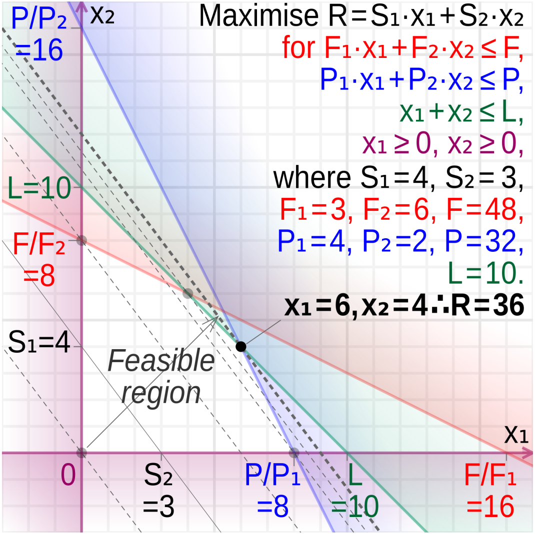
Lecture 5 Polyhedron Theory

Can Li

ChE 597: Computational Optimization
Purdue University

Geometric intuition

Recall from the last lecture the farmer planning problem.



The optimal solution is obtained at a **vertex**.

Polyhedron theory

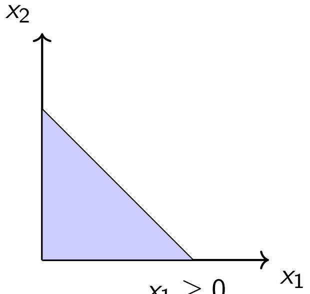
In this lecture, we will provide the mathematical treatment as well as the geometric intuition of "vertices", "faces", "rays", of the polyhedron set $Ax \geq b$.
These concepts are key to the development of the simplex algorithm for solving linear programs

Polyheron Definition

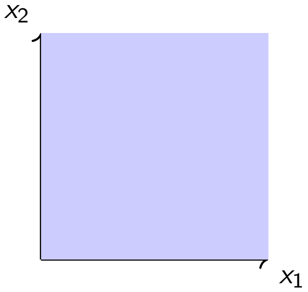
Definition A polyhedron is a set that can be described in the form $\{x \in \mathbb{R}^n \mid Ax \geq b\}$, where A is an $m \times n$ matrix and b is a vector in $\mathbb{R}^m$.

Remark Note that sets in the form $\{x \in \mathbb{R}^n \mid Ax \leq b\}$, $\{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ are also polyhedra.

A polyhedron can be bounded or unbounded.



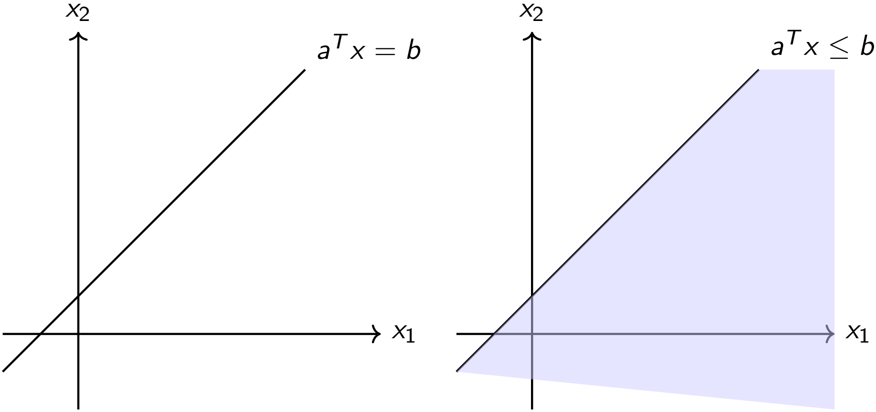
$$x_1 \geq 0$$
$$x_2 \geq 0$$
$$x_1 + x_2 \leq 2$$



$$x_1 \geq 0$$
$$x_2 \geq 0$$

Hyperplane and halfspace

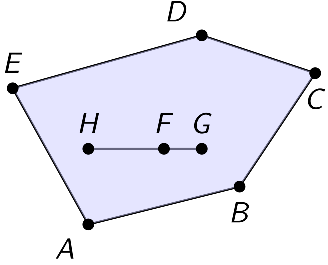
Definition Let a be a nonzero vector in $\mathbb{R}^n$ and let b be a scalar.
(a) The set $\left\{x \in \mathbb{R}^n \mid a^T x = b\right\}$ is called a hyperplane.
(b) The set $\left\{x \in \mathbb{R}^n \mid a^T x \geq b\right\}$ is called a halfspace.



(a) Hyperplane (b) Halfspace

Extreme point

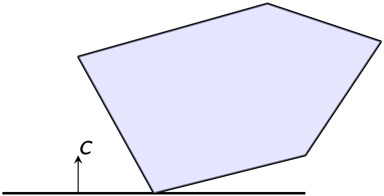
Definition Let P be a polyhedron. A vector $x \in P$ is an extreme point of P if we cannot find two vectors $y, z \in P$, both different from x, and a scalar $\lambda \in [0, 1]$, such that $x = \lambda y + (1 - \lambda)z$. In other words, x cannot be expressed as the convex combination of two other points in P.



A,B,C,D,E are the extreme points of P. F is not an extreme point because it is a convex combination of H and G.

Vertex

Definition Let P be a polyhedron. A point $x^* \in P$ is a vertex of P if there exists some c such that $c^T x^* < c^T x$ for all x satisfying $x \in P$ and $x \neq x^*$.
In other words, x^* is a vertex of P if and only if P is on one side of a hyperplane (the hyperplane $\left\{x \mid c^T x = c^T x^*\right\}$) which meets P only at the point x^*;



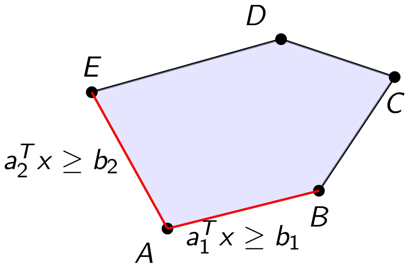
Remark 1 : Intuitively, vertex is equivalent to extreme point.
Remark 2 : It is hard to verify whether a point is an extreme point/vertex just by using their definitions. Another definition called "basic feasible solution" is much easier to verify.

Active constraints

Before defining a basic feasible solution, we need to define active constraints.

Definition Consider a polyhedron $P \subset \mathbb{R}^n$ defined in terms of the constraints $Ax \geq b$ where A is an $m \times n$ matrix and b is a vector in $\mathbb{R}^m$. We say that the corresponding constraint $a_i^T x \geq b_i$ is **active** or **binding** at x^* iff $a_i^T x^* = b_i$.

Illustrative example of active constraints



The figure shows the two active constraints at extreme point A.

$a_1^T x \geq b_1$ and $a_2^T x \geq b_2$

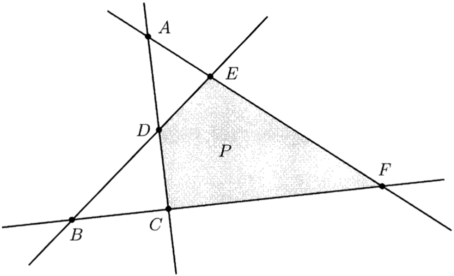
Remark For $x \in \mathbb{R}^n$, n linearly independent constraints uniquely defines a point. The solution to n linearly independent linear equations $a_i^T x_i = b_i$ are unique.

Basic feasible solution

Definition Consider a polyhedron P defined by linear equality and inequality constraints, and let x^* be an element of $\mathbb{R}^n$.

1. The vector x^* is a **basic solution** if:
   - All equality constraints are active;
   - Out of the constraints that are active at x^*, there are n of them that are linearly independent.
2. If x^* is a basic solution that satisfies all of the constraints, we say that it is a **basic feasible solution**.

Illustrative example of basic solution and basic feasible solution



The points A, B, C, D, E, F are all basic solutions because at each one of them, there are two linearly independent constraints that are active. Points C, D, E, F are basic feasible solutions.

Equivalence of extreme point, vertex, basic feasible solution

Theorem Let P be a nonempty polyhedron and let $x^* \in P$. Then, the following are equivalent:
(a) x^* is a vertex;
(b) x^* is an extreme point;
(c) x^* is a basic feasible solution.

Vertex $\Rightarrow$ Extreme point

Suppose that $x^* \in P$ is a vertex. Then, by definition, there exists some $c \in \mathbb{R}^n$ such that $c^T x^* < c^T y$ for every y satisfying $y \in P$ and $y \neq x^*$. If $y \in P, z \in P, y \neq x^*, z \neq x^*$, and $0 \leq \lambda \leq 1$, then $c^T x^* < c^T y$ and $c^T x^* < c^T z$, which implies that $c^T x^* < c^T(\lambda y + (1 - \lambda)z)$ and, therefore, $x^* \neq \lambda y + (1 - \lambda)z$. Thus, x^* cannot be expressed as a convex combination of two other elements of P and is, therefore, an extreme point

Extreme point $\Rightarrow$ Basic feasible solution

Suppose that $x^* \in P$ is not a basic feasible solution. We will show that x^* is not an extreme point of P. Let $I = \left\{ i \mid a_i^T x^* = b_i \right\}$. Since x^* is not a basic feasible solution, there do not exist n linearly independent vectors in the family $a_i, i \in I$. Thus, the vectors $a_i, i \in I$, lie in a proper subspace of $\mathbb{R}^n$, and there exists some nonzero vector $d \in \mathbb{R}^n$ such that $a_i^T d = 0$, for all $i \in I$. Let ϵ be a small positive number and consider the vectors $y = x^* + \epsilon d$ and $z = x^* - \epsilon d$. Notice that $a_i^T y = a_i^T x^* = b_i$, for $i \in I$. Furthermore, for $i \notin I$, we have $a_i^T x^* > b_i$ and, provided that ϵ is small, we will also have $a_i^T y > b_i$. (It suffices to choose ϵ so that $\epsilon \left| a_i^T d \right| < a_i^T x^* - b_i$ for all $i \notin I$.) Thus, when ϵ is small enough, $y \in P$ and, by a similar argument, $z \in P$. We finally notice that $x^* = (y + z)/2$, which implies that x^* is not an extreme point.

Basic feasible solution ⇒ Vertex

Let x^* be a basic feasible solution and let $I = \left\{i \mid a_i^T x^* = b_i\right\}$. Let $c = \sum_{i \in I} a_i$. We then have

$$c^T x^* = \sum_{i \in I} a_i^T x^* = \sum_{i \in I} b_i.$$

Furthermore, for any $x \in P$ and any i, we have $a_i^T x \geq b_i$, and

$$c^T x = \sum_{i \in I} a_i^T x \geq \sum_{i \in I} b_i \quad (**)$$

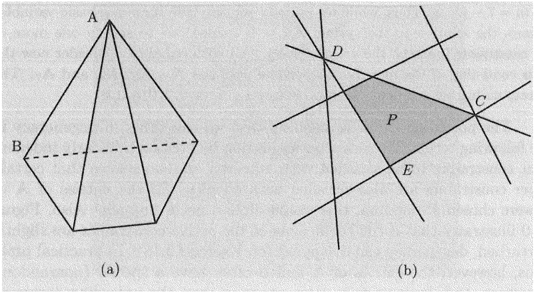
This shows that x^* is an optimal solution to the problem of minimizing $c^T x$ over the set P. Furthermore, equality holds in (**) if and only if $a_i^T x = b_i$ for all $i \in I$. Since x^* is a basic feasible solution, there are n linearly independent constraints that are active at x^*, and x^* is the unique solution to the system of equations $a_i^T x = b_i, i \in I$. It follows that x^* is the unique minimizer of $c^T x$ over the set P and, therefore, x^* is a vertex of P.

Some Corollaries

Corollary 1 The number of extreme points/vertex/BFS in P is upper bounded by $\binom{m}{n}$. We have at most $\binom{m}{n}$ ways of choosing n linearly independent constraints among the m constraints.
Corollary 2 A nonempty polyhedron $Ax \geq b$ has at least one extreme point if and only if there exist n linearly independent vectors out of the m constraint coefficients $a_1, \ldots, a_m$.

Degeneracy

A basic solution $x \in \mathbb{R}^n$ is said to be degenerate if more than n of the constraints are active at x.



The points A and C are degenerate basic feasible solutions. The points B and E are nondegenerate basic feasible solutions. The point D is a degenerate basic solution.

Degeneracy in standard form polyhedra

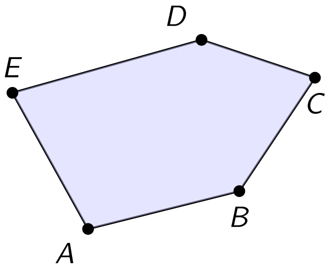
Consider the standard form polyhedron $P = \{x \in \mathbb{R}^n \mid Ax = b, x \geq 0\}$ and let x be a basic solution. Let m be the number of rows of A. The vector x is a degenerate basic solution iff more than $n - m$ of the components of x are zero, i.e., there are more than n active constraints.

Faces

Definition A face of the nonempty polyhedral set P is a nonempty subset of P where a subset of the inequalities are active.

Examples of faces

- The set P itself is a face (no inequalities are active).
- An extreme point is a face (n linearly independent inequalities are active).
- Edges AB, BC, CD, DE, and EA in the figure are faces (1 inequality is active).

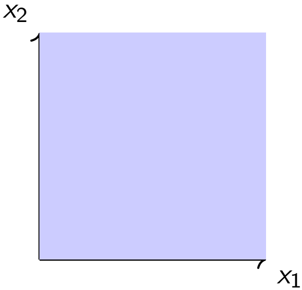


Polyhedral cone

Definition A set $C \subset \mathbb{R}^n$ is a cone if $\lambda x \in C$ for all $\lambda \geq 0$ and all $x \in C$.

Definition A polyhedron of the form $P = \{x \in \mathbb{R}^n \mid Ax \geq 0\}$ is easily seen to be a nonempty cone and is called a **polyhedral cone**.

A simple example of a polyhedral cone:



$$x_1 \geq 0$$
$$x_2 \geq 0$$

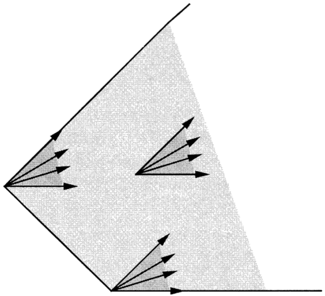
Rays and recession cones

Definition Consider a nonempty polyhedron

$$P = \{x \in \mathbb{R}^n \mid Ax \geq b\},$$

and let us fix some $y \in P$. We define the **recession cone** at y as the set of all directions d along which we can move indefinitely away from y, without leaving the set P. More formally, the recession cone is defined as the set

$$\{d \in \mathbb{R}^n \mid A(y + \lambda d) \geq b, \text{ for all } \lambda \geq 0\}.$$



Rays and recession cones

Definition Consider a nonempty polyhedron

$$P = \{x \in \mathbb{R}^n \mid Ax \geq b\},$$

and let us fix some $y \in P$. We define the **recession cone** at y as the set of all directions d along which we can move indefinitely away from y, without leaving the set P. More formally, the recession cone is defined as the set

$$\{d \in \mathbb{R}^n \mid A(y + \lambda d) \geq b, \text{ for all } \lambda \geq 0\}.$$

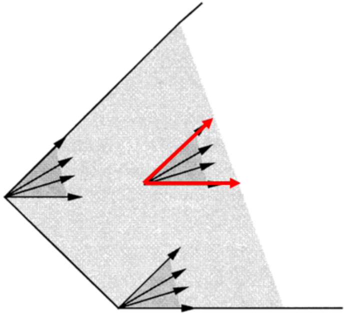
It is easily seen that this set is the same as

$$\{d \in \mathbb{R}^n \mid Ad \geq 0\},$$

and is a polyhedral cone.
This shows that the recession cone is independent of the starting point y. The nonzero elements of the recession cone are called the **rays** of the polyhedron P.

Extreme rays

Definition Let $C \subseteq \mathbb{R}^n$ be a polyhedral cone. A nonzero $d \in C$ is an extreme ray of C if there do not exist linearly independent $u, v \in C$ and positive scalars λ and γ such that $d = \lambda u + \gamma v$.



Extreme rays

Theorem Let $C \subseteq \mathbb{R}^n$ be given by $\{x \in \mathbb{R}^n : Ax \geq 0\}$ for some $A \in \mathbb{R}^{m \times n}$. Let $d \in C$ be nonzero. Let $A^= x = 0$ denote the subsystem of $Ax \geq 0$ consisting of all the inequalities active at d. Then d is an extreme ray of C if and only if $\operatorname{rank}(A^=) = n - 1$.
Proof Necessity Suppose that $\operatorname{rank}(A^=) < n - 1$. There exists a nonzero vector y in the nullspace of $A^=$ such that d and y are linearly independent. For a sufficiently small $\epsilon > 0, d \pm \epsilon y \in C$. Note that $d - \epsilon y$ and $d + \epsilon y$ are linearly independent and $d = \frac{1}{2}(d - \epsilon y) + \frac{1}{2}(d + \epsilon y)$, implying that d is not an extreme ray.
Sufficiency Suppose that $d = \lambda u + \gamma v$ for some linearly independent $u, v \in C$ and scalars $\lambda, \gamma > 0$. Then

$$0 = A^= d = \lambda A^= u + \gamma A^= v \geq 0$$

Hence, equality holds throughout. Since $\lambda, \gamma > 0$, we must have $A^= u = A^= v = 0$. Hence, u and v are linearly independent vectors in the nullspace of $A^=$, implying that $\operatorname{rank}(A^=) < n - 1$.

Minkowski-Weyl Theorem

Theorem (Representation of polyhedra) A polyhedron P can be represented as

$$P = \Big\{ x \in \mathbb{R}^n : x = \sum_{k \in K} \lambda_k v^k + \sum_{j \in J} \mu_j r^j$$

$$\text{with } \sum_{k \in K} \lambda_k = 1, \quad \lambda_k \geq 0 \; \forall k \in K, \mu_j \geq 0 \; \forall j \in J \Big\},$$

where $\left\{ v^k \right\}_{k \in K}$ is the set of extreme points of P and $\left\{ r^j \right\}_{j \in J}$ is the set of extreme rays of P.

In words, a polyhedron can be represented as the **Minkowski sum** of the convex hull of its extreme points and the conic hull of its extreme rays.

Definition Minkowski sum of two sets A and B is
$A + B = \{a + b \mid a \in A, b \in B\}$

Representation of bounded polyhedra

Corollary A bounded polyhedron can be represented as the convex hull of its extreme points.

When will the optimal objective value be $-\infty$

Corollary Consider the problem of minimizing $c^T x$ over a pointed polyhedral cone $C = \left\{x \in \mathbb{R}^n \mid a_i^T x \geq 0, i = 1, \ldots, m\right\}$. The optimal cost is equal to $-\infty$ if and only if some extreme ray d of C satisfies $c^T d < 0$.

Existence of an optimal extreme point

Corollary Suppose we are minimizing. When P is nonempty and the optimal objective value is not $-\infty$, the optimal solution can always be obtained at an extreme point.

Proof. Using the Minkowski theorem

$$P = \left\{ x \in \mathbb{R}^n : x = \sum_{k \in K} \lambda_k v^k + \sum_{j \in J} \mu_j r^j \right.$$

$$\left. \text{with } \sum_{k \in K} \lambda_k = 1, \quad \lambda_k \geq 0 \ \forall k \in K, \mu_j \geq 0 \ \forall j \in J \right\},$$

If the optimal cost is not $-\infty$, we must have $c^T r^j \geq 0$ for all the extreme rays. We can pick one extreme point that minimizes $c^T v^k$. This extreme point is an optimal solution.

Remark This is the basis of the simplex algorithm for solving linear programs.

Reference

1. Chapter 2. Bertsimas, D., & Tsitsiklis, J. N. (1997). Introduction to linear optimization (Vol. 6, pp. 479-530). Belmont, MA: Athena scientific.