Chapter 5

Linear programming

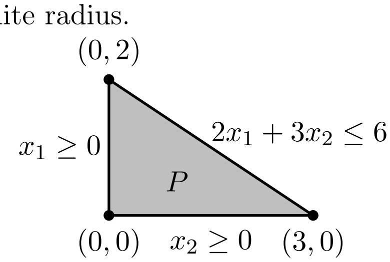
One of the main tools in combinatorial optimization is **linear programming**. We want to quickly review the key concepts and results. Since most statements and proofs are known from course 407, from time to time we will be satisfied with informal proof sketches.

To clearify notation, the set of all real numbers will be denoted as $\mathbb{R}$ and the set of all integers by $\mathbb{Z}$. A function f is said to be real valued if its values are real numbers. For instance a vector $c = (c_1, \ldots, c_n) \in \mathbb{R}^n$ defines the linear function $f : \mathbb{R}^n \to \mathbb{R}$ such that $x \mapsto c \cdot x := c_1x_1 + c_2x_2 + \cdots + c_nx_n$. Linear programs always have linear objective functions $f(x) = c \cdot x$ as above. Note that this is a real valued function since $c \cdot x \in \mathbb{R}$.

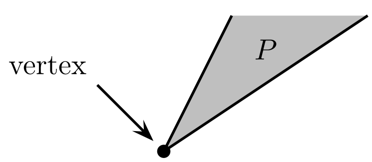
A **polyhedron** $P \subseteq \mathbb{R}^n$ is the set of all points $x \in \mathbb{R}^n$ that satisfy a finite set of linear inequalities. Mathematically,

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

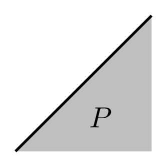
for some matrix $A \in \mathbb{R}^{m \times n}$ and a vector $b \in \mathbb{R}^m$. A polyhedron can be presented in many different ways such as $P = \{x \in \mathbb{R}^n \,:\, Ax = b, x \geq \mathbf{0}\}$ or $P = \{x \in \mathbb{R}^n \,:\, Ax \geq b\}$. All these formulations are equivalent. A polyhedron is called a **polytope** if it is bounded, i.e., can be enclosed in a ball of finite radius.



polytope/polyhedron

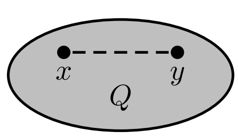


polyhedron

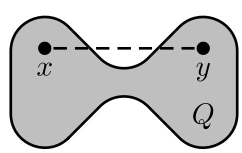


polyhedron

Definition 7. A set $Q \subseteq \mathbb{R}^n$ is **convex** if for any two points x and y in Q, the line segment joining them is also in Q. Mathematically, for every pair of points $x, y \in Q$, the **convex combination** $\lambda x + (1 - \lambda)y \in Q$ for every λ such that $0 \leq \lambda \leq 1$.



convex

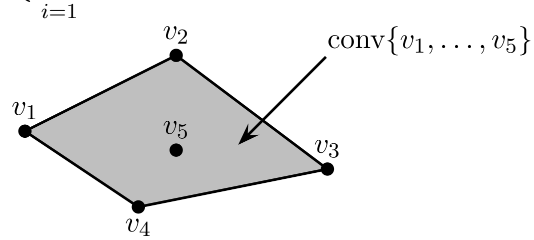


not convex

Obviously, polyhedra are convex.

Definition 8. A **convex combination** of a finite set of points $v_1, \ldots, v_t$ in $\mathbb{R}^n$, is any vector of the form $\sum_{i=1}^t \lambda_i v_i$ such that $0 \leq \lambda_i \leq 1$ for all $i = 1, \ldots, t$ and $\sum_{i=1}^t \lambda_i = 1$. The set of all convex combinations of $v_1, \ldots, v_n$ is called the **convex hull** of $v_1, \ldots, v_n$. We denote it by

$$\mathrm{conv}\{v_1, \ldots, v_n\} = \Big\{ \sum_{i=1}^n \lambda_i v_i \mid \lambda_1 + \ldots + \lambda_n = 1; \lambda_i \geq 0 \ \forall i = 1, \ldots, n \Big\}$$



Theorem 28. *Every polytope P is the convex hull of a finite number of points (and vice versa).*

For a convex set $P \subseteq \mathbb{R}^n$ (such as polytopes or polyhedra) we call a point $x \in P$ an **extreme point** / **vertex** of P if there is no vector $y \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ with both $x + y \in P$ and $x - y \in P$.

A **linear program** is the problem of maximizing or minimizing a linear function of the form $\sum_{i=1}^n c_i x_i$ over all $x = (x_1, \ldots, x_n)$ in a polyhedron P. Mathematically, it is the problem

$$\min \Big\{ \sum_{i=1}^n c_i x_i \mid Ax \leq b \Big\}$$

for some matrix A and vector b (alternatively max instead of min).

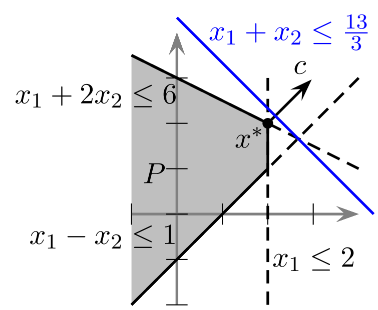
5.1 Separation, Duality and Farkas Lemma

The number 1 key concept in linear optimization is **duality**. We want to motivate this with an example. Consider the linear program

$$\max\{x_1 + x_2 \mid x_1 + 2x_2 \leq 6,\ x_1 \leq 2,\ x_1 - x_2 \leq 1\}$$

(which is of the form $\max\{cx \mid Ax \leq b\}$). First, let us make the following observation: if we add up non-negative multiples of feasible inequalities, then we obtain again an inequality that is valid for each solution x of the LP. For example we can add up the inequalities in the following way:

$$\begin{array}{rcccl}
\frac{2}{3} \cdot (& x_1 & +2x_2 & \leq & 6) \\
0 \cdot (& x_1 & & \leq & 2) \\
\frac{1}{3} \cdot (& x_1 & -x_2 & \leq & 1) \\
\hline
& x_1 & +x_2 & \leq & \frac{13}{3} \approx 4.33
\end{array}$$



Accidentially, the feasible inequality $x_1 + x_2 \leq \frac{13}{3}$ that we obtain has the objective function as normal vector. Hence for each $(x_1, x_2) \in P$ we must have $cx \leq \frac{13}{3}$, which provides an **upper bound** on the value of the LP. Inspecting the picture, we quickly see that optimum solution is

$x^* = (2,2)$ with objective function $cx^* = 4$. Now, let's generalize our observations. Consider the following pair of linear programs

$$\begin{array}{rll} \textbf{primal} & (P): & \max\{cx \mid Ax \leq b\} \\ \textbf{dual} & (D): & \min\{by \mid yA = c,\ y \geq \mathbf{0}\} \end{array}$$

The dual LP is searching for inequalities $(yA)x \leq yb$ that are feasible for any primal solution x; moreover it is looking for a feasible inequality so that the normal vector $yA = c$ is the objective function, so that yb is an upper bound on the primal LP. In other words: the dual LP is searching for the best upper bound for the primal LP.

Theorem 29 (Weak duality Theorem)**.** *One has $(P) \leq (D)$. More precisely if we have (x,y) with $Ax \leq b$, $yA = c$ and $y \geq \mathbf{0}$, then $cx \leq by$.*

Proof. One has

$$\underbrace{c}_{=yA} x = \underbrace{y}_{\geq \mathbf{0}} \underbrace{Ax}_{\leq b} \leq yb.$$

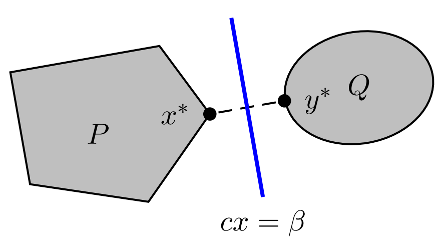
$\square$

In the remainder of this subsection we will show that always $(P) = (D)$, that means we can always combine a feasible inequality that gives a tight upper bound. But first, some tools:

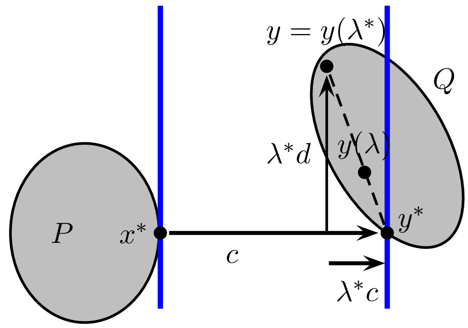
Theorem 30 (Hyperplane separation Theorem)**.** *Let $P, Q \subseteq \mathbb{R}^n$ be convex, closed and disjoint sets with at least one of them bounded. Then there exists a hyperplane $cx = \beta$ with*

$$cx < \beta < cy \quad \forall x \in P\ \forall y \in Q$$

Proof. Let $(x^*, y^*) \in P \times Q$ be the pair minimizing the distance $\|x^* - y^*\|_2$ (this must exist for the following reason: suppose that P is bounded; then P is compact. Then the distance function $d(x) := \min\{\|y - x\|_2 \mid y \in Q\}$ is well-defined and continuous, hence a minimum is attained on P). Then the hyperplane through $\frac{1}{2}(x^* + y^*)$ with normal vector $c = y^* - x^*$ separates P and Q.



To see this, suppose for the sake of contradiction that Q has a point y with $cy < cy^*$. Then we can write this point as $y = x^* + (1 - \lambda^*)c + \lambda^* d$ where d is a vector that is orthogonal to c. Since Q is convex, also the point $y(\lambda) = x^* + (1 - \lambda)c + \lambda d$ for $0 \leq \lambda \leq \lambda^*$ is in Q.



We want to argue that for $\lambda > 0$ small enough, the point $y(\lambda)$ is closer to x^* than y^*.

$$\|x^* - y(\lambda)\|_2^2 = \|(1-\lambda)c + \lambda d\|_2^2 \stackrel{c \perp d}{=} (1-\lambda)^2\|c\|_2^2 + \lambda^2\|d\|_2^2 = \|c\|_2^2 - 2\lambda\|c\|_2^2 + \lambda^2(\|c\|_2^2 + \|d\|_2^2) < \|c\|_2^2$$

if we choose $\lambda > 0$ small enough since the coefficient in front of the linear λ term is negative. $\square$

This theorem is the finite dimensional version of the **Hahn-Banach separation theorem** from functional analysis.

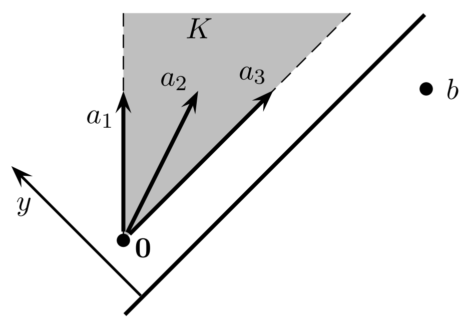
The separation theorem quickly provides us the very useful **Farkas Lemma** which is like a duality theorem without objective function. The lemma tells us that out of two particular linear systems, precisely one is going to be feasible.

Lemma 31 (Farkas Lemma)**.** *One has* $\big(\exists x \geq \mathbf{0} : Ax = b\big) \quad \dot\vee \quad \big(\exists y : y^T A \geq \mathbf{0} \text{ and } yb < 0\big)$.

Proof. First let us check that it is impossible that such x, y exist at the same time since

$$0 \leq \underbrace{yA}_{\geq \mathbf{0}} \underbrace{x}_{\geq \mathbf{0}} = y \underbrace{b}_{=Ax} < 0$$

For the other direction, assume that there is no $x \geq \mathbf{0}$ with $Ax = b$. We need to show that there is a proper y. Consider the cone $K := \{Ax \mid x \geq \mathbf{0}\} = \{\sum_{i=1}^n a_i x_i \mid x_1, \dots, x_n \geq 0\}$ of valid right hand sides, where $a_1, \dots, a_n$ are the columns of A. By assumption we know that $b \notin K$.



But K is a closed convex set, hence there is a hyperplane $y^T c = \gamma$ that separates K and $\{b\}$, i.e.

$$\forall a \in K: \quad y^T a > \gamma > y^T b$$

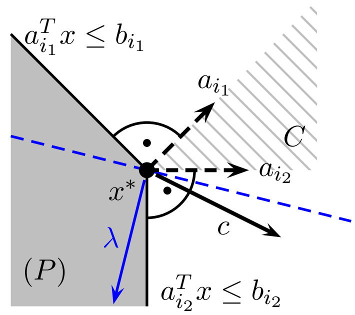
As $\mathbf{0} \in K$ we must have $\gamma < 0$. Moreover all non-negative multiples of columns are in K, that means $a_i x_i \in K$ for all $x_i \geq 0$, thus $y^T(x_i)a_i > \gamma$ for each $i \in [n]$, which implies that even $y^T a_i \geq 0$ for each i. This can be be written more compactly as $y^T A \geq \mathbf{0}$. $\square$

Theorem 32. *One has* $(P) = (D)$*. More precisely, one has*

$$\max\{cx : Ax \leq b\} = \min\{by : yA = c,\ y \geq \mathbf{0}\}$$

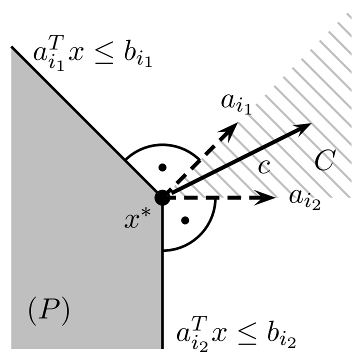
given that both systems have a solution.

Proof sketch. Suppose that (P) is feasible. Let x^* be an optimum solution[1]. Let $a_1, \ldots, a_m$ be rows of A and let $I := \{i \mid a_i^T x^* = b_i\}$ be the *tight* inequalities.



Suppose for the sake of contradiction that $c \notin \{\sum_{i \in I} a_i y_i \mid y_i \geq 0\ \forall i \in I\} =: C$. Then there is a direction $\lambda \in \mathbb{R}^n$ with $c^T\lambda > 0$, $a_i^T \lambda \leq 0\ \forall i \in I$. Walking in direction λ improves the objective function while we do not walk into the direction of constraints that are already tight. In other words, there is a small enough $\varepsilon > 0$ so that $x^* + \varepsilon\lambda$ is feasible for (P) and has a better objective function value — but x^* was optimal. That is a *contradiction!*

Hence we know that indeed $c \in C$, hence there is a $y \geq \mathbf{0}$ with $y^T A = c^T$ and $y_i = 0\ \forall i \notin I$ (that means we only use tight inequalities in y).



Now we can calculate, that the **duality gap** is 0:

$$y^T b - c^T x^* = y^T b - \underbrace{y^T A}_{=c} x^* = y^T(b - Ax^*) = \sum_{i=1}^{m} \underbrace{y_i}_{=0 \text{ if } i \notin I} \cdot \underbrace{(b_i - a_i^T x^*)}_{=0 \text{ if } i \in I} = 0$$

□

Note that moreover, if (P) is unbounded, then (D) is infeasible. If (D) is unbounded then (P) is infeasible. On the other hand, it is possible that (P) and (D) are both infeasible.

[1] That's why this is only a proof sketch. For a polytope it is easy to argue that P is compact and hence there must be an optimum solution. If P is unbounded, but the objective function value is bounded, then one needs more technical arguments that we skip here.

Theorem 33 (Complementary slackness)**.** *Let* (x^*, y^*) *be feasible solutions for*

$$(P): \ \max\{c^T x \mid Ax \le b\} \quad \textit{and} \quad (D): \ \min\{by \mid y^T A = c; y \ge \mathbf{0}\}$$

Then (x^*, y^*) *are both optimal if and only if*

$$(A_i x^* = b_i \vee y_i^* = 0) \quad \forall i$$

Proof. We did already prove this in the last line of the duality theorem! □

5.2 Algorithms for linear programming

In this section, we want to briefly discuss the different methods that are known to solve linear programs.

5.2.1 The simplex method

The oldest and most popular one is the **simplex method**. Suppose the linear program is of the form

$$\max\{cx \mid Ax \le b\}.$$

We may assume that the underlying polyhedron $P = \{x \in \mathbb{R}^n \mid Ax \le b\}$ has vertices[2]. For a set of row indices $I \subseteq [m]$, let A_I be the submatrix of A that contains all the rows with index in I. A compact way of stating the simplex algorithm is as follows:

Simplex algorithm

Input: $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ and a starting basis $I \in \binom{[m]}{n}$

Output: opt. solution x attaining $\max\{cx \mid Ax \le b\}$

(1) $x = A_I^{-1} b_I$

(2) IF $y := cA_I^{-1} \ge \mathbf{0}$ THEN RETURN x is optimal

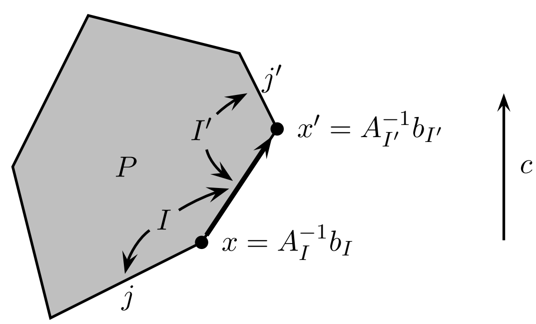
(3) select $j \in I$ and $j' \notin I$ so that for $I' := I \backslash \{j\} \cup \{j'\}$ the following 3 conditions are satisfied

  (i) $\text{rank}(A_{I'}) = n$

  (ii) the point $x' = A_{I'}^{-1} b_{I'}$ lies in P

  (iii) $cx' \ge cx$

(4) UPDATE $I := I'$ and GOTO (1)



one simplex step

The maintained set I of indices is usually called a **basis** and the maintained solution x is always a **vertex** of P. In other words, the simplex method moves from one vertex to the next one so that the

[2]This is equivalent to saying that the $\ker(A) = \{\mathbf{0}\}$. A simple way to obtain this property is to substitute a variable $x_i \in \mathbb{R}$ by $x_i = x_i^+ - x_i^-$ and adding the constraints $x_i^+, x_i^- \ge 0$ to the constraint system. Now the kernel of the constraint matrix is empty.