Informally, a face is a subset of points of the polyhedron which can have the same behavior according to an objective function (which is just a direction of the space). By convention, the whole polyhedron P is also a face. We also often consider that the empty set is a face. Note that each face is again a polyhedron. (Why? Hint: show that a face satisfies all the constraints of P plus some new constraints.)

Theorem 9. *A set F is a face of P if and only if*

(i) F is a nonempty subset of P.

(ii) $F = \{x \in P$ such that $A'x = b'\}$ where $A'x \leq b'$ is a subsystem of $Ax \leq b$.

Note that by definition, for any face F there is a linear objective function $c^T x$ such that the vectors in F are precisely the optimal solutions for the problem $\max_{x \in P} c^T x$.

Any face of P is obtained by turning some subset of the inequalities into equalities. (NB It is easier to show one direction of the theorem: if F satisfies (i,ii) then it is a face.)

The *dimension* $dim(P)$ of a polyhedron P is the maximum number of affinely independent points in P minus 1. A *d-dimensional face* of a polyhedron P is a face of the polyhedron of dimension d (well-defined since the face is a polyhedron). The 0-dimensionnal faces are the extremal points of the polyhedron (also called *vertices*), the 1-dimensionnal faces are called the *edges*. Given a polytope of dimension k, the faces of dimension $(k-1)$ are called *facets* of the polytope P.

Exercise 5. *List the faces of a* 3*-dimensional cube.*

Sketch of solution. There is 1 face of dimension 3: the cube itself. There are 6 faces of dimension 2 which are the facets (usually called faces in the common language) of the cube. Then there are 12 edges and there are 8 vertices. So the number of faces is $1 + 6 + 12 + 8 = 27$. Give a supporting hyperplane for each of these faces... □

Exercise 6. *Let $P = \{x \in \mathbb{R}^n$ such that $x_i \geq 0$ for $i = 1, \dots, n\}$. For every k, how many faces of dimension k does P have?*

A face is called a *minimal face* if it does not strictly contain another face. In most examples, the vertices are minimal faces (consider for instance the faces of the cube). There are however polyhedra for which the minimal faces are not vertices (these counter-examples cannot be polytopes, why?). For instance $P = \{(x_1, x_2) \in \mathbb{R}^2 : x_1 \leq 0\}$. The only face of this polyhedron (apart from itself) is the line $x_1 = 0$.

There exists a characterization of minimal faces similar to the characterization of Theorem 9.

Theorem 10. *A set F is a minimal face of P if and only if*

(i) F is a nonempty subset of P.

(ii) $F = \{x$ such that $A'x = b'\}$ where $A'x \leq b'$ is a subsystem of $Ax \leq b$.

Note that we do not answer anymore x to be in P, which makes a *huge* difference. In other words, if we find a subsystem such that $A'x = b$ automatically implies $x \in P$, then all the solutions of this subsystem form a minimal face of P. When F has dimension 0, it contains a single point which is a vertex.

Let x be an optimal solution to a given linear program. A constraint is *binding* (or active or tight) at x if it is satisfied with equality at x. It morally means that we cannot "push" anymore in the direction of the constraint without leaving the polyhedron (and then violating the constraint). Stated differently, a constraint $c^t x \leq b$ (or $c^t x = b$) is *tight* for a solution x^* if $c^t x^* = b$. In other words, it means that x^* is in the hyperplane defined by $c^t x = b$.

Corollary 11. *If v is a vertex of a polyhedron P, then there is some linearly independent system of equalities $A'x = b'$ which is a subsystem of $Ax = b$, for which v is the unique solution.*

Note moreover that we precisely know the size of this system: it has size n.

Vertices and solutions of a LP

Exercise 7. *Consider the linear programming problem of minimizing $c^t x$ over a non-empty, polyhedron P. Suppose moreover that there exists an optimal solution with bounded value. Then, there exists an optimal solution which is an extreme point of P .*

4.2 Basic feasible solutions

4.3 Basic and nonbasic variables

Let P be a polyhedron with (i) inequality constraints and (ii) equality constraints. Let us assume that the number of variables (ie the dimension of the space) is n and that the number of variables is m.

A constraint $c^t x \leq b$ (or $c^t x = b$) is *tight* for a solution y if $c^t y = b$. In other words, it means that y is in the hyperplane defined by $c^t x = b$.

Definition 3. *Let P be a polyhedron defined by inequality or equality constraints, and let $y \in \mathbb{R}^n$. The vector y is a basic solution if:*

- *All equality constraints are tight;*
- *Among the constraints that are tight at y, n of them are linearly independent.*

If $y \in P$ and y is a basic feasible solution (BFS), then we say that y is a basic feasible solution.

Similarly, the second condition can be rephrased as follows: the submatrix of the constraint matrix induced by the set of tight constraint is a matrix of rank ℓ (where $\ell = \min(n, m)$ with n the number of variables and m the number of constraints).

If more than n constraints are active at a point y, then x is said to be degenete. Degenerate basic feasible solutions might lead to complication when applying the simplex algorithm. For now, we assume that all basic feasible solutions in P are non-degenerate. (We will address issues that arise from degenerate solutions in the next lecture.)

4.4 Equivalent definitions of vertices

Equivalent definition of vertices. So finally we have seen three alternative definitions of vertex of

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

- $z \in P$ is a extreme point if and only if it does not lie on a line segment $[x, y]$ where $x, y \in P$ (*i.e.* it is not a convex combination of points of P).
- $z \in P$ is a basic feasible solution if and only if $rank(A_z) = n$ (where n is the number of columns of A) where A_z is the submatrix of A consisting of the rows a_i for which $a_i^T z = b_i$ *i.e.* all of the tight rows of the system $Ax \leq b$ for the vector z.
- $z \in P$ is a vertex if and only if there is a (objective) vector $c \in \mathbb{R}^n$ such that z is the unique optimal solution to $\max\{c^T x : x \in P\}$.

In the following we will use one or another of these formulations depending on the property we want to exhibit.

Exercise 8. *Let P be a polyhedron. The following are equivalent:*

- *x^* is a vertex;*
- *x^* is an extreme point;*
- *x^* is a basic feasible solution.*

5 Extended formulations

A constraint of a polyhedron is *useless* if deleting this constraint changes the polyhedron. The complexity of a polyhedron is related to its number of non-useless constraints. The more such constraints we have, the longer the Simplex algorithm (that follows constraints of the polyhedron) is. A branch of optimization is devoted to see if we can decrease the number of constraints of a polyhedron.

In particular, one may ask the following counter-intuitive question: *by increasing the dimension of the space (by a small amount), is it possible to drastically decrease the number of constraints?* Slightly more formally, assume that we have an exponential number of constraints in terms of the number of variables. Is it possible to decrease the number of constraints by increasing (by a polynomial factor) the number of variables?

Even more surprising, the answer to this question is **YES**. The rest of this subsection is devoted to show it on an example.

Lift of polytopes A polytope P is a *lift* of a polytope Q if P is the image of Q under an *affine projection* π. In other words, a vector $(x_1, \dots, x_k)$ is in the polytope P if and only if there exists a vector of the form $(x_1, \dots, x_k, y_1, \dots, y_\ell)$ in Q.

Lifts are an important tool in combinatorial optimization since, if we can optimize a (linear) function on the polytope Q that depends only of the k first coordinate, then we can perform the same optimization on P. Indeed, we can effiently find an optimal solution on Q and then restrict this solution to the k first coordinate in order to obtain a solution for P. Conversly, if we want to maximize a linear function on P, then we can optimize the same linear function on Q and then obtain a solution on P. In other words:

$$\max_{x \in P} w^t x = \max_{y \in Q} \overline{w}^t y \qquad \text{where } \overline{w} = (w, 0, 0, \dots, 0)$$

We then say that Q is an *extended formulation* of P.

What is the interest of lifts and extended formulation? If we can find a higher dimension definition of our polytope that is simpler than the definition in the original polytope, then we may want to solve our problem on this higher dimensional polytope than in the original one and then obtain a more efficient algorithm. But does such a formulation exist? In general it does not necessarily exist, but it might exist. Let us prove the existence of (compact) extended formulations on several examples.

Example: the cross polytope. The *cross polytope* is the polytope defined as follows:

$$C_d = \{x \text{ such that } \|x\|_1 = 1\} = \{x \in \mathbb{R}^d \text{ such that } \pm x_1 \pm x_2 \ldots \pm x_d \leq 1\}$$

Exercise 9. *Show that this polytope is defined by 2^d contraints. In other words, no constraints is useless. And then the number of facets of C_d is exponential.*

But you can remark that if we consider the polytope Q_d defined as:

$$Q_d = \{(x, y) \in \mathbb{R}^{2d} \text{ such that } \sum_i y_i = 1, \text{ and } -y_i \leq x_i \leq y_i \text{ and } y_i \geq 0\}$$

then a well-chosen of Q_d is exactly C_d. It will shown during the "TD". So C_d has an extended formulation with the extended formulation whose number of faces is exponentially smaller.