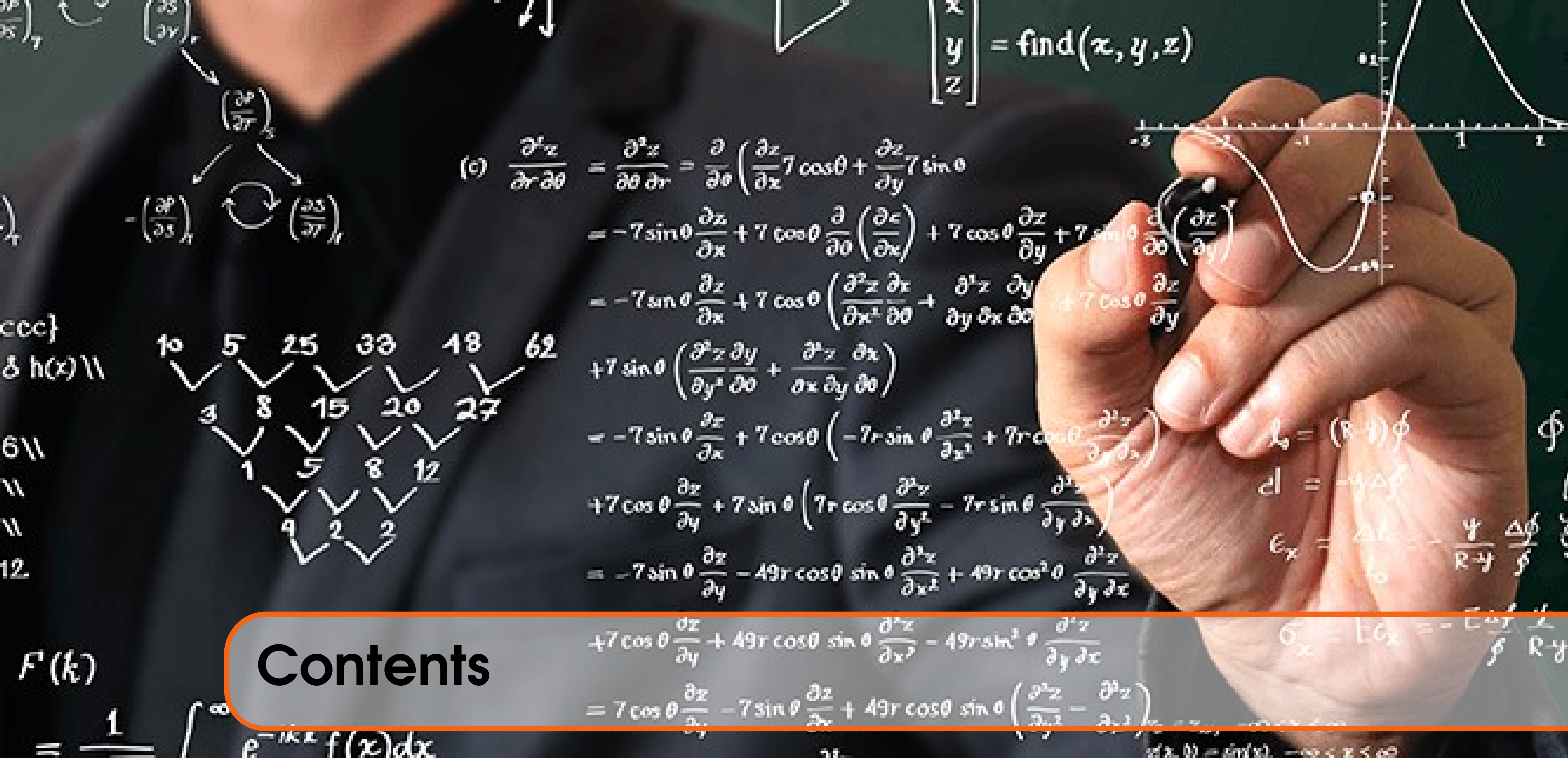
CO 255 Course Notes

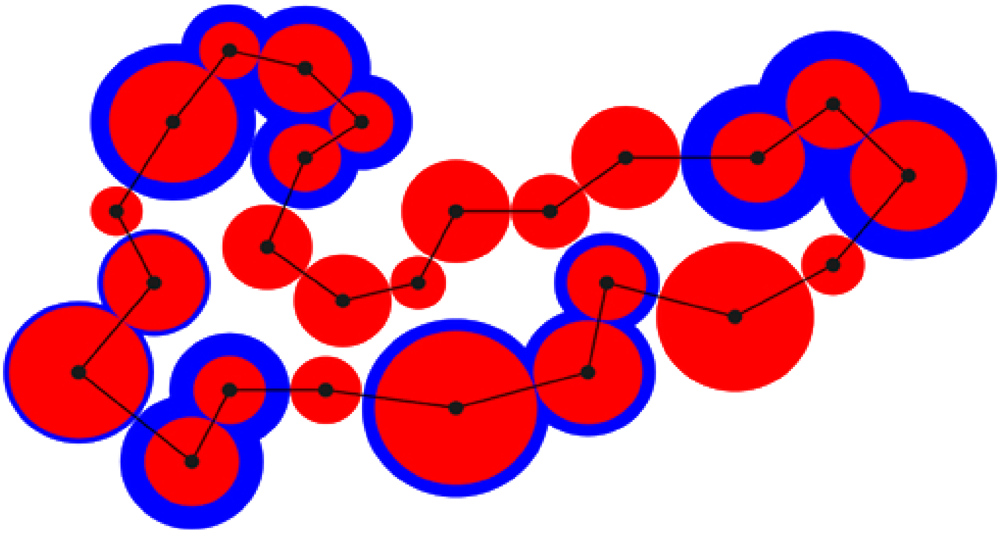
University of Waterloo

The One And Only
Waterloo 76er
Bill Zhuo



Contents

CO255 Main Contents



1. Introduction

1.1 Optimization Problem

Definition 1.1.1 — Optimization Problem. Given a set S, **the feasible region**, and a function $f : S \to \mathbb{R}$, **the objective function**, solve

$$\min\{f(x) : x \in S\}$$

or

$$\max\{f(x) : x \in S\}$$

R We note that

$$\max\{f(x) : x \in S\} = -\min\{-f(x) : x \in S\}$$

Thus, we shall focus our attention with minimization for most of the time.

The optimization problem itself might not be well-posed. For example,

1. **Infeasible:** $S = \emptyset$
2. **Unbounded:** there may exist $x \in S$ with $f(x)$ arbitrarily small

Even if the optimization is both feasible and bounded, it may not be well-posed. For example,

$$\min\{x : x > 1\}$$

cannot be attained since it has an infimum of 1 but it is outside the corresponding range. Wait, what's an infimum?

1.2 Infimum and Supremum

These are common definitions in analysis, also known as the greatest lower bound and least upper bound. Whenever we are working with a complete space, such as $\mathbb{R}$, we have the so-called **greatest lower bound/least upper bound property** in real analysis. We shall not get into it in a technical fashion.

Consider

$$\max\{z \in \mathbb{R} : z \leq f(x), \forall x \in S\}$$

. If the optimization problem is feasible and bounded, then the above maximization has an optimal solution.

Definition 1.2.1 — Infimum.

$$\inf\{f(x) : x \in S\} = \begin{cases} \infty & \text{infeasible} \\ -\infty & \text{unbounded} \\ \max\{z \in \mathbb{R} : z \leq f(x), \forall x \in S\} & \text{otherwise} \end{cases}$$

R The following requires some technical details but true.

$$\sup\{f(x) : x \in S\} = -\inf\{-f(x) : x \in S\}$$

Definition 1.2.2 — Optimal Value. We let $\mathbf{OPT}(1) = \inf\{f(x) : x \in S\}$ and $\mathbf{OPT}(2) = \sup f(x) : x \in S$.

1.3 Some Optimization Problems

Definition 1.3.1 — Linear Programming.

$$f(x) = C^\top \text{ and } S = \{x \in \mathbb{R}^n : Ax \leq b\}$$

where $A \in \mathbb{R}^{mn}, b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$.

Definition 1.3.2 — Integer Linear Programming.

$$f(x) = C^\top \text{ and } S = \{x \in \mathbb{Z}^n : Ax \leq b\}$$

where $A \in \mathbb{R}^{mn}, b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$.

Definition 1.3.3 — Convex Optimization. $S \subseteq \mathbb{R}^n$ is convex and $f : S \to \mathbb{R}$ is convex.

Definition 1.3.4 — Convex Set. $S \subseteq \mathbb{R}^n$ is convex if, for each $x, y \in S$ and $\lambda \in [0,1]$, we have

$$\lambda x + (1-\lambda)y \in S$$

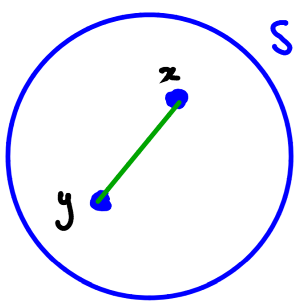


Figure 1.3.1: *Convex Set*

Definition 1.3.5 — Convex Function. $f : S \to \mathbb{R}$ is convex if for each $x, y \in S$ and $\lambda \in [0,1]$, we have

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

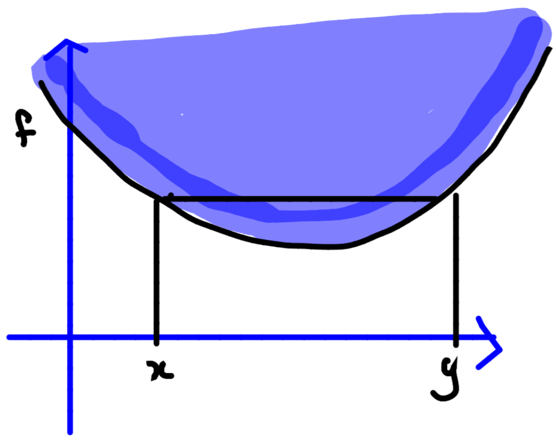


Figure 1.3.2: *Convex Function*

Definition 1.3.6 — Convex Hull. The convex hull of $S \subseteq \mathbb{R}^n$, denoted by $\mathbf{conv}(S)$, is the (unique) minimal convex set that contains S.

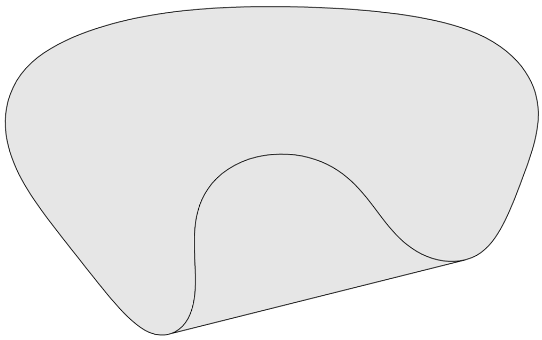


Figure 1.3.3: *Convex Hull*

R We may show its uniqueness later.

Consider an optimization problem $\min\{f(x) : x \in S\}$ where $S \subseteq \mathbb{R}^n$ and $f : \mathbb{R}^n \to \mathbb{R}$. We can "reduce" this to a convex optimization problem with a linear objective function.

1. **Step 1: Linearize the objective function**
   Let $\hat{S} = \left\{ \begin{bmatrix} x \\ y \end{bmatrix} : x \in S, y = f(x) \right\} \subseteq \mathbb{R}^{n+1}$. Then,

$$\min\{f(x) : x \in S\} = \min\left\{ y : \begin{bmatrix} x \\ y \end{bmatrix} \in \hat{S} \right\}$$

2. **Step 2: convexify S**
   If $f : \mathbb{R}^n \to \mathbb{R}$ is linear, then

$$\min\{f(x) : x \in S\} = \min\{f(x) : x \in \mathbf{conv}(S)\}$$

■ **Example 1.1 — A two-player game.** Given $A\mathbb{R}^{mn}$, Rose chooses $i \in \{1,\cdots,m\}$ and Colin chooses $j \in \{1,\cdots,n\}$ (independently), then Colin pays Rose a_{ij}. Say we have

$$A = \begin{bmatrix} 2 & -2 \\ 1 & 5 \end{bmatrix}$$

1. If Rose chooses 1, then her return is ≥ -2
2. If Rose chooses 2, then her return is ≥ 1
3. If Rose chooses with equal probability then her expected return is

$$\geq \min\left\{\frac{1}{2}\times 2+\frac{1}{2}, \frac{1}{2}\times -2+\frac{1}{2}\times 5\right\}$$

Rose will choose her strategy maximizing her expected return in the worst case that Colin actually knows her strategy. So the problem can be formulated to be

$$\begin{aligned} \max \quad & \min_{i\in\{1,\cdots,n\}} \sum_{j=1}^{m} p_i a_{ij} \\ s.t. \quad & p_1+\cdots+p_m = 1 \\ & p_1,\cdots,p_m \geq 0 \end{aligned}$$

or equivalently,

$$(R) \begin{aligned} \max \quad & z \\ s.t. \quad & z \leq \sum_{j=1}^{m} p_i a_{ij}, i \in \{1,\cdots,n\} \\ & p_1+\cdots+p_m = 1 \\ & p_1,\cdots,p_m \geq 0 \end{aligned}$$

We note that (R) is actually a LP. Likewise, Colin wants to choose his strategy by the following LP:

$$(C) \begin{aligned} \min \quad & z \\ s.t. \quad & z \geq \sum_{i=1}^{m} q_i a_{ij}, j \in \{1,\cdots,n\} \\ & q_1+\cdots+q_n = 1 \\ & q_1,\cdots,q_n \geq 0 \end{aligned}$$

Note that $\mathbf{OPT}(R) \leq \mathbf{OPT}(C)$. Surprisingly, (R) and (C) have the same optimal value. Hence, it does not harm either Rose or Colin to reveal their strategies. ■

■ **Example 1.2 — Weighted Bipartite Matching.** Given n jobs, n workers, and a utility a_{ij} of worker i performing job j, find an assignment of workers to jobs of maximum total utility.

1. **Variables:** $x_{ij} \in \{0,1\}$ where $x_{ij} = 1$ indicates assigning worker i to job j.
2. **Formulation:**

$$(P) \begin{aligned} \max \quad & \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}x_{ij} \\ s.t. \quad & \sum_{i=1}^{n} x_{ij} = 1, j \in \{1,\cdots,n\} \\ & \sum_{j=1}^{n} x_{ij} = 1, i \in \{1,\cdots,n\} \\ & x_{ij} \in \{0,1\}, i,j \in \{1,\cdots,n\} \end{aligned}$$

This is an integer linear program (IP). The **linear relaxation** is the linear program (P') obtained by replaing the last constraint with

$$0 \le x_{ij} \le 1, \quad i,j \in \{1,\cdots,n\}$$

Note that $\mathbf{OPT}(P) \le \mathbf{OPT}(P')$. But surprisingly, in this case, $\mathbf{OPT}(P) = \mathbf{OPT}(P')$. ■

■ **Example 1.3 — 3D-Matching.** Given $A \in \mathbb{R}^{n^3}$ where a_{ijk} is the utility of worker i performing job j on machine k, find an assignment of maximum total utility. Using similar variables in the 2D case, we have the following formulation:

$$\begin{aligned}
& \max && \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} a_{ijk}x_{ijk} \\
& s.t. && \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ijk} = 1, k \in \{1,\cdots,n\} \\
(P) & && \sum_{j=1}^{n}\sum_{k=1}^{n} x_{ijk} = 1, i \in \{1,\cdots,n\} \\
& && \sum_{i=1}^{n}\sum_{k=1}^{n} x_{ijk} = 1, j \in \{1,\cdots,n\} \\
& && x_{ijk} \in \{0,1\}, i,j,k \in \{1,\cdots,n\}
\end{aligned}$$

■

R 3D-matching problem is NP-hard! So, in general, IP is NP-hard. LP can be solved efficiently meaning in polynomial time.
We also note that, we can replace $z \in \{0,1\}$ with $z(z-1) = 0$. So, quadratic programming is also NP-hard in general.

■ **Example 1.4 — Integer Solutions to Diophantine Equations.** Consider the following formulation:

$$\begin{aligned}
& \min && \sin(\pi x)^2 + \sin(\pi y)^2 + \sin(\pi z)^2 \\
(P) & s.t. && x^3 + y^3 = z^3 \\
& && x,y,z \ge 1
\end{aligned}$$

Note that the optimal value is bounded by 0 and a feasible solution (x,y,z) has objective value zero if and only if x,y,z are positive integers satisfying $x^3 + y^3 = z^3$. This is actually the Fermat Last Theorem and we know there is no such triple. Thus, we cannot actually get 0 as the solution. ■

Definition 1.3.7 — Diophantine Equation. A Diophantine equation is an equation of the form

$$p(x_1,\cdots,x_n) = 0$$

where p is a polynomial with integer coefficients.

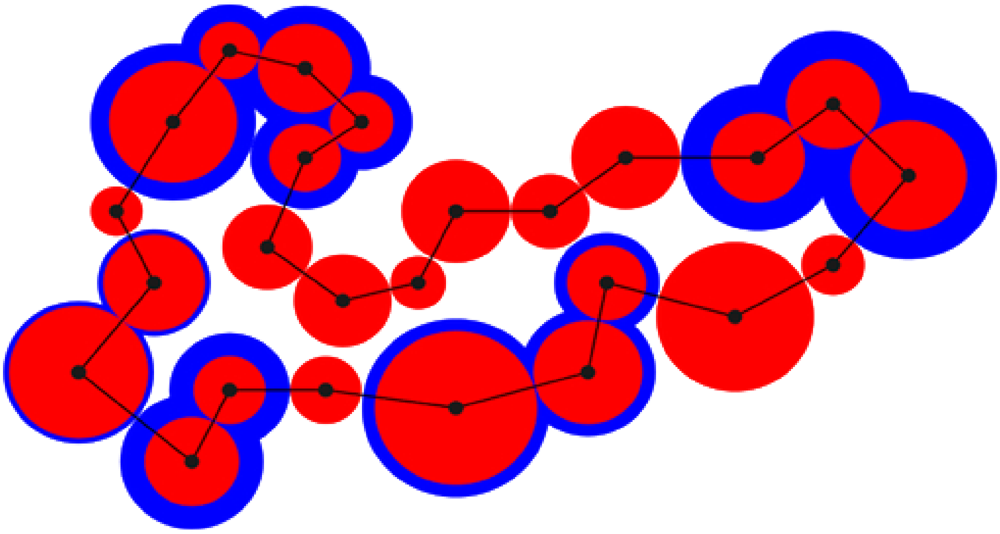
Exercise 1.1 — Hilbert's 10^{th} Problem. Given a Diophantine equation, decide whether it has an integer solution. We can formulate this into

$$(P)\begin{aligned}
& \min && \sin(\pi x)^2 + \sin(\pi y)^2 + \sin(\pi z)^2 \\
& s.t. && p(x_1,\cdots,x_n) = 0
\end{aligned}$$

■

R

1. Many famous problems are instances
   (a) The Four-Colour Theorem
   (b) Reimann Hypothesis
   (c) Goldbach's Conjecture
2. The problem is undecidable even for a variable polynomials.
3. Optimization is hard. We need restrictive assumptions to develop theory and algorithms, even for convex optimization.



2. Convex Basics

2.1 Geometry of Convex Sets

Lemma 2.2 If S is a collection of convex sets, then the intersection of all sets in S is convex.

Lemma 2.3 For any set $X \subseteq \mathbb{R}^n$, there is a unique minimal convex set containing X.

Proof. Take the intersection of all convex sets containing X. ■

Lemma 2.4 For $c \in \mathbb{R}^n$ and $S \subseteq \mathbb{R}^n$, the following problems have the same optimal values:

1. (1) $\min(c^\top x : x \in S)$
2. (2) $\min(c^\top x : x \in \mathbf{conv}(S))$

Proof. Since $S \subseteq \mathbf{conv}(S)$, $\mathbf{OPT}(2) \leq \mathbf{OPT}(1)$. Let $\alpha = \mathbf{OPT}(1)$ and let $H = \left\{x \in \mathbb{R}^n : c^\top x \geq \alpha\right\}$. By definition, $S \subseteq H$. Moreover, since H is convex, $\mathbf{conv}(S) \subseteq H$. It follows that $\mathbf{OPT}(2) \geq \alpha = \mathbf{OPT}(1)$. ■

Theorem 1 — Caratheodory's Theorem (Convex Hull). For $S \subseteq \mathbb{R}^n$ and $x \in \mathbf{conv}(S)$, there is a set $S_0 \subseteq S$ of at most $n+1$ points in S such that $x \in \mathbf{conv}(S_0)$.

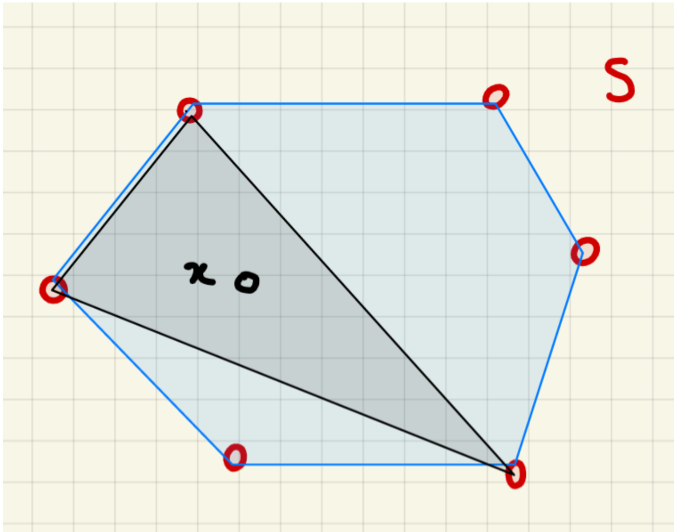


Figure 2.4.1: $2+1$ *points*

■ **Example 2.1 — $\mathbb{R}^2$ Case.** ■

2.5 Geomtry of Cones

Definition 2.5.1 — Cone. $K \subseteq \mathbb{R}^n$ is a cone if

1. $0 \in K$
2. K is convex
3. For each $x \in K$ and $\lambda \in \mathbb{R}_+$, we have $\lambda x \in K$. Here $\mathbb{R}_+ = (0, \infty)$.

For $S \subseteq \mathbb{R}^n$, we let $\mathbf{cone}(S)$ denote the smallest cone containing S.

Definition 2.5.2 — Convex Combination. Let S be a set. A convex combination is of the form

$$\sum_{i=1}^{n} \lambda_i s_i, s_i \in S$$

and $\sum_{i=1}^{n} \lambda_i = 1$ with $\lambda_i \geq 0$.

Theorem 2 — Equivalent Convex Hull Definitions. Let S be a set. Then,

$$\mathbf{conv}(S) = \{\text{all convex combinations generated by elements in } S\} =: V(S)$$

Proof.

Claim 1: $V(S)$ is a convex set containing S

It is clear that $S \subseteq V(S)$ since we can just take $\lambda = 1$ for any $s \in S$. Now, let $x, y \in S$, so

$$x = \sum_{i=1}^{n} \lambda_i s_i, s_i \in S, \qquad y = \sum_{j=1}^{m} \theta_j t_j, t_j \in S$$

where $\sum_{i=1}^{n}\lambda_i = \sum_{j=1}^{m}\theta_j = 1$. Then, for $M \in [0,1]$, we have

$$\begin{aligned} Mx + (1-M)y &= M\sum_{i=1}^{n}\lambda_i s_i + (1-M)\sum_{j=1}^{m}\theta_j t_j \\ &= \sum_{i=1}^{n} M\lambda_i s_i + \sum_{j=1}^{m}(1-M)\theta_j t_j \end{aligned}$$

since we know that $s_i, t_j \in S$, it suffices to check the sum of the coefficients. Then, we note that

$$\sum_{i=1}^{n} M\lambda_i + \sum_{j=1}^{m}(1-M)\theta_j = M\times 1 + (1-M)\times 1 = 1$$

Thus, the claim is true. Claim 1 implies that $\mathbf{conv}(S) \subseteq V(S)$ as the convex hull is originally defined to be the minimal convex set that contains S.

Claim 2: A convex set contains all its convex combinations

This requires an inductive proof on the number coefficients. It is clear that when we only have 1 or 2 coefficients in the convex combination, these are contained in the convex set due to definition of the convex set. Thus, we proceed with the inductive step. Say any convex combination of the form

$$v = \sum_{i=1}^{t}\lambda_i v_i, v_i \in V$$

is contained in V where V is the convex set here and $\sum_{i=1}^{t}\lambda_i = 1$. We want to use strong induction here, so say this is true for all $1 \le t \le n$. Then, consider

$$v' = \sum_{i=1}^{n+1}\lambda_i v_i, v_i \in V, \sum_{i=1}^{n+1}\lambda_i = 1$$

. We can rewrite this as

$$v' = \sum_{i=1}^{n}\lambda_i v_i + \lambda_{n+1}v_{n+1}$$

. We note that $\sum_{i=1}^{n}\lambda_i = 1-\lambda_{n+1} < 1 \implies \sum_{i=1}^{n}\frac{\lambda_i}{1-\lambda_{n+1}} = 1$ as we are in the $n+1$-step (if $\lambda_{n+1} = 0$, not so interesting). We apply this observation to v' to get

$$v' = (1-\lambda_{n+1})\underbrace{\sum_{i=1}^{n}\frac{\lambda_i}{1-\lambda_{n+1}}v_i}_{r} + \lambda_{n+1}v_{n+1}$$

. By our induction hypothesis, $r \in V$ since it is a convex combination. Apply the induction hypothesis again, $v' \in V$ as well. Thus, by induction, we have shown this claim is true.
Claim 2 implies that $V(S)$ is a subset of any convex set containing S including $\mathbf{conv}(S)$. Thus, $V(S) \subseteq \mathbf{conv}(S)$. And we have shown the theorem

$$V(S) = \mathbf{conv}(S)$$

■

Exercise 2.1 Let $S \subseteq \mathbb{R}^n$. For each vector $v \in \mathbb{R}^n$ we let v^+ denote the vector $\begin{bmatrix} 1 \\ v \end{bmatrix}$ in $\mathbb{R}^{n+1}$, and let $S^+ := \{x^+ : x \in S\}$. For any vector $v \in \mathbb{R}^n$, prove that $v \in \text{conv}(S)$ if and only if $v^+ \in \text{cone}(S^+)$.

1. Suppose $v \in \textbf{conv}(S)$. By our equivalent definition above, we can write it as
$$v = \sum_{i=1}^{n} \lambda_i s_i, s_i \in S, \sum_{i=1}^{n} \lambda_i = 1$$
then we have
$$v^+ = \begin{bmatrix} 1 \\ \sum_{i=1}^{n} \lambda_i s_i \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \begin{bmatrix} 1 \\ s_i \end{bmatrix} = \sum_{i=1}^{n} \lambda_i s_i^+$$
where $s^+ \in S^+$. We note that $\textbf{cone}(S^+)$ contains S^+ and it is convex, which means it should contain all of its convex combinations. Thus, $v^+ \in \textbf{cone}(S^+)$.
2. Suppose $v^+ \in \textbf{cone}(S^+)$. By Problem 4, we know that there exists finite $T^+ \subseteq S^+$ such that $v^+ \in \textbf{cone}(T^+)$. Say
$$T^+ = \{s_1^+, \cdots, s_n^+\}$$
Then, by Problem 3, there exists non-negative real numbers $\lambda_1, \cdots, \lambda_n$ such that
$$\begin{bmatrix} 1 \\ v \end{bmatrix} = v^+ = \sum_{i=1}^{n} \lambda_i s_i^+ = \sum_{i=1}^{n} \lambda_i \begin{bmatrix} 1 \\ s_i \end{bmatrix}, s_i \in S$$
to make this equation work, we have
$$\sum_{i=1}^{n} \lambda_i = 1$$
. This implies that v is a convex combination of s_i. Thus, $v \in \textbf{conv}(S)$. And we are done.

■

Definition 2.5.3 — Conic Combination. Let S be a set. A conic combination is of the form
$$\sum_{i=1}^{n} \lambda_i s_i, s_i \in S$$
and $\lambda_i \geq 0, \forall i$.

Theorem 3 — Equivalent Cone Definition. Let S be a set. Then,
$$\textbf{cone}(S) = \{\text{all conic combinations generated by elements in } S\} =: C(S)$$

Proof. **Claim 1: $C(S)$ is a cone containing S**

It is clear that $S \subseteq C(S)$ since we can just take $\lambda = 1$ for any $s \in S$.

1. $0 \in C(S)$ since we can let all $\lambda_i = 0$
2. Let $a \in \geq 0$ and let $x \in C(S)$. Say
$$x = \sum_{i=1}^{n} \lambda_i s_i$$

for $\lambda_i \geq 0, \forall i$. Then,

$$ax = \sum_{i=1}^{n} a\lambda_i s_i = \sum_{i=1}^{n} \theta_i s_i$$

where $\theta_i = a\lambda_i \geq 0, \forall i$. Thus, $ax \in C(S)$.

Claim 1 is true. This implies that $\mathbf{cone}(S) \subseteq C(S)$ since $C(S)$ might include more than $\mathbf{cone}(S)$ has.

Claim 2: A cone contains all of its conic combinations

We prove this by induction as well.

1. **Base case:** The case for $n = 1$ is directly from the definition of a cone. We shall show the $n = 2$ case as the base case. Let $x, y \in \mathbf{cone}(S)$ and $a, b > 0$ (otherwise, we are back in the $n = 1$ case). Then, since a cone is convex, we have

$$\left(\frac{a}{a+b}\right)x + \left(\frac{b}{a+b}\right)y \in \mathbf{cone}(S)$$

Then, by the property of a cone, we have

$$(a+b)\left[\left(\frac{a}{a+b}\right)x + \left(\frac{b}{a+b}\right)y\right] = ax + by \in \mathbf{cone}(S)$$

2. **Inductive Step:** suppose the statement is true for all $1 \leq i \leq n$. We have

$$\sum_{i=1}^{n+1} \lambda_i x_i = \sum_{i=1}^{n} \lambda_i x_i + \lambda_{n+1} x_{n+1}, \lambda_i > 0, x_i \in \mathbf{cone}(S)$$

By the induction hypothesis and the base case, we have

$$\underbrace{\sum_{i=1}^{n} \lambda_i x_i}_{\in \mathbf{cone}(S)} + \lambda_{n+1} x_{n+1} \in \mathbf{cone}(S)$$

Thus, by induction, our claim 2 is true. This implies that $C(S) \subseteq \mathbf{cone}(S)$ since all the conic combinations of S is certainly contained in all the conic combinations of $\mathbf{cone}(S)$. By claim 1 and 2, we have shown the equivalent definitions. ■

Exercise 2.2 Let $S_1 \subseteq \mathbb{R}^n$ be a finite set. Prove that, for each $x \in \mathbf{cone}(S_1)$, there is a linearly independent subset S_0 of S_1 such that $x \in \mathbf{cone}(S_0)$.

Proof. By Problem 3, we know that if $x \in \mathbf{cone}(S_1)$ with $S_1 = \{s_1, \cdots, s_n\}$, we can write it as

$$x = \sum_{i=1}^{n} \lambda_i s_i, \lambda_i \geq 0$$

. We shall assume $x \neq 0$, since every cone contains 0. say,

$$x = \sum_{i=1}^{m} \lambda_i s_i$$

is the minimal (in the sense of number of s_i used) conic combination with $\lambda_i > 0$. We claim $s_1, \cdots, s_m$ should be linearly independent. For the sake of contradiction, suppose they are not, then by flipping the sign if necessary, we have

$$\sum_{i=1}^{m} \theta_i s_i = 0$$

where at least one of the $\theta_i > 0$. There can be some negative θ_i. If they are all negative, we will just multiply (-1) to both sides. Without loss of generality, say $\theta_1, \cdots, \theta_k > 0$ with $1 \leq k \leq m$. Then, pick $M_i = \frac{\lambda_i}{\theta_i}, \forall 1 \leq i \leq k$. Then, pick

$$M = \min_{1 \leq i \leq k} \{M_i\}$$

Then, note that

$$\sum_{i=1}^{m} (\lambda_i - M\theta_i) s_i = \sum_{i=1}^{m} \lambda_i s_i - M \sum_{i=1}^{m} \theta_i s_i = x$$

and $\lambda_i - M\theta_i = 0$ for at least one index between $1 \leq i \leq m$ while all the others are positive. This contradicts the minimality construction of x in the first place. Thus, our original $S_0 = \{s_1, \cdots, s_m\}$ is linearly independent and, by Problem 3, $x \in \textbf{cone}(S_0)$. ■

The exercise above gives us

Theorem 4 — Caratheodory's Theorem for Cones. For $S \subseteq \mathbb{R}^n$ and $x \in \textbf{cone}(S)$, there is a set $S_0 \subseteq S$ of linearly independent vectors in S such that $x \in \textbf{cone}(S_0)$.

Proof. Put everything above together. ■

2.6 Convex Sets from Convex Functions

Definition 2.6.1 — Convex Function. $f : S \to \mathbb{R}$ is convex if for each $x, y \in S$ and $\lambda \in [0,1]$, we have

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

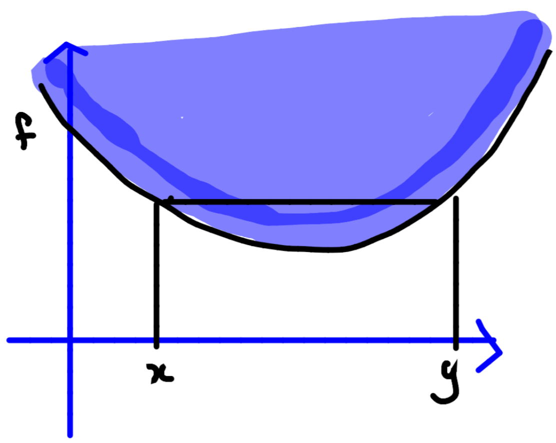


Figure 2.6.1: *Convex Function*

Definition 2.6.2 — Level Sets. If $f : S_0 \to \mathbb{R}$ is a convex function and $\alpha \in \mathbb{R}$, then the set $S = \{x \in S_0 : f(x) \le \alpha\}$ is convex. We call this S a level set of S.

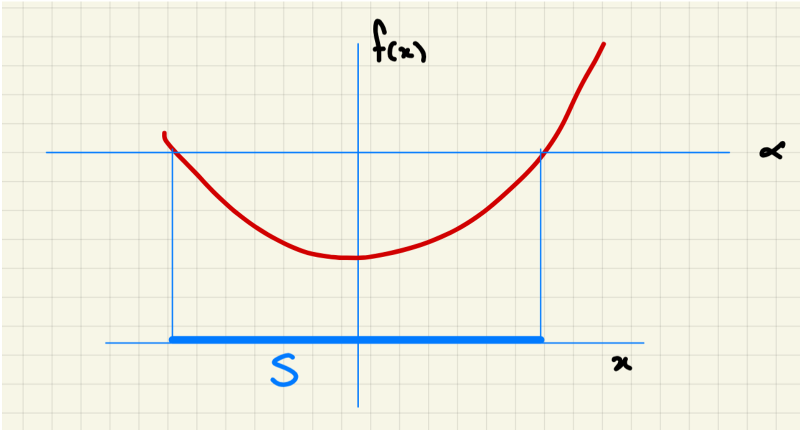


Figure 2.6.2: *Level Set*

R Level sets of non-convex functions can be convex. For example, $f(x) = \sqrt{|x|}$.

Definition 2.6.3 — Epigraphs of Convex Functions. The epigraph of a function $f : S_0 \to \mathbb{R}$, where $S_0 \subseteq \mathbb{R}^n$, is the subset of $\mathbb{R}^{n+1}$ defined by

$$\textbf{epi}(f) := \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{n+1} : z \ge f(x), x \in S_0 \right\}$$

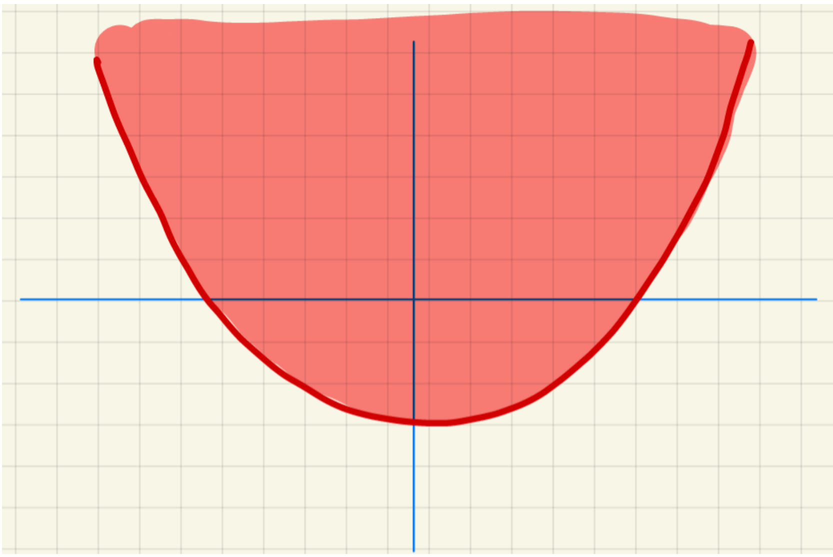


Figure 2.6.3: *Epigraph of f*

Lemma 2.7 $f : S_0 \to \mathbb{R}$ is convex if and only if $\textbf{epi}(f)$ is convex.

Proof. 1. Suppose f is convex so S_0 is a convex set. Let $\begin{bmatrix} x_1 \\ z_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ z_2 \end{bmatrix} \in \textbf{epi}(f)$ where $f(z_i) \ge$

$x_i \in S_0$ and let $\lambda \in [0,1]$. We check

$$\lambda \begin{bmatrix} x_1 \\ z_1 \end{bmatrix} + (1-\lambda) \begin{bmatrix} x_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} \lambda \lambda x_1 + (1-\lambda)x_2 \\ \lambda z_1 + (1-\lambda) z_2 \end{bmatrix} \in \mathbf{epi}(f)$$

where $\lambda x_1 + (1-\lambda)x_2 \in S_0$. And

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda z_1 + (1-\lambda) z_2$$

2. The converse is left as an exercise.

■

■ **Example 2.2 — AM-GM Inequality.** For $x_1, \cdots, x_n > 0$ in $\mathbb{R}$,

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$$

Proof. Note that

$$\frac{cx_1 + \cdots + cx_n}{n} = c\left(\frac{x_1 + \cdots + x_n}{n}\right)$$

and

$$\sqrt[n]{(cx_1)\cdots(cx_n)} = c\sqrt[n]{x_1 \cdots x_n}$$

so, up to scaling, we assume that $\sqrt[n]{x_1 \cdots x_n} = 1$. Consider the problem

$$\begin{array}{lll} & \min & \frac{1}{n}x_1 + \cdots + \frac{1}{n}x_n \\ (P) & s.t. & \sqrt[n]{x_1 \cdots x_n} \geq 1 \\ & & x_i > 0, i \in \{1, \cdots, n\} \end{array}$$

To prove AM-GM inequality, it suffices to prove that $\mathbf{OPT}(P) \geq 1$, or equivalently, $x = [1 \cdots 1]^\top$ is optimal. Well, the current (P) is not a nice program. We note that

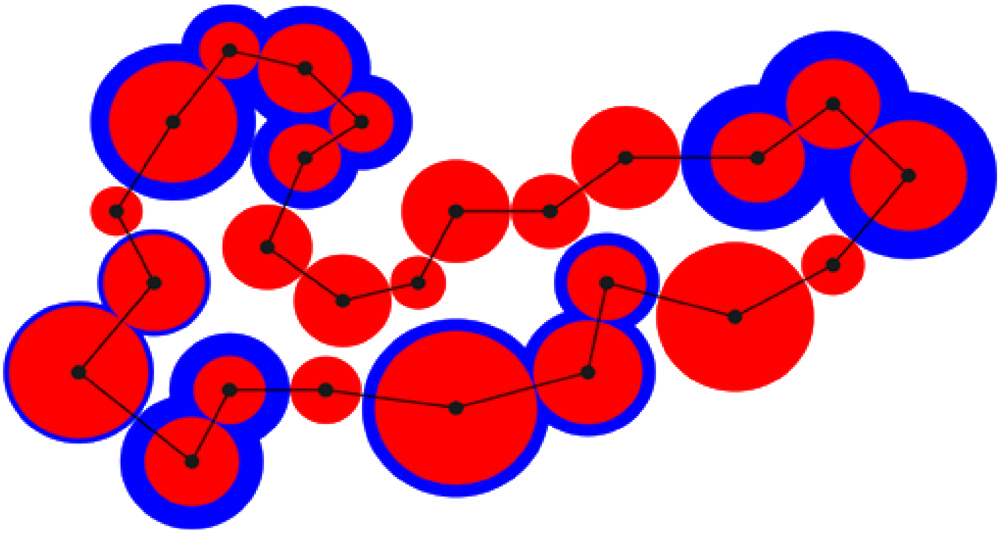
$$\sqrt[n]{x_1 \cdots x_n} \geq 1 \iff x_1 \cdots x_n \geq 1 \iff -\ln(x_1) - \cdots - \ln(x_n) \leq 0$$

and here $f(x) = -\ln(x)$ is convex on $S_0 = \{x > 0 : x \in \mathbb{R}\}$. Thus, we can formulate the problem into a convex optimization problem.

$$(\tilde{P}) \quad \min\left(c^\top x : f(x) \leq 0, x \in S_0\right)$$

where $c = \left[\frac{1}{n} \cdots \frac{1}{n}\right]^\top$ and $S_0 = \{x \in \mathbb{R}^n : x > 0\}$ and for $x \in S_0$, $f(x) = -\ln(x_1) - \cdots - \ln(x_n)$. The feasible region of $(\tilde{P})$ is the level set of a convex function. ■

■



3. Linear Programming

3.1 Feasibility of Linear Program

Given $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, does there exist $x \in \mathbb{R}^n$ such that $Ax \geq b$?

■ **Example 3.1 — Fourier-Motzkin Elimination.** Consider the following system of linear inequalities

$$(1)\begin{cases} x_2 \leq 2 \\ x_1 - x_2 \geq -1 \\ 2x_1 + x_2 \geq 2 \\ x_1 - 3x_2 \leq 1 \end{cases}$$

we rewrite it as

$$\begin{aligned} &x_2 \leq 2 && \textbf{does not use } x_1 \\ &x_1 \geq x_2 - 1 && \textbf{lower bounds on } x_1 \\ &x_1 \geq -\frac{1}{2}x_2 + 1 && \\ &x_1 \leq 3x_2 + 1 && \textbf{upper bound on } x_1 \end{aligned}$$

which is equivalent to

$$(1')\begin{cases} x_2 \leq 2 \\ \max(x_2 - 1, -\frac{1}{2}x_2 + 1) \leq x_1 \leq 3x_2 + 1 \end{cases}$$

we now eliminate x_1 to get

$$(2)\begin{cases} x_2 \leq 2 \\ x_2 - 1 \leq 3x_2 + 1 \\ -\frac{1}{2}x_2 + 1 \leq 3x_2 + 1 \end{cases}$$

we note that x_2 satisfies (2) if and only if there exists $x_1 \in \mathbb{R}$ such that (x_1, x_2) satisfies (1). ■

■ **Example 3.2**

$$(1)\begin{cases} x_1 & +3x_2 & & -x_4 & \geq & 6 \\ -2x_1 & & +3x_3 & +x_4 & \geq & 4 \\ -x_1 & +2x_2 & +x_3 & & \geq & 3 \\ 2x_1 & +5x_2 & -x_3 & & \geq & 4 \end{cases}$$

isolate x_1,

$$(1')\begin{cases} x_1 & \geq & 6 & -3x_2 & +x_4 \\ x_1 & \leq & -4 & +\frac{3}{2}x_3 & +\frac{1}{2}x_4 \\ x_1 & \leq & -3 & +2x_2 & +x_3 \\ x_1 & \geq & 2 & -5x_2 & +x_3 \end{cases}$$

eliminate x_1,

$$(2)\begin{cases} 6-3x_2+x_4 \leq -4+\frac{3}{2}x_3+\frac{1}{2}x_4 \\ 6-3x_2+x_4 \leq -3+2x_2+x_3 \\ 2-5x_2+x_3 \leq -4+\frac{3}{2}x_3+\frac{1}{2}x_4 \\ 2-5x_2+x_3 \leq -3+2x_2+x_3 \end{cases}$$

■

How fast is Fourier-Motzkin Elimination?

Given a system of m inequalities, if we eliminate one variable we get $O(m^2)$ inequalities. After eliminating k variables, we get $O(m^{2^k})$ inequalities. This is not efficient.

3.2 Polyhedra

Definition 3.2.1 — Polyhedron. A polyhedron is a set of the form $\{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$.

Definition 3.2.2 — Polytope. A polytope is a bounded polyhedron.

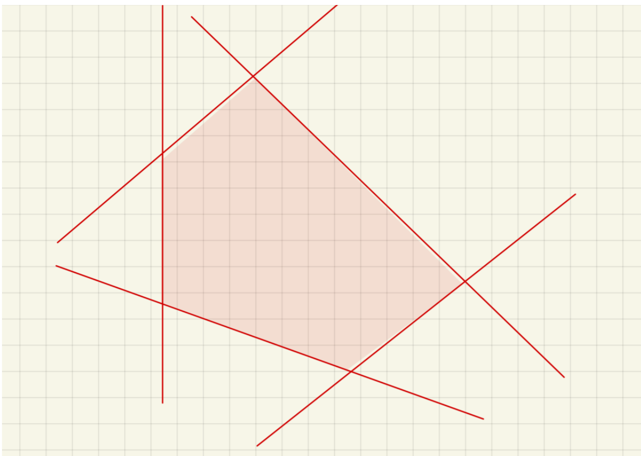


Figure 3.2.1: *Polytope*

Definition 3.2.3 — Projection. Let $P \subseteq \mathbb{R}^n$ and let $l < n$, and

$$P' = \left\{ [x_1 \cdots x_l]^\top : x \in P \right\}$$

we call P' the projection of P onto $x_1, \cdots, x_l$.

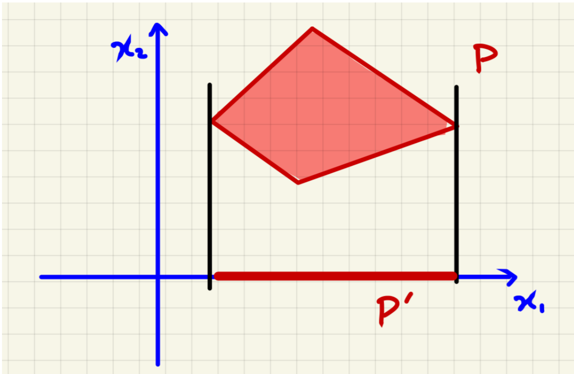


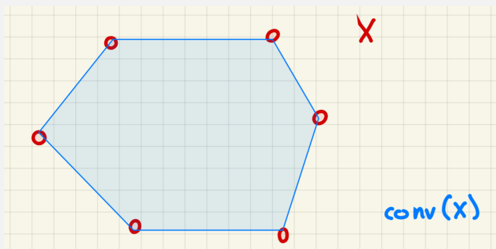
Figure 3.2.2: *P' is the projection of P onto x_1*

■ **Example 3.3** ■

Theorem 5 If $P \subseteq \mathbb{R}^n$ is a polyhedron and P' is the projection of P onto $x_1, \cdots, x_l$, then P' is a polyhedron.

Proof. This follows from Fourier-Motzkin elimination. ■

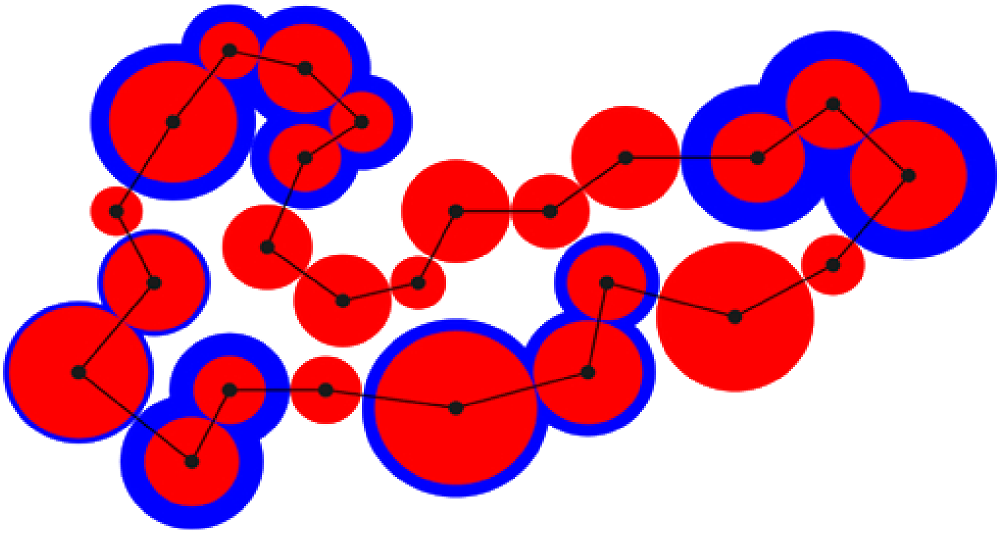
Corollary 3.2.1 If $X \subseteq \mathbb{R}^n$ is a finite set, then $\mathbf{conv}(X)$ is a polytope.



Proof. Since X is finite, $\mathbf{conv}(X)$ is bounded, so it remains to show that $\mathbf{conv}(X)$ is a polyhedron. Let $X = \{a_1, \cdots, a_m\}$ and $A = [a_1 \cdots a_m]$. Recall that $x \in \mathbf{conv}(X)$ if and only if there exists $\lambda \in \mathbb{R}^m$ such that $x = A\lambda$ with $\lambda \geq 0$ and $\lambda_1 + \cdots + \lambda_m = 1$. Let

$$P^+ = \left\{ \begin{bmatrix} x \\ \lambda \end{bmatrix} \in \mathbb{R}^{m+n} : x = A\lambda, \lambda \geq 0, \lambda_1 + \cdots + \lambda_m = 1 \right\}$$

Note that P^+ is a polyhedron and $\mathbf{conv}(X)$ is the projection of P^+ onto x, so $\mathbf{conv}(X)$ is a polyhedron. ■



4. Magic of Polyhedra

4.1 Certificate of Infeasibility

Recall that (do not really recal tbh)

Theorem 6 — Fundamental Theorem of Linear Algebra. Let $\mathbb{F}$ be a field. For $A \in \mathbb{F}^{m\times n}$ and $b \in \mathbb{F}^m$ exactly one of the following systems has a solution:
1. (1) $(Ax = b, x \in \mathbb{R}^n)$
2. (2) $(y^\top A = 0, y^\top b = 1, y \in \mathbb{R}^m)$

R That is, if $Ax = b$ is infeasible, then we can obtain the equation $0 = 1$ by taking a linear combination of the rows.

Definition 4.1.1 — Implied Inequalities. A linear inequality $a^\top x \geq a_0$ is implied by a system $Ax \geq b$ if there exists a non-negative vector $y \in \mathbb{R}^m$ such that $a = A^\top y$ and $a_0 = y^\top b$.

R This definition is non-standard. The more standard definition has $a_0 \leq y^\top b$.

Theorem 7 — Farkas' Lemma. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Exactly one of the following systems has a solution:
1. (1) $(Ax \geq b, x \in \mathbb{R}^n)$
2. (2) $(y^\top A = 0, y^\top b = 1, y \geq 0, y \in \mathbb{R}^m)$

Proof. **Claim 1:(1) and (2) cannot both hold**

Proof. If (1) and (2) hold, then

$$0 = (y^\top A)x = y^\top (Ax) \geq y^\top b = 1$$

yields a contradiction. ■

Note that (2) is equivalent to (2') that $0 \geq 1$ is implied by $Ax \geq b$. ■

Claim 2: If $A'x \geq b'$ is a set of inequalities implied by $Ax \geq b$ and $A''x \geq b''$ is a set of inequalities implied by $A'x \geq b'$, then $A''x \geq b''$ is implied by $Ax \geq b$

Proof. A non-negative combination of non-negative numbers is non-negative. ■

Claim 3: The system obtained from $Ax \geq b$ by Fourier-Motzkin Elimination is implied

Proof. Well, "proof by example" (ewwwwwww). Consider the inequalities:

$$\begin{aligned} 2x_1 + 2x_2 + x_3 &\geq 6 \quad (1.1) \\ -3x_1 + 6x_2 + 2x_3 &\geq 8 \quad (1.2) \end{aligned}$$

rewrite as

$$\begin{aligned} x_1 + x_2 + \frac{1}{2}x_3 &\geq 3 \quad \frac{1}{2}(1.1) \\ -x_1 + 2x_2 + \frac{2}{3}x_3 &\geq \frac{8}{3} \quad \frac{1}{3}(1.2) \end{aligned}$$

adding these gives

$$3x_2 + \frac{7}{6}x_3 \geq 5\frac{2}{3} \qquad \frac{1}{2}(1.1) + \frac{1}{3}(1.2)$$

Use an "easy" induction argument using Claim 2 and 3 to show if $Ax \geq b$ is infeasible, then $0 \geq 1$ is implied. ■

R That is, if $Ax \geq b$ has no solution, then we can obtain the inequality $0 \geq 1$ as a non-negative combination of the constraints.

■ **Example 4.1**

$$\begin{aligned} x + 2y &\leq 2 \quad (1) \\ x - y &\geq 0 \quad (2) \\ 3x + 2y &\leq 6 \quad (3) \\ y &\geq 1 \quad (4) \end{aligned}$$

rewrite as

$$\begin{aligned} y &\geq 1 \qquad (4) \\ x - y &\geq 0 \qquad (2) \\ -x - 2y &\geq -2 \quad -(1) \\ -x - \frac{2}{3}y &\geq -2 \quad -\frac{1}{3}(3) \end{aligned}$$

eliminate x and rewrite as

$$\begin{aligned} y &\geq 1 \qquad (4) \\ -y &\geq -\frac{2}{3} \quad \frac{1}{3}(2) - \frac{1}{3}(1) \\ -y &\geq -\frac{3}{2} \quad \frac{3}{8}(2) - \frac{1}{8}(3) \end{aligned}$$

Eliminate y,

$$\begin{aligned} 0 &\geq \frac{1}{3} \qquad (4) + \frac{1}{3}(2) - \frac{1}{3}(1) \\ 0 &\geq -\frac{1}{2} \quad (4) + \frac{3}{8}(2) - \frac{1}{8}(3) \end{aligned}$$

The system (1)(2)(3)(4) is infeasible as we get $0 \geq 1$ from the combination $3(4)+(2)-(1)$. ■

Corollary 4.1.1 — Another form of the Farkas Lemma. Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. Then exactly one of the following systems has a solution:

1. (1) $(Ax = b, x \geq 0)$
2. (2) $(y^\top A \geq 0, y^\top b = -1)$

Proof.

1. **Easy Part:** if x satisfies (1) and y satisfies (2), then

$$0 \leq (y^\top A)x = y^\top (Ax) = y^\top b = -1$$

yields a contradiction. Thus, we cannot have both (1) and (2) hold.

2. **Hard Part:** we will prove that if (1) has no solution, then (2) has a solution. We can rewrite (1) as

$$(Ax \geq b, Ax \leq b, x \geq 0)$$

or equivalently

$$(1') \quad \begin{bmatrix} A \\ -A \\ I \end{bmatrix} x \geq \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix}$$

If (1') has no solution, then, by the Farkas Lemma, exist non-negative vectors $y_1, y_2 \in \mathbb{R}^m$ and $z \in \mathbb{R}^n$ such that

$$\begin{bmatrix} y_1^\top & y_2^\top & z^\top \end{bmatrix} \begin{bmatrix} A \\ -A \\ I \end{bmatrix} = 0 \qquad \begin{bmatrix} y_1^\top & y_2^\top & z^\top \end{bmatrix} \begin{bmatrix} b \\ -b \\ 0 \end{bmatrix} = 1$$

That is

$$y_1^\top A - y_2^\top A + z = 0$$

and

$$y_1^\top b - y_2^\top b = 1$$

we can rewrite this as

$$(y_2 - y_1)^\top A \geq 0$$
$$(y_2 - y_1)^\top b = -1$$

Letting $y = y_2 - y_1$ gives a solution to (2) as required.

■

R

Geometric Interpretation of Farkas Lemma

Suppose that $A = [a_1 \cdots a_n]$. The following are equivalent:

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

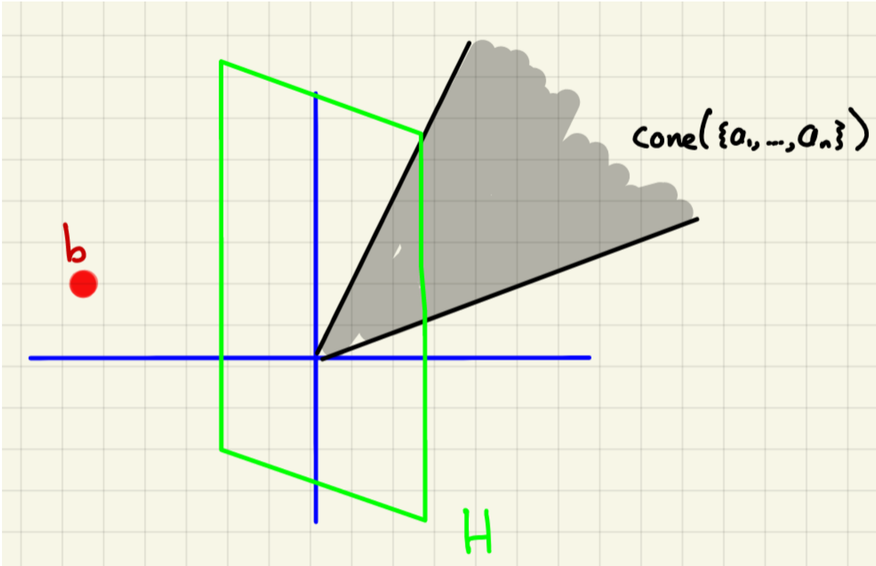
$$Ax = b, x \geq 0$$

2. (1') $b \in \mathbf{cone}(\{a_1, \cdots, a_n\})$

Proof. Consider a solution to (2) $(y^\top A \geq 0, y^\top b = -1)$. That is

$$y^\top a_1 \geq 0, \cdots, y^\top a_n \geq 0$$

but $y^\top b = -1$. Let $H = \{x \in \mathbb{R}^n : y^\top x \geq 0\}$. Note that H is a closed half-space containing $\{a_1, \cdots, a_n\}$ but not b.



■

Thus, we can state the alternative form of the Farkas Lemma as

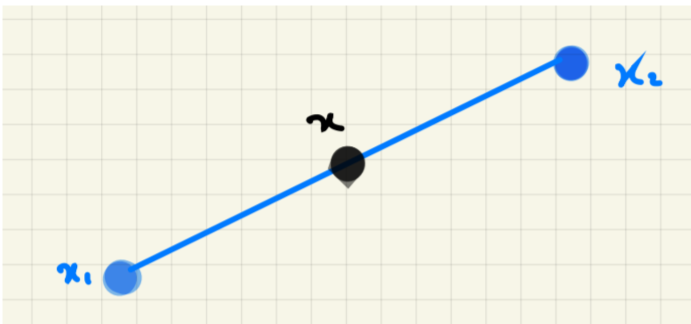
Corollary 4.1.2 Let $a_1, \cdots, a_n \in \mathbb{R}^M$. Then exactly one of the following is true:

1. (1) $b \in \textbf{cone}(\{a_1, \cdots, a_n\})$
2. (2) There is a closed half-space containing $a_1, \cdots, a_n$ but not b.

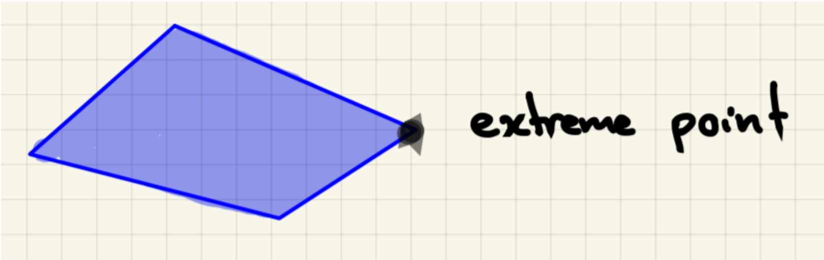
4.2 Geometry of Polyhedra

The points that we are going to take are the extreme points.

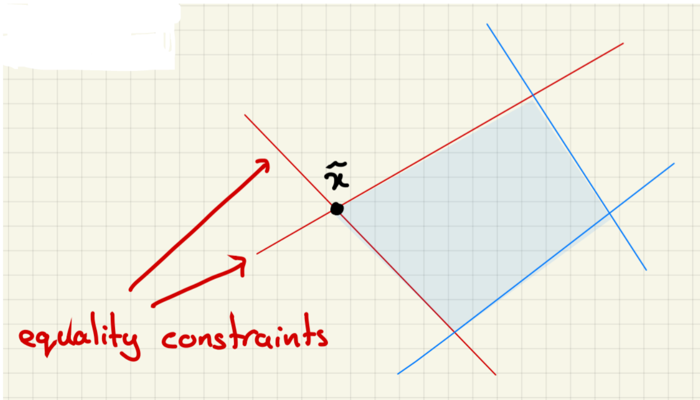
Definition 4.2.1 — Extreme Points. Let $S \subseteq \mathbb{R}^n$ be a convex set and $x \in S$. We call $x \in S$ on an extreme point of S if there are no two distinct points x_1, x_2 in S such that $x \in \{\lambda x_1 + (1-\lambda)x_2 : 0 < \lambda < 1\}$.



Equivalently, $S \backslash \{x\}$ is convex



Definition 4.2.2 — The Equality Subsystem. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ and let $\hat{x} \in P$. The equality subsystem for $\hat{x}$ is the subsystem of $Ax \geq b$ which $\hat{x}$ satisfies with equality.



Theorem 8 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, let $\hat{x} \in P$, and let $A^= x \geq b^=$ be the equality subsystem. If $\operatorname{rank}(A^=) = n$, then $\hat{x}$ is an extreme point.

Proof. Suppose that $\hat{x}$ is not an extreme point, so there exists distinct $x_1, x_2 \in P$ and $\lambda \in (0,1)$ such that $\hat{x} = \lambda x_1 + (1-\lambda)x_2$. Then, note that

$$b^= = A^=\hat{x} = \lambda A^= x_1 + (1-\lambda)A^= x_2 \geq \lambda b^= + (1-\lambda)b^= = b^=$$

Therefore, $A^= x_1 = b^=$ and $A^= x_2 = b^=$. Thus, $A^=(x_2 - x_1) = 0$. And since $x_2 - x_1 \neq 0$, we have $\operatorname{rank}(A^=) < n$. ■

Theorem 9 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, let $\hat{x} \in P$, and let $A^= x \geq b^=$ be the equaility subsystem. If $\operatorname{rank}(A^=) < n$, then $\hat{x}$ is not an extreme point.

Proof. If $\operatorname{rank}(A^=) < n$, then there is a non-zero vector $d \in \mathbb{R}^n$ such that $A^= d = 0$. Now $A^=(\hat{x} + \lambda d) = b^=$ for all $\lambda \in \mathbb{R}$. Then, there exists $\varepsilon > 0$ such that $\hat{x} + \varepsilon d, \hat{-}\varepsilon d \in P$ and hence $\hat{x}$ is not an extreme point. ■

Corollary 4.2.1 — Characterization of Extreme Points. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, let $\hat{x} \in P$, and let $A^= x \geq b^=$ be the equaility subsystem. $\operatorname{rank}(A^=) = n$ if and only if $\hat{x}$ is an extreme point.

Theorem 10 A set $P \subseteq \mathbb{R}^n$ is a polytope if and only if $P = \mathbf{conv}(X)$ for some finite set $X \subseteq \mathbb{R}^n$.

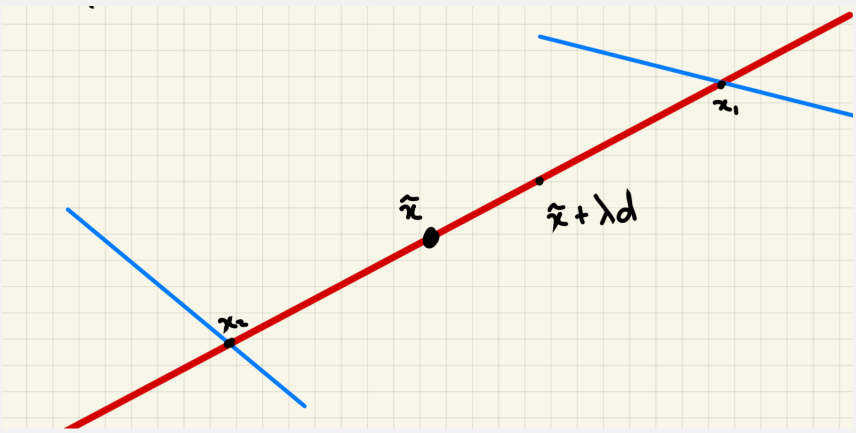
Proof. We will prove:

1. (1) Polyhedra have only finitely many extreme points

   Proof. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Now consider an extreme point $\hat{x}$ and its associated equality subsystem $A^= x \geq b^=$. By the theorem, $\mathrm{rank}(A^=) = n$. Therefore, $\hat{x}$ is the unique solution to $A^= x = b^=$. There are only 2^m subsystems of $Ax \geq b$, so there are at most 2^m extreme points. ■

2. (2) Every polytope is the convex hull of its extreme points

   Proof. Let X be the set of extreme points of a polytope $P = \{x \in \mathbb{R}^n : Ax \geq b\}$. Clearly, $\mathbf{conv}(X) \subseteq P$. Assume by way of contradiction that there exists $\hat{x} \in P \backslash \mathbf{conv}(X)$ and choose such that $\hat{x}$ whose equality subsystem $A^= x \geq b^=$ contains as many constraints as possible. Since $\hat{x} \notin \mathbf{conv}(X)$, $\hat{x}$ is not an extreme point and hence $\mathrm{rank}(A^=) < n$. So, there exists a non-zero vector d such that $A^= d = 0$.

   Let $\lambda^+ = \max(\lambda \in \mathbb{R} : \hat{x} + \lambda d \in P)$ and $\lambda^- = \min(\lambda \in \mathbb{R} : \hat{x} + \lambda d \in P)$. Note that these exist since P is closed and bounded. Now, let $x_1 = \hat{x} + \lambda^+ d$ and $x_2 = \hat{x} + \lambda^- d$. Since $A^= d = 0$, we have $A^= x_1 = b^=$ and $A^= x_2 = b^=$. Therefore, by our choice of λ^- and λ^+, we have $\lambda^- < 0 < \lambda^+$ and both x_1 and x_2 have more equality constraints than $\hat{x}$ does. By our choice of $\hat{x}$, we have $x_1, x_2 \in \mathbf{conv}(X)$. But $\hat{x}$ is on the line segment of x_1, x_2, so $\hat{x} \in \mathbf{conv}(X)$ as well. ■

■

4.2.1 Supporting Hyperplanes

Definition 4.2.3 — Hyperplane. A hyperplane of $\mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : a^\top x = a_0\}$ where $a \in \mathbb{R}^n \backslash \{0\}$ and $a_0 \in \mathbb{R}$.

Definition 4.2.4 — Supporting Hyperplane. A supporting hyperplane for a set $S \subseteq \mathbb{R}^n$ is a hyperplane $H = \{x \in \mathbb{R}^n : a^\top x = a_0\}$ such that

1. S is contained in either $\{x \in \mathbb{R}^n : a^\top x \geq a_0\}$ or $\{x \in \mathbb{R}^n : a^\top \leq a_0\}$
2. $S \cap H = \emptyset$

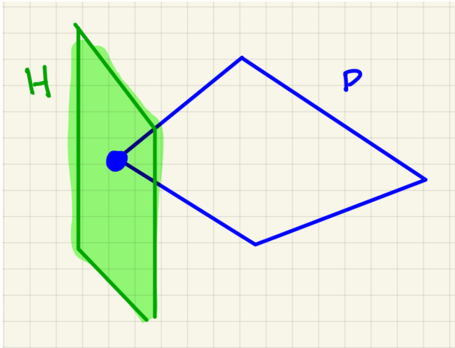


Figure 4.2.1: *Supporting Hyperplane*

R Note that, if H is a supporting hyperplane for P and $P \cap H = \{\hat{x}\}$, then $\hat{x}$ is an extreme point. The converse may not hold. Consider

$$\begin{aligned} S = &\left\{x \in \mathbb{R}^2 : 0 \leq x_1 \leq 1, -1 \leq x_2 \leq 0\right\} \cup \\ &\left\{x \in \mathbb{R}^2 : (x_1 - 1)^2 + x_2^2 \leq 1\right\} \end{aligned}$$

and $\hat{x} = (0,0)$.

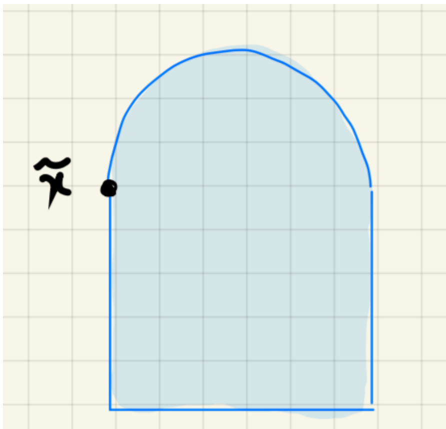


Figure 4.2.2: *Set S*

$\hat{x}$ is an extreme point, but there is no supporting hyperlane H with $H \cap S = \{\hat{x}\}$.

Theorem 11 If $\hat{x}$ is an extreme point of a polyhedron $P \subseteq \mathbb{R}^n$, then there is a supporting hyperplane H such that $P \cap H = \{\hat{x}\}$.

Proof. Support that $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ and let $a_1^\top x \geq b_1, \cdots, a_k^\top \geq b_k$ be the equality constraints for $\hat{x}$. Let $a = a_1 + \cdots + a_k$ and $a_0 = b_1 + \cdots + b_k$, and $H = \left\{x \in \mathbb{R}^n : a^\top x = a_0\right\}$. Consider any point $\hat{x} \in P \cap H$. We have

$$\begin{aligned} a_0 &= a^\top \hat{x} \\ &= a_1^\top \hat{x} + \cdots + a_k^\top \hat{x} \\ &\geq b_1 + \cdots + b_k \\ &= a_0 \end{aligned}$$

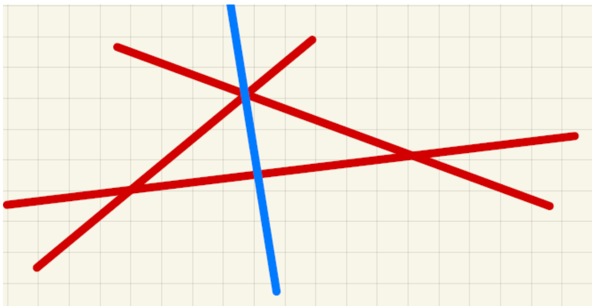
Thus, $a_1^\top \hat{x} = b_1, \cdots, a_k^\top \hat{x} = b_k$. However, the equality system for $\hat{x}$ has a unique solution, namely $\hat{x}$. and hence $H \cap P = \{\hat{x}\}$. ■

4.3 Application

Recall from linear algebra. If a system of m linear equations in n variables is infeasible, then there is an infeasible subcollection of at most $n+1$ of the equations. Say

$$Ax = b \xrightarrow{\text{Take } A_1 \text{ being maximal lin. ind.}} \begin{cases} A_1 x = b_1 \\ A_2 x = b_2 \end{cases} \xrightarrow{\text{spans } A_2} 0x = \hat{b}$$

Since the system is infeasible, $\hat{b} \neq 0$. Thus, we just need at most n variables from $A_1 x = b_1$ and one more in $A_2 x = b_2$ to make an infeasible subcollection.
Equivalently, if $H_1, \cdots, H_m$ are hyperplanes in $\mathbb{R}^n$ with empty intersection, then there is a subcollection of at most $n+1$ of these hyperplanes with empty intersection.



This is a special case of the Helly's Theorem.

Theorem 12 If a system of m linear inequalities in n variables is infeasible, then there is an infeasible subcollection of at most $n+1$ of the inequalities.

Proof. Consider an infeasible system of inequalities

$$a_1^\top x \geq b_1, \cdots, a_m^\top x \geq \alpha_m$$

By the Farkas Lemma, there exists non-negative numbers $y_1, \cdots, y_m \in \mathbb{R}$ such that $a_1 y_1 + \cdots + a_m y_m = 0$ and $b_1 y_1 + \cdots + b_m y_m = 1$. That is

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \textbf{cone}\left(\left\{ \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \cdots, \begin{bmatrix} a_m \\ b_m \end{bmatrix} \right\}\right)$$

Recall Caratheodory's theorem for cones. So, up to possible reordering, we may assume that

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} \in \textbf{cone}\left(\left\{ \begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \cdots, \begin{bmatrix} a_k \\ b_k \end{bmatrix} \right\}\right)$$

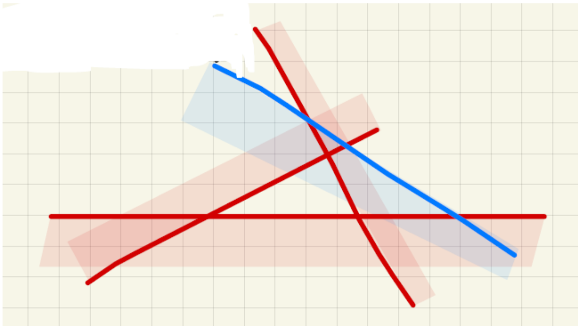
where $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \cdots, \begin{bmatrix} a_k \\ b_k \end{bmatrix}$ are linearly independent. Since $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}, \cdots, \begin{bmatrix} a_k \\ b_k \end{bmatrix} \in \mathbb{R}^{n+1}$, we have $k \leq n+1$.

By the Farkas Lemma, the system

$$a_1^\top x \geq b_1, \cdots, a_k^\top x \geq \alpha_k$$

is infeasible. ■

Equivalently, if $H_1, \cdots, H_m$ are closed half-spaces in $\mathbb{R}^n$ with empty intersection, then there is a subcollection of at most $n+1$ of these half-spaces with empty intersection.

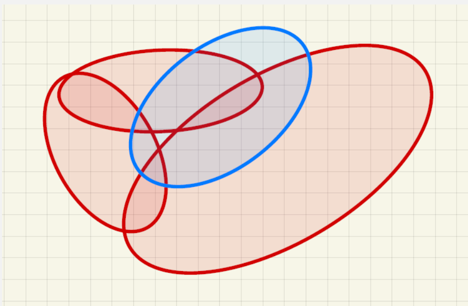


Another special case of the Helly's theorem, but we need this to prove Helly's theorem.

Corollary 4.3.1 — Helly's Theorem for Polyhedra. If $H_1, \cdots, H_m$ are polyhedra in $\mathbb{R}^n$ with empty intersection, then there is a subcollection of at most $n+1$ of these polyhedra with empty intersection.

Proof. Each of the polyhedra is iteself an intersection of finitely many closed half-spaces, so this result follows. ■

Theorem 13 — Helly's Theorem. If $S_1, \cdots, S_n \subseteq \mathbb{R}^n$ are convex sets with $S_1 \cap \cdots \cap S_m = \emptyset$, then there is a subcollection of at most $n+1$ of these sets that has empty intersection.



Proof. Suppose that the result fails for $S_1, \cdots, S_m$. Then there is a set $X \in \mathbb{R}^n$ of at most $\binom{m}{n+1}$ points such that each subcollection of $n = 1$ of the sets $S_1, \cdots, S_m$ contain a common element in X.
Let $P_1 = \textbf{conv}(S_1 \cap X), \cdots, P_m = \textbf{conv}(S_1 \cap X)$. Note that $P_1, \cdots, P_m$ are polytopes since each is the convex hull of a finite set. And $P_1 \cap P_2 \cap \cdots \cap P_m \subseteq S_1 \cap S_2 \cap \cdots \cap S_m = \emptyset$. However any subcollection of at most $n+1$ of $P_1, \cdots, P_m$ contain a common element in X. This contradicts Helly's Theorem for polyhedra. ■

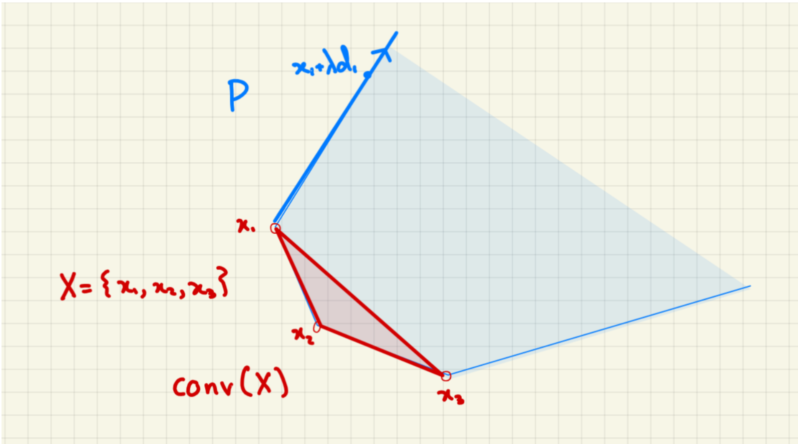
Question:
Would this still hold if we took infinitely many sets?

Answer: No, consider $S_1 = [1,\infty), S_2 = [2,\infty), S_3 = [3,\infty), \cdots$.

Exercise 4.1 Prove that if $S_1, S_2, \cdots$ is an infinite sequence of compact sets in $\mathbb{R}^n$ that have empty intersection, then there is a finite subcollection with empty intersection. ■

4.4 Unbounded Polyhedra

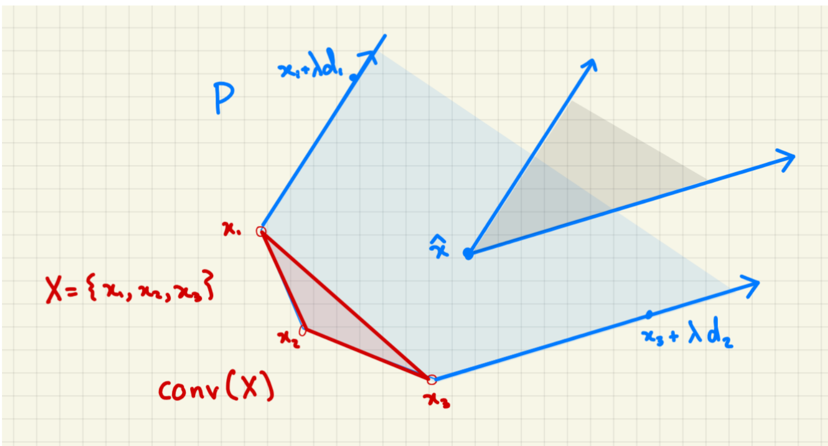
What about unbounded polyhedra?



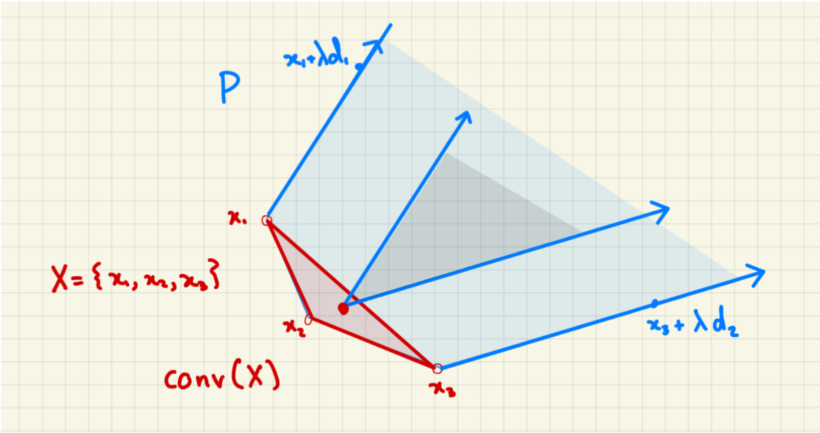
Lemma 4.5 — A2-P1. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ be a polyhedra. Let $\hat{x} \in P$ and let $\hat{d} \in \mathbb{R}^n$. Then, the ray $\{\hat{x} + \lambda \hat{d} : \lambda \geq 0\}$ is contained in $P \iff A\hat{d} \geq 0$.

R This does not depend on the point $\hat{x} \in P$.

■ **Example 4.2 — Decomposition of a Unbounded Polyhedra.** We note that for any $\hat{\in} P \backslash \mathbf{conv}(X)$, we can create a cone in the way displayed below.



We can shift this into $\mathbf{conv}(X)$ as displayed below.



Thus,

$$P = \mathbf{conv}(\{x_1, x_2, x_3\}) + \mathbf{cone}(\{d_1, d_2\})$$

■

Theorem 14 A set $P \subseteq \mathbb{R}^n$ is a polyhedron if and only if $P = \mathbf{conv}(X) + \mathbf{cone}(D)$ for some finite sets $X, D \subseteq \mathbb{R}^n$.

Proof. 1. Suppose $X, D \subseteq \mathbb{R}^n$ are finite sets.

Lemma 4.6 If $X \subseteq \mathbb{R}^n$ is a finite set, then $\mathbf{conv}(X)$ is a polytope.

this result can lead to

Lemma 4.7 If $D \subseteq \mathbb{R}^n$ is a finite set, then $\mathbf{cone}(X)$ is a polyhedron.

moreover,

Lemma 4.8 If $P_1, P_2 \subseteq \mathbb{R}^n$ are polyhedra, then $P_1 + P_2$ is a polyhedra.

and we are done with this direction.

2. Now, suppose $P \subseteq \mathbb{R}^n$ is a polyhedron.

Definition 4.8.1 — Line.

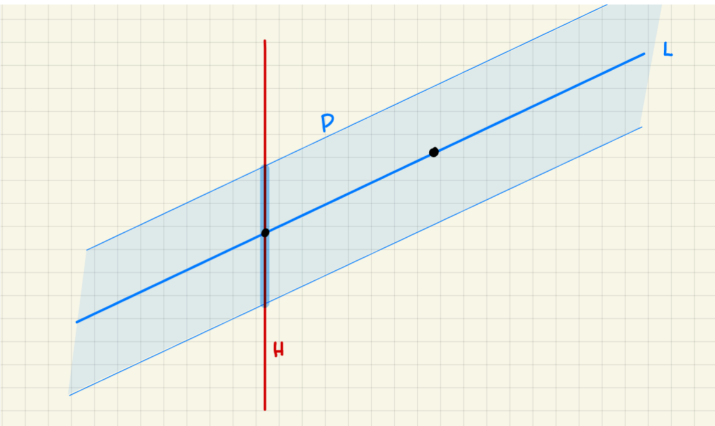
$$\mathbf{line}(\hat{x}, \hat{d}) := \{\hat{x} + \lambda \hat{d} : \lambda \in \mathbb{R}\}$$

we know that

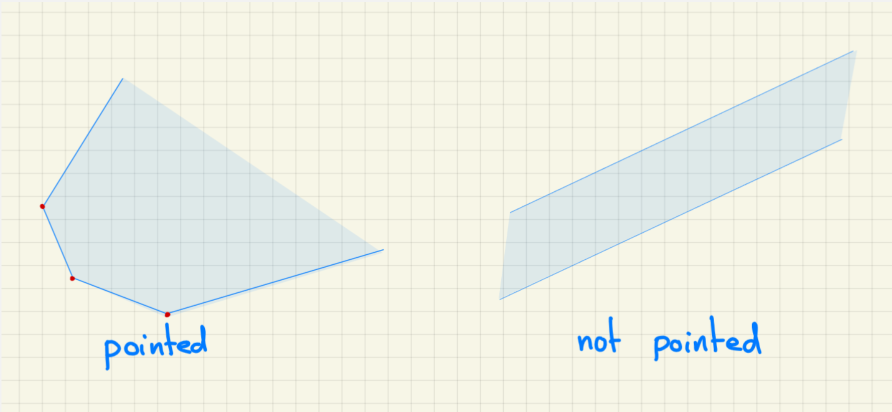
Lemma 4.9 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ be a polyhedron. Let $\hat{x} \in P$ and let $\hat{d} \in \mathbb{R}^n$. Then P contains $\mathbf{line}(\hat{x}, \hat{d}) \iff A\hat{d} = 0$.

Corollary 4.9.1 If a polyhedron contains a line, then it has no extreme points.

R If $P \subseteq \mathbb{R}^n$ is a polyhedron containing a line $L = \mathbf{line}(\hat{x}, \hat{d})$ and $H \subseteq \mathbb{R}^n$ is a hyperplane not containing L, then $P = (P \cap H) + \{\lambda \hat{d} : \lambda \in \mathbb{R}\} = (P \cap H) + \mathbf{cone}(\{\hat{d}, -\hat{d}\})$.



Definition 4.9.1 — Pointed Polyhedra. A polyhedron is pointed if it is non-empty and does not contain a line.



We will see that

Theorem 15 A polyhedron is pointed if and only if it has an extreme point.

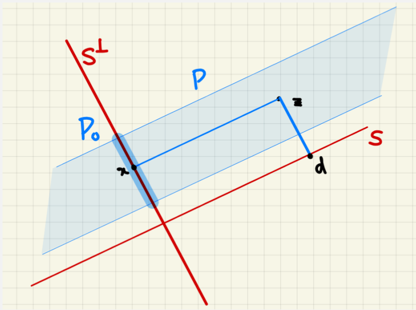
Now, we can reduce to the proof to pointed polyhedra.

Theorem 16 If $P \subseteq \mathbb{R}^n$ is a non-empty polyhedra, then there is a pointed polyhedra $P_0 \subseteq P$ and a finite set $D \subseteq \mathbb{R}^n$ such that $P = P_0 + \mathbf{cone}(D)$

Proof. Suppose that $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Define $S = \{d \in \mathbb{R}^n : Ad = 0\}$ and $S^\perp := \left\{x \in \mathbb{R}^n : x^\top d = 0, \forall d \in S\right\}$, and let $P_0 = P \cap S^\perp$.

Claim 1: P_0 does not contain a line
recall that P contains a $\mathbf{line}(\hat{x}, \hat{d})$ if and only if $\hat{x} \in P$ and $\hat{d} \in S \backslash \{0\}$. However, $S \cap S^\perp = \{0\}$, so $P_0 = P \cap S^\perp$ cannot contain such a line. ■

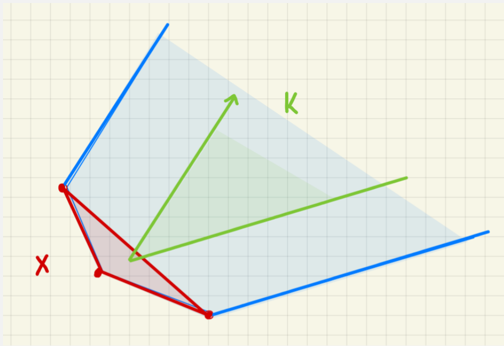
Claim 2: $P = P_0 + S$
From linear algebra, each point $z \in \mathbb{R}^n$ can be written uniquely as $x + d$ where $x \in S^\perp$ and $d \in S$ (since $S \oplus S^\perp = \mathbb{R}^n$)



Now, $z \in P$ if and only if $x \in P$. Thus, $P = P_0 + S$.

Now, for pointed polyhedra, we can actually choose X to be the set of extreme points. To be continued...

Theorem 17 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ be a pointed polyhedron where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$, let X be the set of extreme points of P, and let $K = \{x \in \mathbb{R}^n : Ax \geq 0\}$. Then, $P = \mathbf{conv}(X) + K$.



Proof. For $x, d \in \mathbb{R}^n$, define $\mathbf{ray}(x,d) := \{x + \lambda d : \lambda \geq 0\}$. From assignment, we know that $\mathbf{ray}(x,d) \subseteq P \iff Ad \geq 0$. Thus, $\mathbf{conv}(X) + K \subseteq P$.
Now, consider $\hat{x} \in P$. Assume that $\hat{x} \notin \mathbf{conv}(X) + K$. Let $A^= x \geq b^=$ be the equality subsystem for $\hat{x}$. Since $\hat{x} \notin X$, we have $\mathrm{rank}(A^=) < n$. Since P is pointed, $\mathrm{rank}(A) = n$; therefore $\hat{x}$ does not satisfy all the constraints $Ax \geq b$ with equality.

(a) **Induction hypothesis:** we may assume that if $\bar{x} \in P$ satisfies more of the constraints $Ax \geq b$ with equality than $\hat{x}$ does, then $\bar{x} \in \mathbf{conv}(X) + K$.
(b) Since $\mathrm{rank}(A^=) < n$, there is a non-zero vector $d \in \mathbb{R}^n$ such that $A^= d = 0$. Note that
   i. Each point on $\mathbf{line}(\hat{x}, d)$ satisfies $A^= x = b^=$
   ii. All points on $\mathbf{line}(\hat{x}, d)$ that are sufficiently close to $\hat{x}$ are $\hat{x}$ are contained in P.
   iii. Since P is pointed, P cannot contain both $\mathbf{ray}(\hat{x}, d)$ and $\mathbf{ray}(\hat{x}, -d)$.

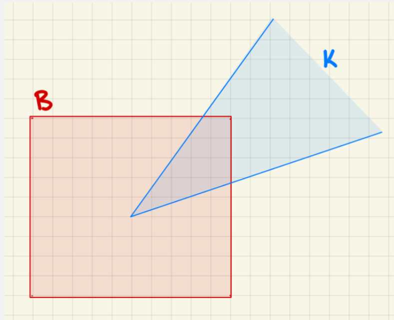
   Without loss of generality, we may assume that $\mathbf{ray}(\hat{x}, -d) \nsubseteq P$.
   i. **Case 1:** $\mathbf{ray}(\hat{x}, d) \nsubseteq P$. Choose $\lambda^- = \min(\lambda \in \mathbb{R} : \hat{x} + \lambda d \in P)$ and $\lambda^+ = \max(\lambda \in \mathbb{R} : \hat{x} + \lambda d \in P)$. These exist since P is closed and since P does not contain either $\mathbf{ray}(\hat{x}, -d)$ or $\mathbf{ray}(\hat{x}, d)$. Let $x_1 = \hat{x} + \lambda^- d$ and $x_2 = \hat{x} + \lambda^+ d$. Note that the equality subsystems for both x_1 and x_2 contain more inequalities than $A^= x \geq b^=$, so, by induction hypothesis, $x_1, x_2 \in \mathbf{conv}(X) + K$. However, it is convex, so $\hat{x} \in \mathbf{conv}(X) + K$.
   ii. **Case 2:**$\mathbf{ray}(\hat{x}, d) \nsubseteq P \subseteq P$. Thus, $d \in K$. Choose $\lambda_d = \max(\lambda \in \mathbb{R} : \hat{x} - \lambda d \in P)$. This is well-defined since P is closed and $\mathbf{ray}(\hat{x}, -d) \nsubseteq P$. Define $\bar{x} = \hat{x} - \lambda_d d$. Now, $\hat{x}$ satisfies more of the constraints $Ax \geq b$ with equality than $\hat{x}$ does. Thus, by induction hypothesis, $\bar{x} \in \mathbf{conv}(X) + K$. However, $\bar{x} = \hat{x} + \lambda_d d$ and $d \in K$. Thus, $\hat{x} \in \mathbf{conv}(X) + K$.

■

A polyhedron cone is a polyhedron that is a cone.

Lemma 4.10 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. If P is a cone, then $P = \{x \in \mathbb{R}^n : Ax \geq 0\}$.

Theorem 18 If $K \subseteq \mathbb{R}^n$ is a polyhedron cone, then there is a finite set $D \subseteq K$ such that $K = \textbf{cone}(D)$.



Proof. Let $B = \{x \in \mathbb{R}^n : -\mathbf{1} \le x \le \mathbf{1}\}$. Note that $B \cap K$ is a polytope and $K = \textbf{cone}(B \cap K)$. Let D be the finite set of extreme points of $B \cap K$. Thus, $B \cap K = \textbf{conv}(D)$ and hence

$$K = \textbf{cone}(B \cap K) = \textbf{cone}(\textbf{conv}(D)) = \textbf{cone}(D)$$

■

We now may assume that $P \neq \emptyset$ (otherwise, we just pick $X = D = \emptyset$). No, we can write

$$P = P_1 + \textbf{cone}(D_1)$$

where P_1 is a pointed polyhedron and $D_1 \subseteq \mathbb{R}^n$ is finite. Then, by the theorem above, we have

$$P_1 = \textbf{conv}(X) + K$$

where X is the set of extreme points of X and K is a polyhedron cone. Then, $K = \textbf{cone}(D_2)$ where $D_2 \subseteq \mathbb{R}^n$ is finite. Thus,

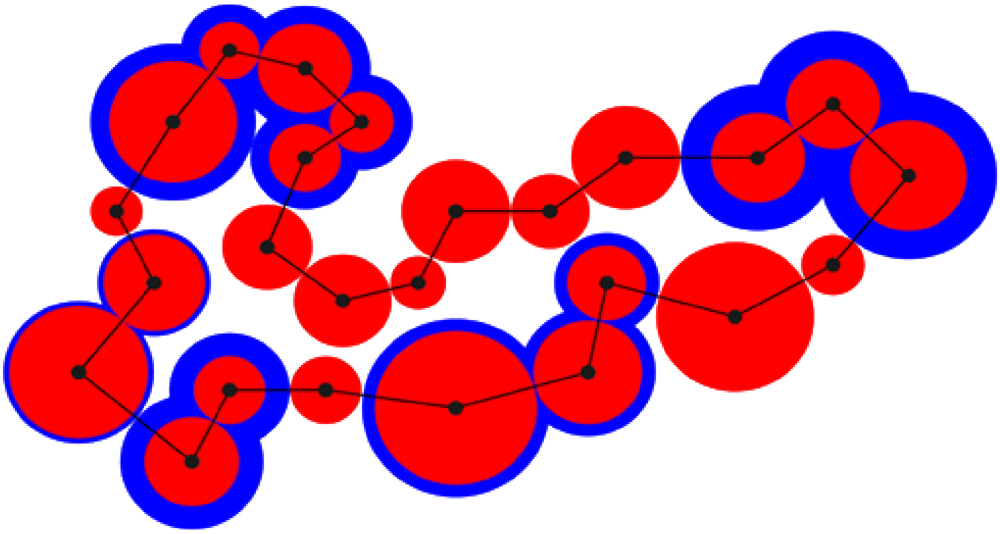
$$P = \textbf{conv}(X) + \textbf{cone}(D_2) + \textbf{cone}(D_1) = \textbf{conv}(X) + \textbf{cone}(D_1 \cup D_2)$$

Finally, WE ARE DONE :D!

■

R Here, for sets $X, Y \subseteq \mathbb{R}^n$, we define

$$X + Y := \{x + y : x \in X, y \in Y\}$$



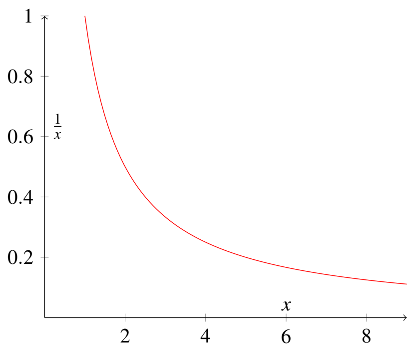
5. Linear Optimization and Duality

Theorem 19 — Fundamental Theorem of Linear Programming. A linear program is either

1. Infeasible
2. Unbounded, or
3. Has an optimal solution

Proof. Delay the proof until we prove it. ∎

(R) This is not necessarily the case for non-linear optimization problems. For example $\min\left(\frac{1}{x} : x \geq 1\right)$ is feasible and bounded but it has no optimal solution.



So given a linear program,

$$(P) \qquad \min\left(c^\top x : Ax \geq b\right)$$

Question: how can we certify infeasibility, unboundedness, and optimality?

Certifying Infeasibility

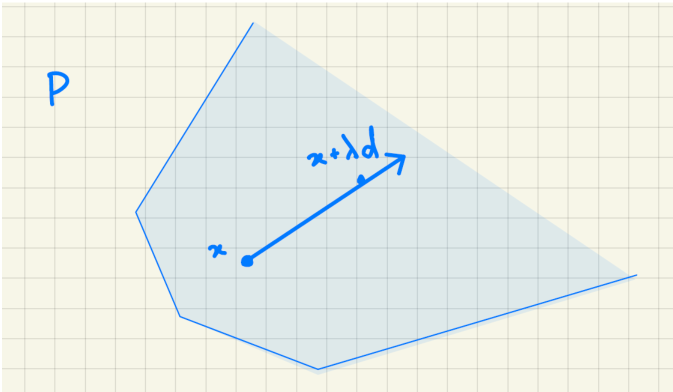
Theorem 20 — Farkas' Lemma. (P) is infeasible if and only if there is a non-negative vector $y \in \mathbb{R}^n$ such that $y^\top A = 0$ and $y^\top b = 1$.

We have done this already.

5.1 Certifying Unboundedness/Boundedness

5.1.1 Certifying Unboundedness

Lemma 5.2 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ be a polyhedron, let $x \in P$ and let $d \in \mathbb{R}^n$. Then, $\textbf{ray}(x,d) \in P \iff Ad \geq 0$.



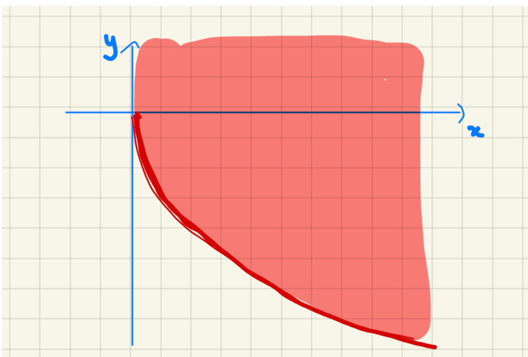
Theorem 21 — Unboundedness Theorem. (P) is unbounded if and only if (P) is feasible and there exists $d \in \mathbb{R}^n$ such that $Ad \geq 0$ and $c^\top d < 0$.

Proof. Delay the proof until we prove it. ■

R This does not extend to non-linear optimization problems. For example,

$$\min\left(y : y \geq -\sqrt{x}, x \geq 0\right)$$

is unbounded, but there is no feasible ray with unbounded objective value.



5.2.1 Certifying Boundedness (Duality)

We can use implied inequalities to certify boundedness of (P).

Example 5.1

$$(P)\begin{cases} \min & x_1 & +x_2 & & & \\ s.t. & 2x_1 & +x_2 & \geq & 4 & (a) \\ & 2x_1 & +3x_2 & \geq & 6 & (b) \\ & x_1 & +4x_2 & \geq & 4 & (c) \end{cases}$$

$\frac{3}{7}(a)+\frac{1}{7}(c)$ gives:

$$x_1+x_2 \geq \frac{16}{7}$$

hence, $\mathbf{OPT}(P) \geq \frac{16}{7}$.
What is the best bound that implied ineuqalities can give us?
Each implied inequality has the form:

$$(2y_1+2y_2+y_3)x_1+(y_1+3y_2+4y_3)x_2 \geq 4y_1+6y_2+4y_3$$

where $y_i \geq 0$. If $2y_1+2y_2+y_3=1$ and $y_1+3y_2+4y_3=1$, then we have the implied inequality

$$x_1+x_2 \geq 4y_1+6y_2+4y_3$$

In order to get the best bound, we actuall need to **maximize** $4y_1+6y_2+4y_3$ subject to

$$\begin{cases} 2y_1+2y_2+y_3=1 \\ y_1+3y_2+4y_3=1 \\ y_1,y_2,y_3 \geq 0 \end{cases}$$

that is

$$(D)\begin{cases} \max & 4y_1 & +6y_2 & +4y_3 & & \\ s.t. & 2y_1 & +2y_2 & +y_3 & = & 1 \\ & y_1 & +3y_2 & +4y_3 & = & 1 \\ & y_1, & y_2, & y_3 & \geq & 0 \end{cases}$$

By our construction, $\mathbf{OPT}(D) \leq \mathbf{OPT}(P)$. In particular, note that $x^*=\left[\frac{3}{2},1\right]^\top$ is a feasible solution to (P) with objective value $\frac{5}{2}$ and $y^*=\left[\frac{1}{4},\frac{1}{4},0\right]^\top$ is a feasible solution to (D) with objective value $\frac{5}{2}$. Thus, $\mathbf{OPT}(D)=\mathbf{OPT}(P)$ in this case. ■

Definition 5.2.1 — Dual. Consider

$$(P) \qquad \min\left(c^\top x : Ax \geq b\right)$$

the dual of (P) is

$$(D) \qquad \max\left(b^\top y : A^\top y = c, y \geq 0\right)$$

Theorem 22 — Weak Duality Theorem. If x is a feasible solution to (P) and y is a feasible solution to (D), then $c^\top x \geq b^\top y$. Equivalently, $\mathbf{OPT}(D) \leq \mathbf{OPT}(P)$.

Proof.

$$c^\top x = (A^\top y)^\top x = (y^\top A)x = y^\top(Ax) \geq y^\top b = b^\top y$$

■

Corollary 5.2.1
1. If (P) is feasible, then (D) is bounded
2. If (D) is feasible, then (P) is bounded
3. If x is feasible for (P), y is feasible for (D), and $c^\top x = b^\top y$, then x is optimal for (P) and y is optimal for (D)

Theorem 23 — Strong Duality Theorem. If (P) has an optimal solution, then (D) has an optimal solution and $\mathbf{OPT}(P) = \mathbf{OPT}(D)$.

Theorem 24 — Uber Theorem. Either

(I) (P) and (D) both have optimal solutions and $\mathbf{OPT}(P) = \mathbf{OPT}(D)$

(II) There exists $y \in \mathbb{R}^m$ such that $(y^\top A = 0, y^\top b = 1, y \geq 0)$ and, hence, (P) is infeasible.

(III) (P) is feasible and there exists $d \in \mathbb{R}^n$ such that $(c^\top d < 0, Ad \geq 0)$ and, hence, (P) is unbounded.

Proof. We assume that neither (I) nor (II) hold. By the Farkas' Lemma, since (II) does not hold, (P) is feasible. We can write (I) as

$$(A) \qquad Ax \geq b, A^\top y = c, y \geq 0, c^\top x \leq b^\top y$$

and consider

$$(B) \qquad A^\top y = zc, Ax \geq zb, b^\top y - c^\top x = 1, y \geq 0, z \geq 0$$

From assignment, we know that exactly one of (A) and (B) has a solution. Since (I) does not hold, there is a solution $\bar{x} \in \mathbb{R}^n, \bar{y} \in \mathbb{R}^m$ and $\bar{z} \in \mathbb{R}$ to

$$(B) \qquad A^\top y = zc, Ax \geq zb, b^\top y - c^\top x = 1, y \geq 0, z \geq 0$$

Claim: $\bar{z} = 0$

Proof. Suppose $\bar{z} > 0$ and let $x' = \frac{1}{\bar{z}}\bar{x}$ and $y' = \frac{1}{\bar{z}}\bar{y}$. Thus, $A^\top y' = c, Ax' \geq b, b^\top y' - c^\top x' = \frac{1}{\bar{z}}(b^\top \bar{y} + c^\top \bar{x}) = \frac{1}{\bar{z}}(b^\top \bar{y} + c^\top \bar{x}) = \frac{1}{\bar{z}} > 0, y' \geq 0$. But then, x', y' satisfies (A) and yields a contradiction. ∎

Now, $\bar{x}, \bar{y}$ satisfy: $A^\top \bar{y} = 0, A\bar{x} \geq 0, b^\top \bar{y} - c^\top \bar{x} > 0, y \geq 0$. Either

1. Case 1: $A^\top \bar{y} = 0, b^\top \bar{y} > 0, \bar{y} \geq 0$, this can be rescaled to have $(y^\top A = 0, y^\top b = 1, y \geq 0)$, which implies (P) infeasible. A contradiction.
2. Case 2: $A\bar{x} \geq 0, c^\top \bar{x} < 0$ as required.

∎

		(D)		
		infeasible	unbounded	optimal
(P)	infeasible	✓	✓	✗
	unbounded	✓	✗	✗
	optimal	✗	✗	✓

Weak Duality

Strong Duality

"Uber Theorem"

Example

Figure 5.2.1: *Summary*

5.3 Maximization

■ **Example 5.2 — Dual (Other Forms).** Consider

$$(P) \qquad \max\left(c^\top x : Ax \le b, x \ge 0\right)$$

the dual of (P) is

$$(D) \qquad \min\left(b^\top y : A^\top y \ge c, y \ge 0\right)$$

■

Theorem 25 — Weak Duality (Maximization). If x is feasible for (P) and y is feasible for (D), then $c^\top x \le b^\top y$.

Proof.

$$c^\top x \le (A^\top x)^\top x = (y^\top A)x = y^\top(Ax) \le y^\top b = b^\top y$$

■

Theorem 26 — Strong Duality (Maximization). If (P) has an optimal solution, then $\mathbf{OPT}(P) = \mathbf{OPT}(D)$.

Proof.

$$(P)\left\{\begin{array}{ll} \max & c^\top x \\ s.t. & Ax \le b \\ & x \ge 0 \end{array}\right. \Longrightarrow (P')\left\{\begin{array}{ll} \min & -c^\top x \\ s.t. & -Ax \ge -b \\ & x \ge 0 \end{array}\right.$$

and $\mathbf{OPT}(P') = -\mathbf{OPT}(P)$. The dual of (P') is

$$(D')\left\{\begin{array}{ll} \max & -b^\top y \\ s.t. & -A^\top y + z = -c \\ & y, z \ge 0 \end{array}\right. \Longrightarrow (D)\left\{\begin{array}{ll} \min & b^\top y \\ s.t. & A^\top y \ge c \\ & y \ge 0 \end{array}\right.$$

also, $\mathbf{OPT}(D) = \mathbf{OPT}(D')$. Then, by the Strong Duality Theorem in the previous version, $\mathbf{OPT}(P') = \mathbf{OPT}(D')$. Thus, $\mathbf{OPT}(D) = \mathbf{OPT}(P)$. ■

■ **Example 5.3**

$$(P)\left\{\begin{array}{lll} \max & 3x_1 - x_2 + x_3 & \\ s.t. & 2x_1 + 2x_2 = 4 & y_1 \\ & x_1 - 2x_2 + 2x_3 \le 3 & y_2 \ge 0 \\ & x_1, x_3 \ge 0 & \end{array}\right.$$

the implied inequalities have the form:

$$(2y_1 + y_2)x_1 + (2y_1 - 2y_2)x_2 + (2y_2)x_3 \le 4y_1 + 3y_2$$

we want

$$3x_1 - x_2 + x_3 \le (2y_1 + y_2)x_1 + (2y_1 - 2y_2)x_2 + (2y_2)x_3$$

the dual of (P) is

$$\left\{\begin{array}{lll} \min & 4y_1 + 3y_2 & \\ s.t. & 2y_1 + y_2 \ge 3 & x_1 \ge 0 \\ & 2y_1 - 2y_2 = -1 & x_2 \\ & 2y_2 \ge 1 & x_3 \ge 0 \\ & y_2 \ge 0 & \end{array}\right.$$

■

Dual Cheat Sheet

(P) min	(D) max
$\geq$ constraint	non-negative variable
$\leq$ constraint	non-positive variable
$=$ constraint	free variable
non-negative variable	$\leq$ constraint
non-positive variable	$\geq$ constraint
free variable	$=$ constraint

Table 5.3.1: *Caption*

5.3.1 Application to Game Theory

Given a matrix $A \in \mathbb{R}^{m\times n}$ for example,

$$A = \begin{bmatrix} 2 & -2 \\ 1 & 5 \end{bmatrix}$$

The Game

Rose choose a row $i \in \{1, \cdots, m\}$ and Colin chooses a column $j \in \{1, \cdots, n\}$, then Rose pays Colin A_{ij}.

Problem: what is the best strategy for Rose and Colin over a sequence of games?

1. Rose will construct a probability distribution $\bar{p}$ on $\{1, \cdots, m\}$ and choose the row according to $\bar{p}$
2. Colin will construct a probability distribution $\bar{q}$ on $\{1, \cdots, n\}$ and choose the column according to $\bar{q}$.

The expected payout is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\bar{p}_i\bar{q}_j$$

How should Rose and Colin choose their distributions?

Suppose that Rose and Colin play repeatedly using distributions $\bar{p}$ and $\bar{q}$ respectively, and Colin guesses Rose's strategy. Then, for each $j \in \{1, \cdots, n\}$, Colin can compute his expected return for choosing column j, $\sum_{i=1}^{m} A_{ij}\bar{p}_i$. Then, Colin would take the best of these to get an expected payout of

$$\max\left\{\sum_{i=1}^{m} A_{ij}\bar{p}_i : 1 \leq j \leq n\right\}$$

To counter Colin, Rose chooses $\bar{p}$ optimizing

$$(R)\quad \begin{array}{ll} \min & \max\left\{\sum_{i=1}^{m} A_{ij}\bar{p}_i : 1 \leq j \leq n\right\} \\ s.t. & p_1 + \cdots + p_m = 1 \\ & p_1, \cdots, p_m \geq 0 \end{array} \quad\Longleftrightarrow\quad (\tilde{R})\quad \begin{array}{ll} \min & z \\ s.t. & z \geq \sum_{i=1}^{m} A_{ij}\bar{p}_i, j = 1, \cdots, n \\ & p_1 + \cdots + p_m = 1 \\ & p_1, \cdots, p_m \geq 0 \end{array}$$

Note that $(\tilde{R})$ is feasible and bounded, it has an optimal solution, say $\bar{p}$.
Similarly, Colin chooses $\bar{q}$ optimizing

$$(C)\quad \begin{array}{ll} \max & \min\left\{\sum_{j=1}^{n} A_{ij}\bar{q}_j : 1 \leq i \leq m\right\} \\ s.t. & q_1 + \cdots + q_n = 1 \\ & q_1, \cdots, q_n \geq 0 \end{array} \quad\Longleftrightarrow\quad (\tilde{C})\quad \begin{array}{ll} \max & z \\ s.t. & z \leq \sum_{j=1}^{n} A_{ij}\bar{q}_j, i = 1, \cdots, m \\ & q_1 + \cdots + q_n = 1 \\ & q_1, \cdots, q_n \geq 0 \end{array}$$

again, since $(\tilde{C})$ is feasible and bounded, it has an optimal solution, say $\bar{q}$. Moreover, by construction,

$$\mathbf{OPT}(C) \leq \sum_{i=1}^{m}\sum_{j=1}^{n} A_{ij}\bar{p}_i\bar{q}_j \leq \mathbf{OPT}(R)$$

Theorem 27 — Amazing Theorem.

$$\mathbf{OPT}(R) = \mathbf{OPT}(C)$$

Proof. Assignment 4 ■

5.4 Certifying Optimality

Theorem 28 — Complementary Slackness Conditions. Consider

$$(P) \qquad \min(c^\top x : Ax \geq b)$$

$$(D) \qquad \max(b^\top y : A^\top y = c, y \geq 0)$$

Let $a_1^\top, \cdots, a_m^\top$ be the rows of A. If $\bar{x}$ is feasible for (P) and $\bar{y}$ is feasible for (D), then

$$c^\top \bar{x} - b^\top \bar{y} = (A^\top \bar{y})^\top \bar{x} - \bar{y}^\top b = \bar{y}^\top (A\bar{x} - b) = \sum_{i=1}^{m} \underbrace{\bar{y}_i}_{\geq 0} \underbrace{(a_i^\top \bar{x} - b_i)}_{\geq 0} \geq 0$$

Moreover, equality holds if and only if,

For each $i \in \{1, \cdots, m\}$, either $a_i^\top \bar{x} = b_i$ or $\bar{y}_i = 0$.

We call this equality condition, the **Complementary Slackness Conditions**.

■ **Example 5.4 — Unique Solution.**

$$(P) \begin{cases} \min & 2x_1 + x_2 & \\ s.t. & x_1 + x_2 \geq 5 & y_1 \geq 0 \\ & x_1 + 2x_2 \geq 6 & y_2 \geq 0 \\ & 3x_1 + x_2 \geq 9 & y_3 \geq 0 \end{cases}$$

The dual of (P) is

$$(D) \begin{cases} \max & 5y_1 + 6y_2 + 9y_3 & \\ s.t. & y_1 + y_2 + 3y_3 = 2 & x_1 \\ & y_1 + 2y_2 + y_3 = 1 & x_2 \\ & y_1, y_2, y_3 \geq 0 & \end{cases}$$

The complementary slackness conditions:

1. $x_1 + x_2 = 5$ or $y_1 = 0$
2. $x_1 + 2x_2 = 6$ or $y_2 = 0$
3. $3x_1 + x_2 = 9$ or $y_3 = 0$

Question: is $\bar{x} = [2, 3]^\top$ optimal for (P)?
We can see that 1 and 3 are satisfied and we can solve

$$\begin{cases} y_1 + y_2 + 3y_3 = 2 \\ y_1 + 2y_2 + y_3 = 1 \\ y_2 = 0 \end{cases} \implies \bar{y} = \begin{bmatrix} \frac{1}{2} \\ 0 \\ \frac{1}{2} \end{bmatrix} \geq 0$$

to certify $\bar{x}$'s optimality. ■

■ **Example 5.5 — Complementary slackness in other forms.**

$$(P)\begin{cases}\max & 3x_1 - x_2 + x_3 & \\ s.t. & 2x_1 + 2x_2 = 4 & y_1 \\ & x_1 - 2x_2 + 2x_3 \leq 3 & y_2 \geq 0 \\ & x_1, x_3 \geq 0 & \end{cases}$$

the dual of (P) is,

$$(D)\begin{cases}\min & 4y_1 + 3y_2 & \\ s.t. & 2y_1 + y_2 \geq 3 & x_1 \geq 0 \\ & 2y_1 - 2y_2 = -1 & x_2 \\ & 2y_2 \geq 1 & x_3 \geq 0 \\ & y_2 \geq 0 & \end{cases}$$

The complementary slackness conditions:

1. $x_1 + 2x_2 + 2x_3 = 3$ or $y_2 = 0$
2. $2y_1 + y_2 = 3$ or $x_1 = 0$
3. $2y_2 = 1$ or $x_3 = 0$

■

Theorem 29 — Optimality Criteria for LP. Let $\bar{x}$ be a feasible solution for (P) and let $A^{=}x \geq b^{=}$ be the equality subsystem for $\bar{x}$. Then $\bar{x}$ is an optimal solution if and only if there is a non-negative vector y such that $c = (A^{=})^\top y$ if and only if $c \in \mathbf{cone}(\text{rows}(A^{=}))$.

Proof. This is a restatement of the complementary slackness condition theorem. ■

Theorem 30 — Weak Cost-Splitting Theorem. Consider

$$(P) \qquad \min(c^\top x : x \in S_1 \cap \cdots \cap S_m)$$

where $c \in \mathbb{R}^n$ and $S_1, \cdots, S_m \subseteq \mathbb{R}^n$. Let $\bar{x} \in S_1 \cap \cdots \cap S_m$. If there exist $c_1, \cdots, c_m \in \mathbb{R}^n$ such that $c = c_1 + \cdots + c_m$ and $\bar{x}$ minimizes $(c^\top x : x \in S_i)$, for each $i \in \{1, \cdots, m\}$, then $\bar{x}$ minimizes (P).

Economic Interpretation

The cost $c^\top \bar{x}$ of $\bar{x}$ can be apportioned in lost $(c_1^\top \bar{x}, \cdots, c_m^\top \bar{x})$ and assigned to the constraints $(\bar{x} \in S_1, \cdots, \bar{x} \in S_m)$.

Theorem 31 — Cost-Splitting for LP. Consider a LP

$$(P) \qquad \min(c^\top x : Ax \geq b)$$

Let $P_i = \left\{x \in \mathbb{R}^n : a_i^\top x \geq b_i\right\}$ where $a_1^\top, \cdots, a_m^\top$ are the rows of A. Let $\bar{x} \in P_1 \cap \cdots \cap P_m$. Now, $\bar{x}$ optimizes $(c_i^\top x : x \in P_i)$ if and only if either $c_i = 0$ or $a_i^\top \bar{x} = b_i$ and $c_i = \lambda_i a_i, \lambda_i \geq 0$.

Physical Interpretation

$$(P) \qquad \min(c^\top x : Ax \geq b)$$

Consider an optimal solution $\bar{x}$ and a cost-splitting $c = c_1 + \cdots + c_m$.

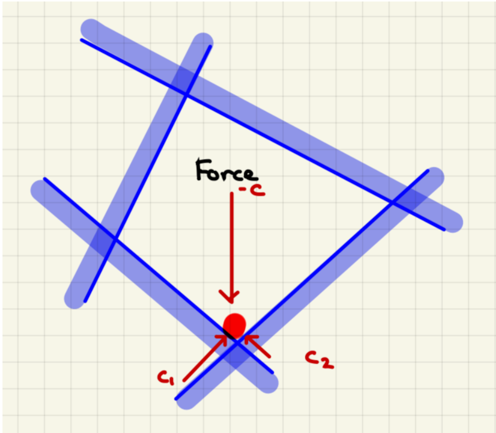


Figure 5.4.1: *Physical interpretation*

1. **Newton's First Law:** an object at rest stays at rest unless it is acted upon by an unbalanced force
2. **Newton's Third Law:** For every action, there is an equal and opposite reaction.

5.5 Solving Linear Optimization Problems

Given $A \in \mathbb{Q}^{m\times n}$ and $b \in \mathbb{Q}^m$ and $c \in \mathbb{Q}^n$. Consider the LP

$$(P) \qquad \min(c^\top x : Ax \geq b)$$

Together with the Uber Theorem, we can do

1. **Feasibility Problem:** Find either,
   (a) find a feasible solution $\bar{x} \in \mathbb{Q}^n$ to (P), or
   (b) find $y \in \mathbb{Q}^m$ such that $(y^\top A = 0, y^\top b = 1, y \geq 0)$
2. **Optimization Problem:** Given a feasible solution $\bar{x} \in \mathbb{Q}^n$, find either
   (a) a feasible solution $\hat{x}$ to (P) and a feasible solution $\hat{y}$ to (D) with $c^\top \hat{x} = b^\top \hat{y}$, or
   (b) a vector $d \in \mathbb{Q}^n$ such that $(c^\top d < 0, Ad \geq 0)$

Claim: the feasibility problem and the optimization problem are equivalent.

Proof. Given a feasible solution $\bar{x} \in \mathbb{Q}^n$, find either

1. a feasible solution $\hat{x}$ to (P) and a feasible solution $\hat{y}$ to (D) with $c^\top \hat{x} = b^\top \hat{y}$, or **equivalently**, solve
$$(Ax \geq b, A^\top y = c, y \geq 0, c^\top x = b^\top y)$$
2. a vector $d \in \mathbb{Q}^n$ such that $(c^\top d < 0, Ad \geq 0)$, **equivalently**, solve
$$(c^\top d = 1, Ad \geq 0)$$

For the converse, consider

$$(AP) \qquad \min(s_1 + \cdots + s)m : Ax + s \geq b, s \geq 0)$$

Lemma 5.6 (P) is feasible if and only if $\mathbf{OPT}(AD) = 0$

Proof. Proved in Assignment 4. ■

Lemma 5.7 If (P) is infeasible, then, from an optimal solution to the dual of (AP), we can get $y \in \mathbb{R}^m$ such that $(A^\top y = 0, b^\top y = 1, y \geq 0)$.

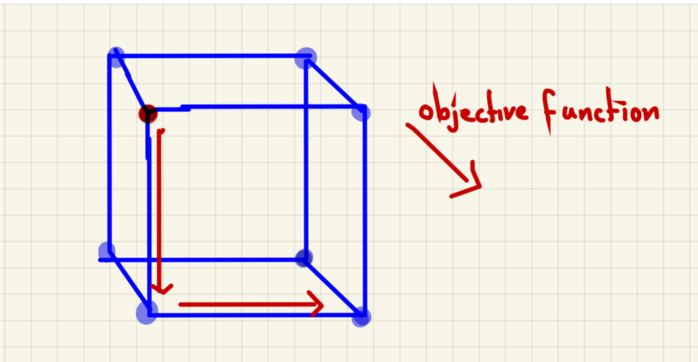
Proof. Proved in Assignment 4. ∎

∎

R We will try to solve the **Optimization Problem**.

5.7.1 Simplex Method - Revised dual Simplex Method and the Perturbation Method

Idea: move from extreme point to extreme point around the boundary improving the objective value.



Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$.

Problem: P **may not have an extreme point**

We are given $\hat{x} \in P$, so $P \neq \emptyset$. Recall that the following are equivalent:

1. P has no extreme point
2. P contains a line
3. $\text{rank}(A) < n$

Solution 1: replace (P) by

$$(\hat{P}) \qquad \min(c^\top(x_1 - x_2) : A(x_1 - x_2) \geq b, x_1, x_2 \geq 0)$$

R This is easy, but it doubles the number of variables unnecessarily.

Solution 2: suppose that $\text{rank}(A) < n$, and let $d \in \mathbb{R}^n$ be a non-zero vector such that $Ad = 0$. Thus, $\textbf{line}(\bar{x}, d) \in P$.

Claim: If $c^\top d \neq 0$, then (P) is unbounded

Proof. By possibly replacing d with $-d$, we may assume that $c^\top d < 0$. Then, by the Unboundedness Theorem, (P) is unbounded. ∎

Thus, we may assume that $c^\top d = 0$.

Claim: If $d_i \neq 0$, then for each $x \in P$, there exists $x' \in P$ such that $c^\top x' = c^\top x$ and $x'_i = 0$

Proof. Let $\lambda = \frac{x_i}{d_i}$ and let $x' = x - \lambda d$. Since $c^\top d = 0$, we have $c^\top x = c^\top x'$ and $x_i = 0$. ∎

Let (P') be the problem obtained from (P) by setting $x_i = 0$. Now, (P') has fewer variables than (P) and, by the claim, $\textbf{OPT}(P) = \textbf{OPT}(P')$.

■ Example 5.6 — Unbounded.

$$(P)\begin{cases} \min & x_1+x_2+x_3 \\ & 3x_1-x_2+x_3 \geq 5 \\ & -x_1+x_2+x_3 \geq 2 \\ & x_1+x_2+3x_3 \geq 11 \\ & x_1+x_3 \geq 4 \end{cases}$$

$\hat{x} = [1,1,3]^\top$ is a feasible solution.

$$A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} \qquad A \underbrace{\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}}_{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Now, $c^\top d = [1,1,1]\begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix} = 2$. Let $d = -d$. So d satisfies $(c^\top d < 0, Ad \geq 0)$ and hence (P) is unbounded. ■

■ Example 5.7 — Reduced (P).

$$(P)\begin{cases} \min & 2x_1-x_2 \\ & 3x_1-x_2+x_3 \geq 5 \\ & -x_1+x_2+x_3 \geq 2 \\ & x_1+x_2+3x_3 \geq 11 \\ & x_1+x_3 \geq 4 \end{cases}$$

$\hat{x} = [1,1,3]^\top$ is a feasible solution.

$$A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & 3 \\ 1 & 0 & 1 \end{bmatrix} \qquad A \underbrace{\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}}_{d} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Now, $c^\top d = [2,-1,1]\begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix} = 0$. Since $d_1 \neq 0, d_2 \neq 0, d_3 \neq 0$, we can set any of x_1, x_2, x_3 to zero. So (P) is equivalent to

$$(P')\begin{cases} \min & 2x_1 \\ & 3x_1+x_3 \geq 5 \\ & -x_1+x_3 \geq 2 \\ & x_1+3x_3 \geq 11 \\ & x_1+x_3 \geq 4 \end{cases}$$

We can repeat this process. ■

From the example above, we can assume that $\text{rank}(A) = n$ and hence that P has an extreme point.

Problem: Given $\bar{x} \in P$, find an extreme point of P

Algorithm 5.1 — Extreme Point Search.

1. **Step 1:** construct the equality subsystem $A^= x \geq b^=$ for $\bar{x}$. If $\text{rank}(A^=) = n$, STOP ($\bar{x}$ is an extreme point)
2. **Step 2:** Find a non-zero vector $d \in \mathbb{R}^n$ such that $A^= d = 0$. If $Ad \geq 0$, replace d with $-d$.
3. **Step 3:** Let $\lambda^- = \max(\lambda \in \mathbb{R} : x + \lambda d \in P)$. Replace $\bar{x}$ with $\bar{x} + \lambda^- d$. Repeat from Step 1.

■ **Example 5.8**

$$(P) \begin{cases} \min & x_1 + x_2 \\ s.t. & x_1 + x_2 + x_3 \geq 4 \\ & 3x_1 - x_2 + x_3 \geq 6 \\ & x_1 + 2x_2 + 2x_3 \geq 5 \\ & x_1 + 2x_2 - x_3 \geq 2 \end{cases}$$

Now, $\bar{x} = [2, 1, 1]^\top$ is feasible. The equality subsystem is

$$\begin{cases} x_1 + x_2 + x_3 \geq 4 \\ 3x_1 - x_2 + x_3 \geq 6 \end{cases}$$

Now,

$$\begin{bmatrix} 1 & 1 & 1 \\ 3 & -1 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}}_{d} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

check

$$Ad = \begin{bmatrix} 0 \\ 0 \\ -1 \\ 0 \end{bmatrix} \not\geq \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Choose λ maximum such that $\bar{x} + \lambda d$ is feasible.

$$A(\bar{x} + \lambda d) = \begin{bmatrix} 4 \\ 6 \\ 6 - \lambda \\ 3 + 5\lambda \end{bmatrix} \geq \begin{bmatrix} 4 \\ 6 \\ 5 \\ 2 \end{bmatrix}$$

Take $\lambda = 1$; our new solution is $\bar{=}[3, 2, -1]^\top$, which satisfies constraints 1,2,3 with equality. We can easily verify that $A^=$ has rank 3 and hence $\bar{x}$ is an extreme point. ■

Exercise 5.1 Revise the algorithm so that is finds either
1. An extreme point $\bar{x}'$ with $c^\top \bar{x}' \leq c^\top \bar{x}$, or
2. a vector $d \in \mathbb{R}^n$ such that $c^\top d < 0, Ad \geq 0$.

■

Problem: given an extreme point $\bar{x} \in P$, solve (P)

Let $A^= x \geq b^=$ be the equality subsystem for $\bar{x}$.

Optimality Condition: $\bar{x}$ is optimal for (P) if and only if there exists a solution to $((A^=)^\top y = c, y \geq 0)$.

R Feasibility problems of this form are, in general, as hard to solve as linear optimization itself!

■ **Example 5.9**

$$(P)\begin{cases} \min & 2x_1+3x_2 \\ s.t. & x_1+x_2\geq 5 \\ & x_1+2x_2\geq 6 \\ & 3x_1+x_2\geq 9 \end{cases}$$

Question: is $\bar{x}=[2,3]^\top$ optimal for (P)?
The equality subsystem is

$$\begin{cases} x_1+x_2\geq 5 \\ 3x_1+x_2\geq 9 \end{cases}$$

solve

$$\begin{cases} y_1+3y_2=2 \\ y_1+y_2=3 \\ y_1,y_2\geq 0 \end{cases} \implies_{\text{unique}} \hat{y}=\begin{bmatrix} \frac{7}{2} \\ -\frac{1}{2} \end{bmatrix}$$

since $\hat{y}\ngeq 0$, our solution $\bar{x}$ is not optimal. ■

R If $A^=$ is square and non-singular, then there is a unique solution $\hat{y}$ to $(A^=)^\top y=c$. Then, $\bar{x}$ is optimal if and only if $\hat{y}\geq 0$. Moreover, since $\bar{x}$ is an extreme point, $\text{rank}(A^=)=n$. We only care about whether $A^=$ is square or not.

Definition 5.7.1 — Degeneracy. Since $\text{rank}(A^=)=n$, $\bar{x}$ is the unique solution to $A^=\hat{x}=b^=$. If $A^=$ has more than n rows, then $A^=x=b^=$ is over determined; in this case, we call $\bar{x}$ degenerate.

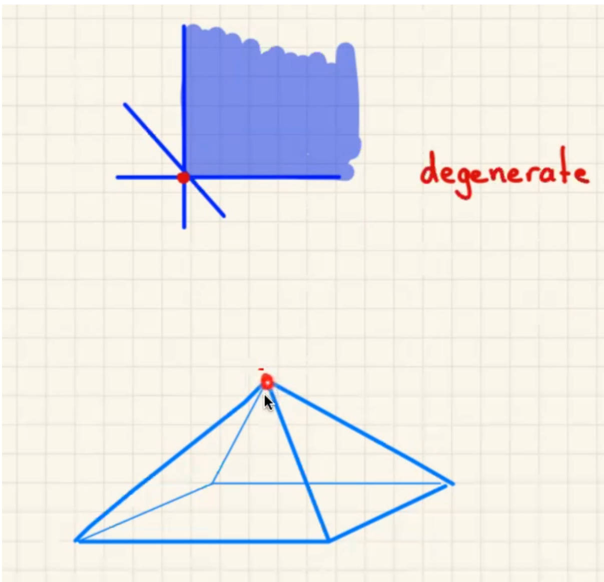


Figure 5.7.1: *You only need 3 planes to determine one point.*

We shall assume that $\bar{x}$ is non-degenerate. But we will void this assumption later.

R If $\hat{y}\geq 0$, then we can extend $\hat{y}$ to an optimal solution for (D) by setting $\bar{y}_i=0$ for all non-equality constraints. (Here we really want to be careful about indices.)

Now, let B be the indices of the equality constraints; thus, $A^= \in \mathbb{R}^{B\times n}$ and $\hat{y} \in \mathbb{R}^B$. Suppose $\hat{y}_j < 0$. Define $e_j \in \mathbb{R}^B$ such that $j-$th component is 1 and 0 otherwise. Let d denote the unique solution to $A^= d = e_j$. Note that

$$c^\top d = (\hat{y}^\top A^=)d = \hat{y}^\top e_j = \hat{y}_j < 0$$

Claim: For sufficiently small $\varepsilon > 0$, $\bar{x} + \varepsilon d$ is feasible for (P)

Proof. It suffices to prove that for each $i \in \{1, \cdots, m\}$, $a_i^\top(\bar{x} + \varepsilon d) \geq b_i$ for sufficiently small $\varepsilon > 0$.

1. If $i \notin B$, then $a_i^\top \bar{x} > b_i$ and hence $a_i^\top(\bar{x} + \varepsilon d) \geq b_i$ for small $\varepsilon > 0$
2. If $i \in B$, then $a_i^\top$ is a row of $A^=$, so

$$a_i^\top(\bar{x} + \varepsilon d) = \begin{cases} b_i + \varepsilon & i = j \\ b_i & \text{otherwise} \end{cases}$$

■

1. $Ad \geq 0$, then (P) is unbounded
2. $Ad \ngeq 0$, choose $\lambda \in \mathbb{R}$ maximum such that $\bar{x} + \lambda d \in P$.

Claim: $\bar{x} + \lambda d$ is an extreme point and $c^\top(\bar{x} + \lambda d) < c^\top \bar{x}$

Proof. Immediate result from the previous claim. ■

The equality subsystem for $\bar{x} + \lambda d$ contains all of the constraints indexed by $B \backslash \{j\}$, which determine $\textbf{line}(\bar{x}, d)$. Moreover, the definition of λ provides an additional equality constraint which is not satisfied by all points on $\textbf{line}(\bar{x}, d)$. So, $\bar{x} + \lambda d$ is uniquely determined by its equality subsystem and hence $\bar{x} + \lambda d$ is extreme.

Algorithm 5.2 — Simplex Algorithm. Given an extreme point $\bar{x}$ to (P):

1. **Check non-degeneracy:** Let $B = \left\{i \in \{1, \cdots, m\} : a_i^\top \bar{x} = b_i\right\}$, let $N = \{1, \cdots, m\} \backslash B$ and let $A^= x \geq b^=$ be the equality subsystem for $\bar{x}$. If $|B| \neq n$, STOP ($\bar{x}$ is degenerate)
2. **Test for optimality:** Solve $(A^=)^\top y = c$ for $\hat{y} \in \mathbb{R}^B$. If $\hat{y} \geq 0$, STOP ($\bar{x}$ is optimal)
3. **Choose leaving constraint:** choose $j \in B$ such that $\hat{y}_j < 0$.
4. **Check unboundedness:** Solve $A^= d = e_j$ for d and let $z = Ad$. If $z \geq 0$, STOP ((P) is unbounded)
5. **Choose entering constraint:** choose $i \in N$ with $z_i < 0$ minimizing $\frac{a_i^\top \bar{x} - b_i}{-z_i}$.
6. **Update:** Let $\lambda = \frac{a_i^\top \bar{x} - b_i}{-z_i}$, replace $\bar{x}$ with $\bar{x} + \lambda d$, and repeat from Step 1.

R If there are no degenerate extreme points, then the algorithm will terminate (since there are at most $\binom{m}{n}$ extreme points) and will solve (P) correctly.

■ **Example 5.10**

$$(P) \begin{cases} \min & x_1 + x_2 & \\ & 2x_1 + x_2 \geq 4 & (1) \\ & 2x_1 + 3x_2 \geq 6 & (2) \\ & x_1 + 4x_2 \geq 4 & (3) \end{cases}$$

Consider the feasible solution $\hat{x} = \left[\frac{12}{5}, \frac{2}{5}\right]^\top$. The equality subsystem for $\bar{x}$ is

$$\begin{cases} 2x_1 + 3x_2 \geq 6 & (2) \\ x_1 + 4x_2 \geq 4 & (3) \end{cases}$$

so $\bar{x}$ is a non-degenerate extreme point. Solve

$$\begin{cases} 2y_2 + y_3 = 1 \\ 3y_2 + 4y_3 = 1 \end{cases} \implies \hat{y}_2 = \frac{3}{5}, \hat{y}_3 = -\frac{1}{5}$$

Then, solve

$$\begin{cases} 2d_1 + 3d_2 = 0 & (2) \\ d_1 + 4d_2 = 1 & (3) \end{cases} \implies d = \begin{bmatrix} -\frac{3}{5} \\ \frac{2}{5} \end{bmatrix}$$

Let $\hat{x} = \bar{x} + \lambda d = \begin{bmatrix} \frac{12}{5} \\ \frac{2}{5} \end{bmatrix} + \lambda \begin{bmatrix} -\frac{3}{5} \\ \frac{2}{5} \end{bmatrix}$ Choose $\lambda \in \mathbb{R}$ maximum such that

$$2\hat{x}_1 + \hat{x}_2 \geq 4 \qquad (1)$$

We get $\lambda = \frac{3}{2}$ and the entering constraint is (1). The new extreme point is $\hat{x} = \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix}$ and the equality subsystem is

$$\begin{cases} 2x_1 + x_2 \geq 4 & (1) \\ 2x_1 + 3x_2 \geq 6 & (2) \end{cases}$$

so $\hat{x}$ is non-degenerate. Solve

$$\begin{cases} 2y_1 + 2y_2 = 1 \\ y_1 + 3y_2 = 1 \end{cases} \implies \hat{y}_1 = \hat{y}_2 = \frac{1}{4}$$

Since $\hat{y} \geq 0$, $\bar{x}$ is an optimal solution for (P). Note that $\hat{y} = \begin{bmatrix} \frac{1}{4} \\ \frac{1}{4} \\ 0 \end{bmatrix}$ is an optimal solution for (D). ■

5.7.2 Pertubation Method

What if we run into degeneracy? Consider

$$(P) \qquad \min(c^\top x : Ax \geq b)$$

and

$$(P') \qquad \min(c^\top x : Ax \geq b')$$

where $b' = \begin{bmatrix} b_1 - \varepsilon \\ b_2 - \varepsilon^2 \\ \vdots \\ b_m - \varepsilon^m \end{bmatrix}$ and ε is indeterminate that we think of as a small positive real number.

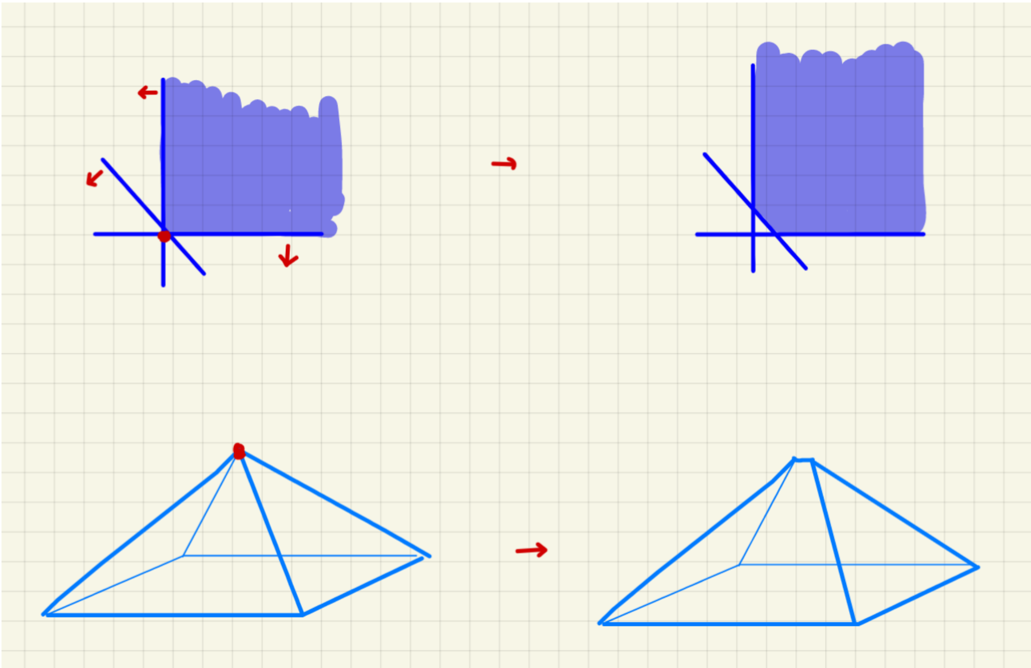


Figure 5.7.2: *Intuition for Pertubation Method*

For polynomial $p(\varepsilon)$ and $q(\varepsilon)$, we write $p(\varepsilon) < q(\varepsilon)$ if $p(\varepsilon') < q(\varepsilon')$ for all sufficiently small $\varepsilon' > 0$.

■ **Example 5.11**

$$1+\varepsilon+1000\varepsilon^2 < 1+2\varepsilon$$

■

Proposition 5.7.1 (P') has no degenerate extreme points.

Proof. Consider an extreme point $\bar{x}$ of (P'). Let

$$X = \left\{ i \in \{1, \cdots, m\} : a_i^\top \bar{x} = b_i' \right\}$$

If $\bar{x}$ is degenerate, then the vectors $\{a_i : i \in X\}$ are linearly dependent. So, there is a nonzero vector $\lambda \in \mathbb{R}^X$ such that $\sum_{i \in X} \lambda_i a_i = 0$. Then,

$$0 = \sum_{i \in X} \lambda_i a_i^\top \bar{x} = \sum_{i \in X} \lambda_i b_i' = \sum_{i \in X} \lambda_i (b_i - \varepsilon^i)$$

However, $\lambda \neq 0$ implies $\sum_{i \in X} \lambda_i (b_i - \varepsilon^i)$ is a non-zero polynomial in ε. This yields a contradiction. ■

R

1. Since (P') is non-degenerate, the Simplex Method will terminate.
2. There is some computational overhead in applying this method, but we can switch between (P) and (P') easily, so we need only use (P') when we are at a degenerate solution for (P).
3. If the equality subsystem has m' constraints, we need only perturb $m' - n$ of them, keeping $\bar{x}$ extreme.

■ **Example 5.12**

$$(P)\begin{cases}\min & x_2 \\ s.t. & x_1+x_2 \geq 8 \quad (1) \\ & x_1-x_2 \geq 2 \quad (2) \\ & -x_1+2x_2 \geq 0 \quad (3) \\ & 2x_1+x_2 \geq 13 \quad (4)\end{cases}$$

$\bar{x} = \begin{bmatrix}5\\3\end{bmatrix}$ is a feasible solution. The equality constraints for $\bar{x}$ are (1), (2), and (4), so $\bar{x}$ is degenerate.

$$A^{=} = \begin{bmatrix}1 & 1\\ 1 & -1\\ 2 & 1\end{bmatrix}$$

We note that $\begin{bmatrix}1 & -1\\ 2 & 1\end{bmatrix}$ is non-singular, so we perturb (1).

$$(P')\begin{cases}\min & x_2 \\ s.t. & x_1+x_2 \geq 8-\varepsilon \quad (1) \\ & x_1-x_2 \geq 2 \quad (2) \\ & -x_1+2x_2 \geq 0 \quad (3) \\ & 2x_1+x_2 \geq 13 \quad (4)\end{cases}$$

Now, $\bar{x}$ is a non-degenenrate extreme point of (P'). Solving

$$\begin{bmatrix}1 & 2\\ -1 & 1\end{bmatrix} y = \begin{bmatrix}0\\1\end{bmatrix}$$

for y gives $y = \begin{bmatrix}-\frac{2}{3}\\ \frac{1}{3}\end{bmatrix}$. That is $y_2 = -\frac{2}{3}$ and $y_4 = \frac{1}{3}$. The leaving constraint is (2). Now, solve $A^{=}d = e_2$.

$$\begin{bmatrix}1 & -1\\ 2 & 1\end{bmatrix} d = \begin{bmatrix}1\\0\end{bmatrix}$$

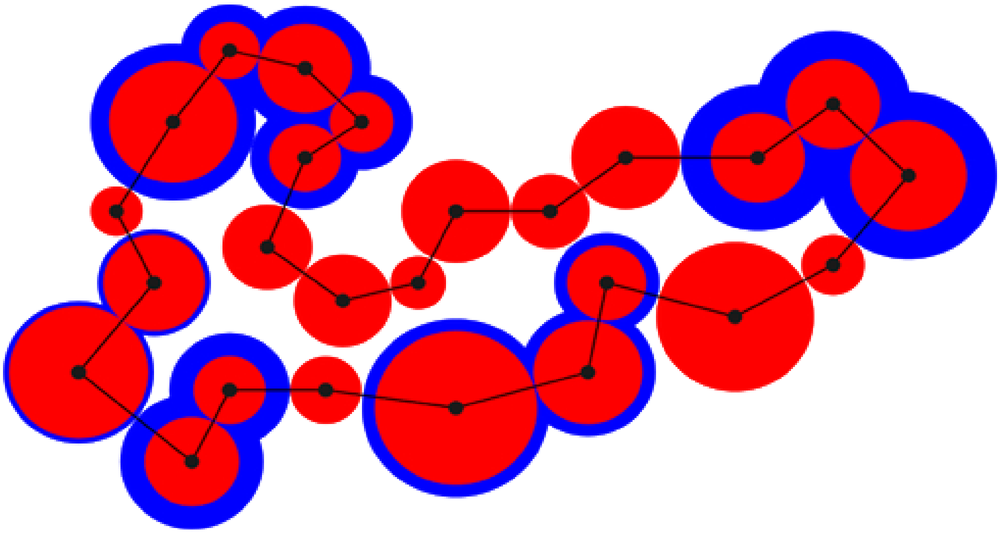
We choose λ maximum such that $A(\bar{x}+\lambda d) \geq b$

$$\begin{bmatrix}1 & 1\\ 1 & -1\\ -1 & 2\\ 2 & 1\end{bmatrix}\begin{bmatrix}5+\frac{1}{3}\lambda\\ 3-\frac{2}{3}\lambda\end{bmatrix} = \begin{bmatrix}8-\frac{1}{3}\lambda\\ 2+\lambda\\ 1-\frac{5}{3}\lambda\\ 13\end{bmatrix} \geq \begin{bmatrix}8-\varepsilon\\ 2\\ 0\\ 13\end{bmatrix}$$

So $\lambda = \min\left\{3\varepsilon, \frac{3}{5}\right\}$. Our new solution is

$$\bar{x} = \begin{bmatrix}5+\varepsilon\\ 3-2\varepsilon\end{bmatrix}$$

and the equality systems are (1) and (4). We can continue. Note that the objective value improved from 3 to $3-2\varepsilon$. When the constant term improves we set the ε's to zero and continue using the Simplex Method until the next time we encounter a degenerate extreme point, at which point we apply a new perturbation. ■



6. Integer Linear Optimization

6.1 Integer LP

Definition 6.1.1 — Integer LP (IP). An integer inlinear optimization problem is a problem of the form:

$$(IP) \qquad \min(c^\top x : Ax \geq b, x \in \mathbb{Z}^n)$$

where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, and $c \in \mathbb{R}^n$.

Definition 6.1.2 — Linear Relaxation. The linear relaxation of (IP) is

$$(LP) \qquad \min(c^\top x : Ax \geq b)$$

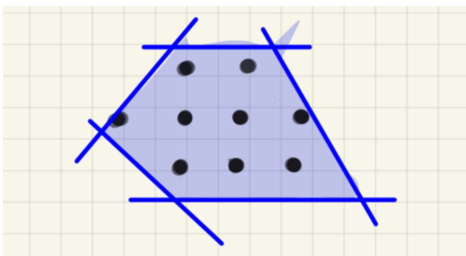


Figure 6.1.1: *Linear relaxation of IP*

R Note that $\mathbf{OPT}(LP) \leq \mathbf{OPT}(IP)$ since each feasible solution to (IP) is feasible for (LP).

Definition 6.1.3 — Integral Polyhedra. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ and let $Z = P \cap \mathbb{Z}^n$. Thus, P is the feasible region for (LP) and Z is the feasible region for (IP). Note that $\mathbf{conv}(Z) \subseteq P$; equality is rate, even when A and b are integral.

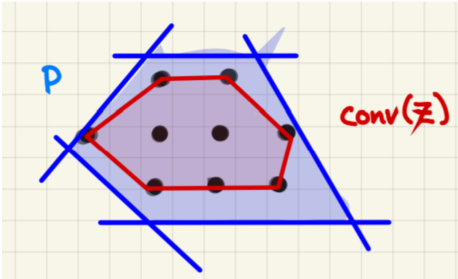


Figure 6.1.2: *Integral Polyhedra*

A polyhedron $P \subseteq \mathbb{R}^n$ is integral if $P = \mathbf{conv}(P \cap \mathbb{Z}^n)$.

R If P is integral, then

$$\min(c^\top x : x \in P \cap \mathbb{Z}^n) = \min(c^\top x : x \in P), \forall c \in \mathbb{R}^n$$

Thus, $\mathbf{OPT}(IP) = \mathbf{OPT}(LP)$.

Recall that a polytope is a bounded polyhedron.

Lemma 6.2 A polytope $P \subseteq \mathbb{R}^n$ is integral if and only if its extreme points are integer valued.

Proof. A polytope is a convex hull of its extreme points. ■

Recall that a polyhedron is pointed if it is non-empty and does not contain a line.

Theorem 32 Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ where $A \in \mathbb{Z}^{m\times n}$ and $b \in \mathbb{Z}^m$. If P is pointed, then P is integral if and only if all of its extreme points are integral.

Proof. Using $P = \mathbf{conv}(X) + \{x \in \mathbb{R}^n : Ax \geq 0\}$ where X is the set of extreme points of P. ■

6.3 Constructing Integral Polyhedra

Definition 6.3.1 — Totally Unimodular. A matrix is totally unimodular (TU) if each of its square submatrices has determinant $0, \pm 1$. In particular, the entries must be $0, \pm 1$.

Lemma 6.4 Let $A \in \{0, \pm 1\}^{m\times n}$ be TU and $b \in \mathbb{Z}^m$. Then, the extreme points of $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ are integer valued.

Proof. Let $\bar{x} \in \mathbb{R}^n$ be an extreme point of P. Then, there is a subsystem $A'x \geq b'$ of $Ax \geq b$ such that A' is $n \times n$ and non-singular. Since A is TU, $\det(A')_= \pm 1$. By Cramer's rule, each entry of $(A')^{-1}$ is integer valued. Hence each entry of $\bar{x} = (A')^{-1}b'$ is integer valued. ■

Proposition 6.4.1 — Constructions for TU Matrices. Let $A \in \{0, \pm 1\}^{m\times n}$ be TU. Then,

1. $A^\top$ is TU
2. $[I, A]$ is TU
3. If A' is obtained from A by scaling a row or column by -1, then A' is TU
4. $[A, -A]$ is TU

Lemma 6.5 If $A \in \{0, \pm 1\}$ is TU, $b \in \mathbb{Z}^m$ and $l, u \in \mathbb{Z}^n$, then $P = \{x \in \mathbb{R}^n : Ax \geq b, l \leq x \leq u\}$ is integral.

Proof. We write $P = \{x \in \mathbb{R}^n : A'x \geq b'\}$ where

$$A' = \begin{bmatrix} A \\ I \\ -I \end{bmatrix} \qquad b' = \begin{bmatrix} b \\ l \\ -u \end{bmatrix}$$

By the constructions, A' is TU. By the previous lemma, each extreme point of P is integral. Moreover, P is a polytope. So P is integral. ■

Theorem 33 If $A \in \{0, \pm 1\}^{m \times n}$ is TU and $b \in \mathbb{Z}^m$, then $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ is integral.

Proof. Let $\bar{x} \in P$ and for each $i \in \{1, \cdots, n\}$. Define $l_i = \lfloor \bar{x}_i \rfloor$ and $u_i = \lceil \bar{x}_i \rceil$. Now,

$$\bar{x} \in \{x \in \mathbb{R}^n : Ax \geq b, l \leq x \leq u\} \subseteq \mathbf{conv}(P \cap \mathbb{Z}^n)$$

Therefore, P is integral. ■

Exercise 6.1 Let $A \in \{0, \pm 1\}^{m \times n}$ be TU and let $b \in \mathbb{Z}^m$.

1. show that $\{x \in \mathbb{R}^n : Ax \geq b, x \geq 0\}$ is integral
2. show that $\{x \in \mathbb{R}^n : Ax = b, x \leq 0\}$ is integral

■

Lemma 6.6 Let $A = \{0, \pm 1\}^{m \times n}$. If each column of A has at most one 1 and at most one -1, then A is TU. For example,

$$\begin{bmatrix} 0 & 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & -1 & 0 \end{bmatrix}$$

Proof. Suppose otherwise and consider a counterexample $A \in \{0, \pm 1\}^{m \times n}$ with $m + n$ minimum. Clearly, $m = n$ and $\det(A) \notin \{0, \pm 1\}$. Since $A \in \{0, \pm 1\}^{m \times n}$, $m \geq 2$. Since we ahve a minimum counterexample, each column has both a 1 and -1. But then the rows of A sum to 0. Hence, $\det(A) = 0$, which yields a contradiction. Suppose we only have 1 or -1 in that column, we can write

$$A = \begin{bmatrix} \pm 1 & \cdots \\ \vdots & A' \end{bmatrix}$$

where $\det(A') \in \{0, \pm 1\}$ implies $\det(A) \in \{0, \pm 1\}$. Again, contradiction. ■

6.7 Application in Graph Theory

Definition 6.7.1 — Graph. A graph $G = (V, E)$ contains vertices and edges.

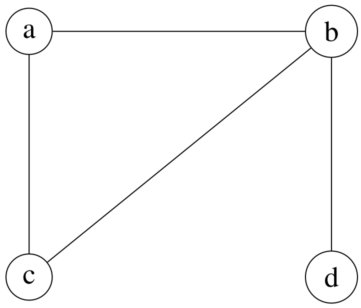


Figure 6.7.1: $G = (V, E)$

In this case, $V = \{a, b, c, d\}$ and $E = \{ab, bc, ac, bd\}$, unordered pairs of vertices. We use **incidence matrix** to encode these information. For this case,

$$A = \begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} ab & bc & ac & bd \end{array} \\ \begin{pmatrix} 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Note that

1. The column-sums are all 2
2. The row-sum for row v is the number of neighbours of v and is denoted $\deg(v)$.
3. A need not be TU, note that

$$\begin{array}{c} \\ a \\ b \\ c \end{array} \begin{array}{c} \begin{array}{ccc} ab & bc & ac \end{array} \\ \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \end{array}$$

   has determinant 2.

Definition 6.7.2 — Bipartite graphs. A graph $G = (V, E)$ is bipartite with bipartition (X, Y) if (X, Y) is a partition of V and each edge has one end in X and one end in Y.

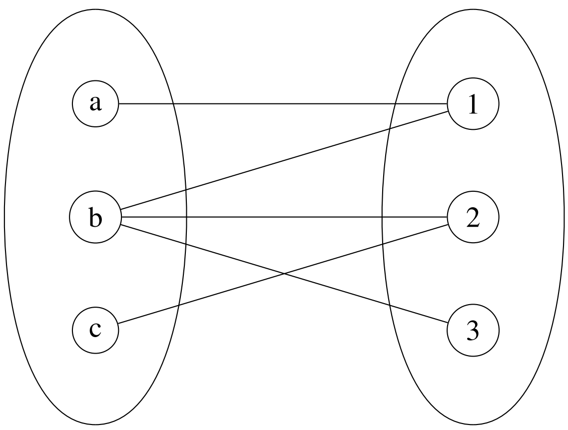


Figure 6.7.2: *Bipartite Graphs*

Theorem 34 The incidence matrix of a bipartite graph is TU.

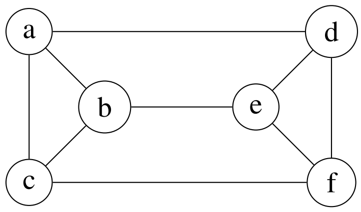
Proof. Let (X,Y) be a bipartition of a graph $G=(V,E)$ and let A be the incidence matrix.

$$A = \begin{array}{c c c} & uv & \\ \begin{array}{c} u \\ v \end{array} & \begin{pmatrix} 1 & \\ 1 & \end{pmatrix} & \begin{array}{c} X \\ Y \end{array} \end{array}$$

Let A' be obtained from A by multiplying the rows indexed by Y by -1. Then, by Lemma 6.6, we have A' is TU and A is TU. ∎

6.7.1 Matching

Definition 6.7.3 — Matching. $M \subseteq E(G)$ is a matching if no two edges in M are incident with a common vertex. A matching M is perfect if each vertex is incident with an edge of M.



Example 6.1
1. $\{ac, df\}$ is a matching
2. $\{ad, be, cf\}$ is a perfect matching

∎

Matching Polyhedra

Let A be the incidence matrix of a graph $G=(V,E)$. Define $M(G)=\left\{x \in \mathbb{R}^E : Ax \leq 1, x \geq 0\right\}$ and $PM(G)=\left\{x \in \mathbb{R}^E : Ax = 1, x \geq 0\right\}$. For $x \in \mathbb{R}^E$, let $\textbf{support}(x)=\{e \in E : x_e \neq 0\}$ Note that

1. $M(G) \cap \mathbb{Z}^E \subseteq \{0,1\}^E$
2. For $x \in \{0,1\}^E$, $x \in M(G)$ if and only if $\textbf{support}(x)$ is a matching
3. For $x \in \{0,1\}^E$, $x \in PM(G)$ if and only if $\textbf{support}(x)$ is a perfect matching
4. if G is bipartite, then A is TU and, hence, $M(G)$ and $PM(G)$ are integral.

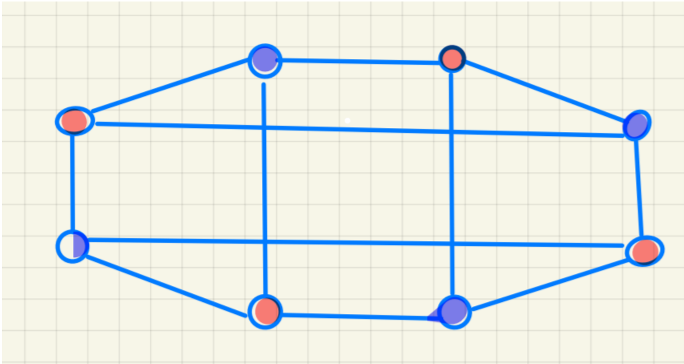
Matching Polytope

Let $\mathfrak{M}(G)$ denote the set of all vectors $x \in \{0,1\}^E$ such that $\textbf{support}(x)$ is a matching and let $\mathfrak{PM}(G)$ denote the set of vectors $x \in \{0,1\}^E$ such that $\textbf{support}(x)$ is a perfect matching.

Theorem 35 If G is a bipartite graph, then
1. $\textbf{conv}(\mathfrak{M}(G)) = \{x \in \mathbb{E}^n : Ax \leq 1, x \geq 0\}$
2. $\textbf{conv}(\mathfrak{PM}(G)) = \{x \in \mathbb{E}^n : Ax = 1, x \geq 0\}$

Application

Definition 6.7.4 — R-regular. A graph G is r-regular if each of its vertices has exactly r neighbours.



Theorem 36 For each $r \geq 1$, if G is an r-regular bipartite graph, then G has a perfect matching.

Proof. Let $\bar{x} = \left[\frac{1}{r}, \cdots, \frac{1}{r}\right]^\top$. Hence $A\bar{x} = 1$ and $\bar{x} \geq 0$. Then, $\bar{x} \in \mathbf{conv}(\mathfrak{PM}(G))$. Hence, $\mathfrak{PM}(G) \neq \emptyset$. ■

Definition 6.7.5 — Multigraphs. A multigraph G is r-regular if each of its vertices is incident with exactly r edges. For multigraphs, we allow parallel edges.

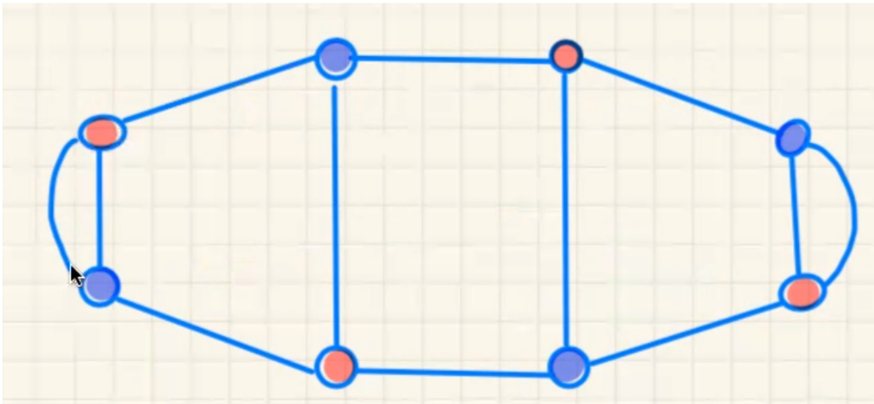


Figure 6.7.3: *Sample multigraph*

Theorem 37 For each $r \geq 1$, if G is an r-regular bipartite multigraph, then G has a perfect matching.

Proof. Same as for graphs. ■

Definition 6.7.6 — r-edge-colouring. An r-edge-colouring is a colouring of the edges with r colours so that no two edges of the same colour are incident with a common vertex.

Corollary 6.7.1 Every r-regular bipartite multigraph is r-edge-colourable.

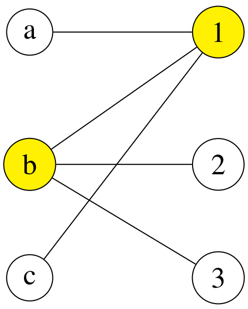
Proof. After we obtain a perfect matching, you colour one for each node and delete the edges. Inductively, you have the result. ■

Exercise 6.2 Show that, if we arrange a deck of cards in a rectangle with 4 rows and 13 columns, then by recording the cards in each column (the cards must stay in their assigned column) we can get each row containing $A, 2, \cdots, K$ in some order.

Figure 6.7.4: *Prof.Geelen teaching students how to make money*

Definition 6.7.7 — Covers. $C \subseteq V$ is a cover if $G - C$ has no edges.

For example,



$C = \{1, b\}$ is a cover. Note that if C is a cover and M is a matching, then $|M| \leq |C|$.

Theorem 38 — Kőnig's Theorem. In a bipartite graph, the maximum size of a matching is equal to the minimum size of a cover.

Proof. Let A be the incidence matrix of a bipartite graph G. Consider

$$(P) \qquad \max\left(\sum_{e \in E} x_e : Ax \leq 1, x \geq 0\right)$$

and its dual

$$(D) \qquad \min\left(\sum_{v \in V} y_v : A^\top y \geq 1, y \geq 0\right)$$

Note that (P) is feasible ($x = 0$) and (D) is feasible ($y = 1$). Hence, (P) and (D) both have optimal

solutions and $\mathbf{OPT}(P) = \mathbf{OPT}(D)$. Moreover, since A is TU, the feasible regions of both (P) and (D) are integral. Hence, (P) and (D) have optimal solutions, $\bar{x}, \bar{y}$.
Moreover, since A is TU, the feasible regions of both (P) and (D) are integral. Hence (P) and (D) have optimal solutions, $\bar{x}$ and $\bar{y}$ say, that are integer valued. Note that $\bar{x} \in \{0,1\}^E$ and $\bar{y} \in \{0,1\}^V$. Let $M = \mathbf{support}(\bar{x})$ and $C = \mathbf{support}(\bar{y})$. Note that M is a matching and C is a cover, and since $\mathbf{OPT}(P) = \mathbf{OPT}(D)$, $|M| = \sum_{e \in E} \bar{x}_e = \sum_{v \in V} \bar{y}_v = |C|$ as required. ■

Finding a maximum matching

Let $G = (V, E)$ be a bipartite graph with bipartition (X, Y), and let M be a matching.

Problem: Find a larger matching if possible.

■ **Example 6.2** Consider the following graph

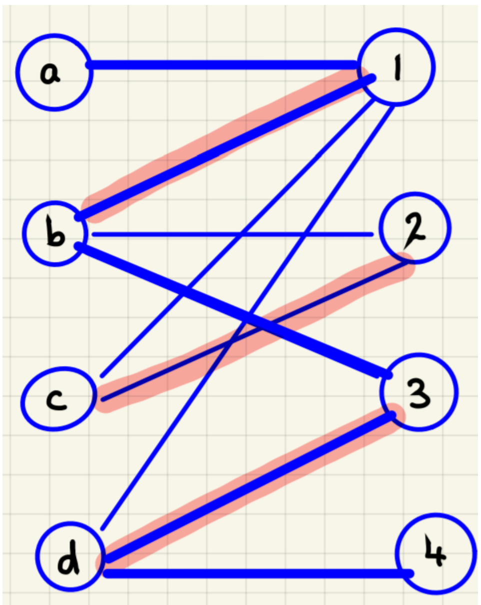


Figure 6.7.5: $M = \{b1, c2, d3\}$. *a and* 4 *are* $M-$*exposed*

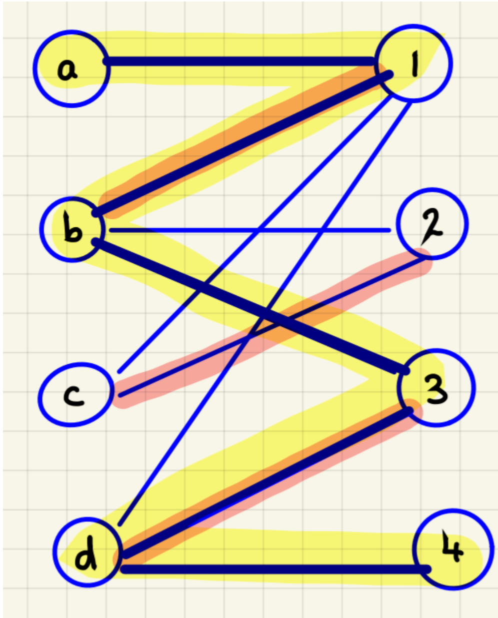


Figure 6.7.6: $P = (a,1,b,3,d,4)$. *Then,* $M' = M\Delta E(P) = \{a1,b3,c2,d4\}$

Let $\vec{G}$ be the directed graph obtained by directing the edges in $E - M$ towards Y and the edges in M towards X. Let $\bar{X}$ be the $M-$exposed vertices in X and $\bar{Y}$ be the $M-$exposed vertices in Y.

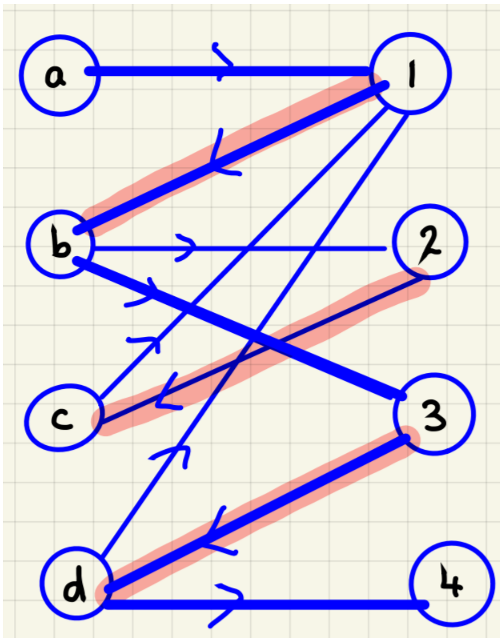


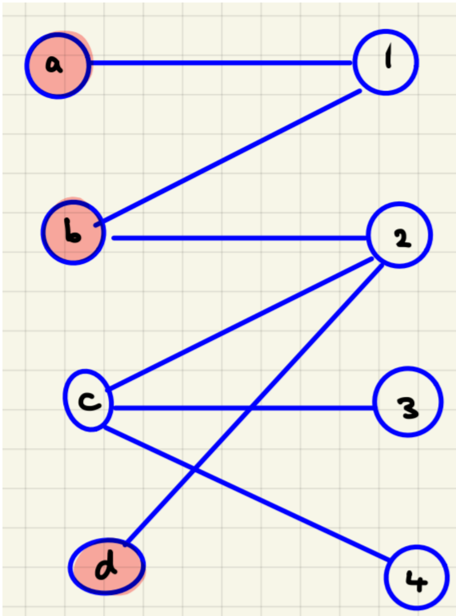
Figure 6.7.7: $\vec{G}$ *directed graph*

Proposition 6.7.2 M is a maximum matching in G if and only if there is no directed path from $\bar{X}$ to $\bar{Y}$ in $\vec{G}$.

■

Condition for Perfect Matchings

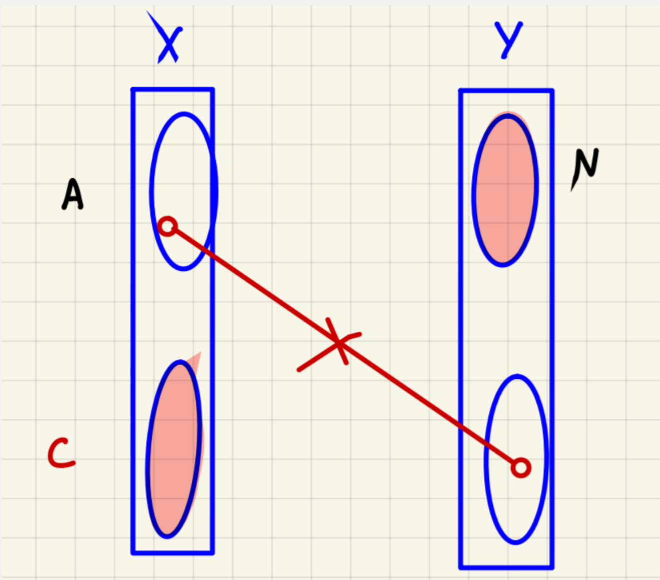
Consider the following example,



Definition 6.7.8 — Neighbour Set. For a graph $G = (V, E)$ and $X \subseteq V$, the neighbour set of X, denoted by $N(X)$, is the set of vertices in $V \backslash X$ that have a neighbour in X.

Theorem 39 — Hall's Theorem. A bipartite graph G with bipartition (X, Y) has a perfect matching if and only if $|X| = |Y|$ and $|N(A)| \geq |A|$ for each $A \subseteq X$.

Proof. The conditions are clearly necessary. Suppose that G has no perfect matching and that $|X| = |Y|$. By Konig's Theorem, G has a cover C with $|C| < |X|$. Let $A = X \backslash C$ and $N = C \backslash X$.



Since C is a cover, $N(A) \subseteq N$. Moreover, since $|X| = |Y|$, we have

$$|A| = |X| - |C| + |N| > |N| \geq |N(A)|$$

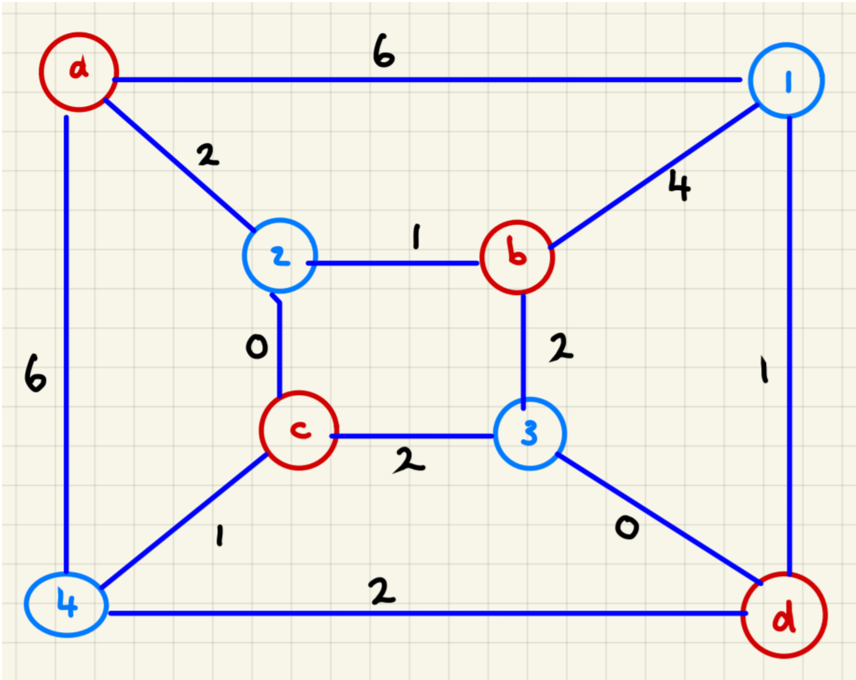
■

For this graph, we have $N(\{a,b,d\}) = \{1,2\}$. Thus, G has no perfect matching.

Minimum cost perfect matching in bipartite graphs

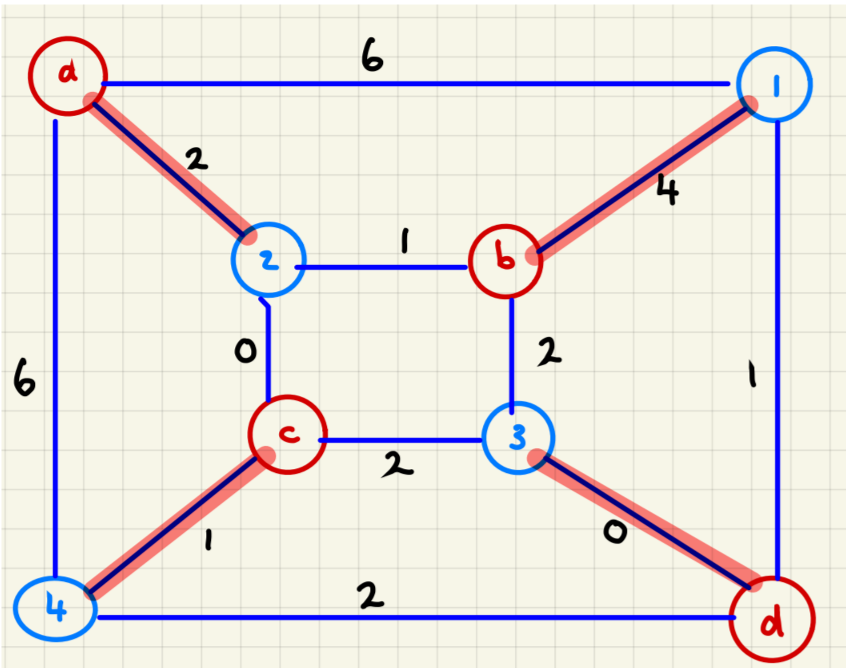
Let $G = (V,E)$ and $C \in \mathbb{Q}^E$.
Problem: find a perfect matching M minimizing $C(M) := \sum_{e \in M} C(e)$

■ **Example 6.3** Consider



■

where $X = \{a,b,c,d\}$ and $Y = \{1,2,3,4\}$. We propose the following perfect matching with $C(M) = 7$, is this the minimum?



Integer Linear Formulation

$$(IP)\begin{cases} \min & \sum_{e\in E} C(e)x_e & \\ s.t. & \sum(x_e : e \text{ incident with } v) = 1 & v\in V(G) \\ & x_e \geq 0 & e \in E \\ & x_e \in \mathbb{Z} & e\in E \end{cases}$$

In matrix form, we have $\min(c^\top x : Ax = \mathbf{1}, x \geq 0, x \in \mathbb{Z}^E)$ where A is the incidence matrix of G. The linear relaxation is

$$(LP) \qquad \min(c^\top x : Ax = \mathbf{1}, x \geq 0)$$

Since G is bipartite, A is TU. Therefore, $\{x \in \mathbb{R}^E : Ax = \mathbf{1}, x \geq 0\}$ is integral and $\mathbf{OPT}(LP) = \mathbf{OPT}(IP)$. The dual of (LP) is

$$(D) \qquad \max(\mathbf{1}^\top y : A^\top y \leq c)$$

that is

$$\begin{cases} \max & \sum_{v\in V} y_v & \\ s.t. & y_u + y_v \leq C(uv) & uv \in E \end{cases}$$

For a feasible solution $\bar{x}$ to (LP) and $\bar{y}$ to (D) the complementary slackness conditions say

For each $uv \in E$, either $\bar{x}_{uv} = 0$ or $\bar{y}_u + \bar{y}_v = C(uv)$

If the coplementary slackness conditions are satisfied, then $\bar{x}$ is optimal for (LP).
Let $\bar{y}$ be a feasible solution to (D), let

$$E^=(\bar{y}) = \{uv \in E : y_u + y_v = C(uv)\}$$

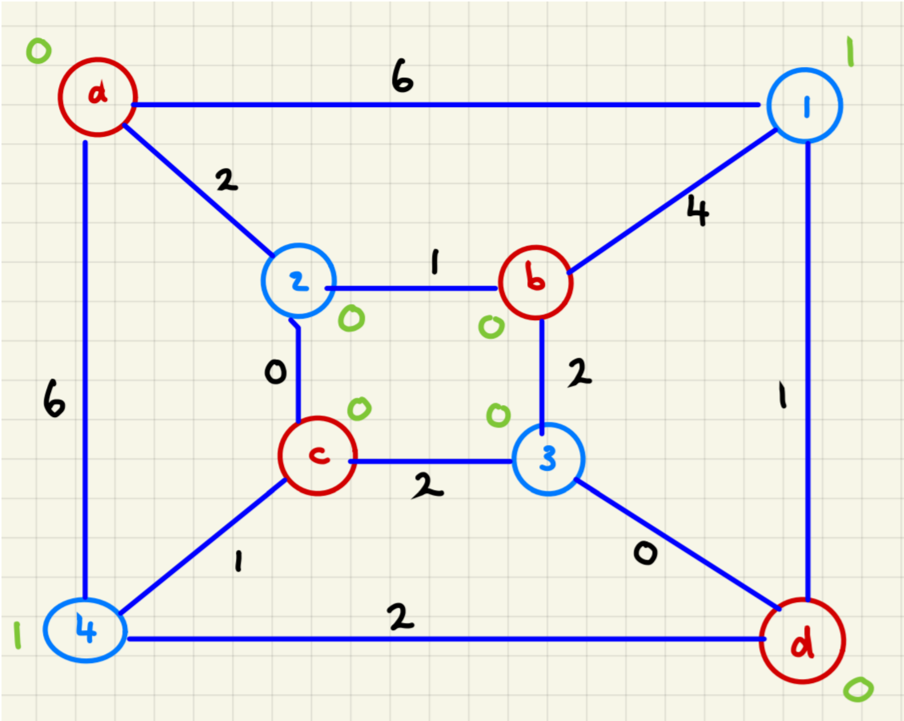
and $G^=(\bar{y}) = (V, E^=(\bar{y}))$ is the **equality subgraph**.
Optimal Condition: if $\bar{M}$ is a perfect matching in $G^=(\bar{y})$, then $\bar{M}$ is a minimum-cost perfect matching.

Algorithm 6.1 — Minimum-cost perfect matching search.
1. **Step 0:** Find a feasible solution $\bar{y}$ to (D)
2. **Step 1:** If $G^=(\bar{y})$ has a perfect matching $\bar{M}$, then STOP and return $\bar{M}$ as the min-cost perfect matching.
3. **Step 2:** Use Hall's Theorem to find a new feasible solution $\hat{y}$ to (D) with

$$\sum_{v\in V} \hat{y}_v > \sum_{v\in V} \bar{y}_v$$

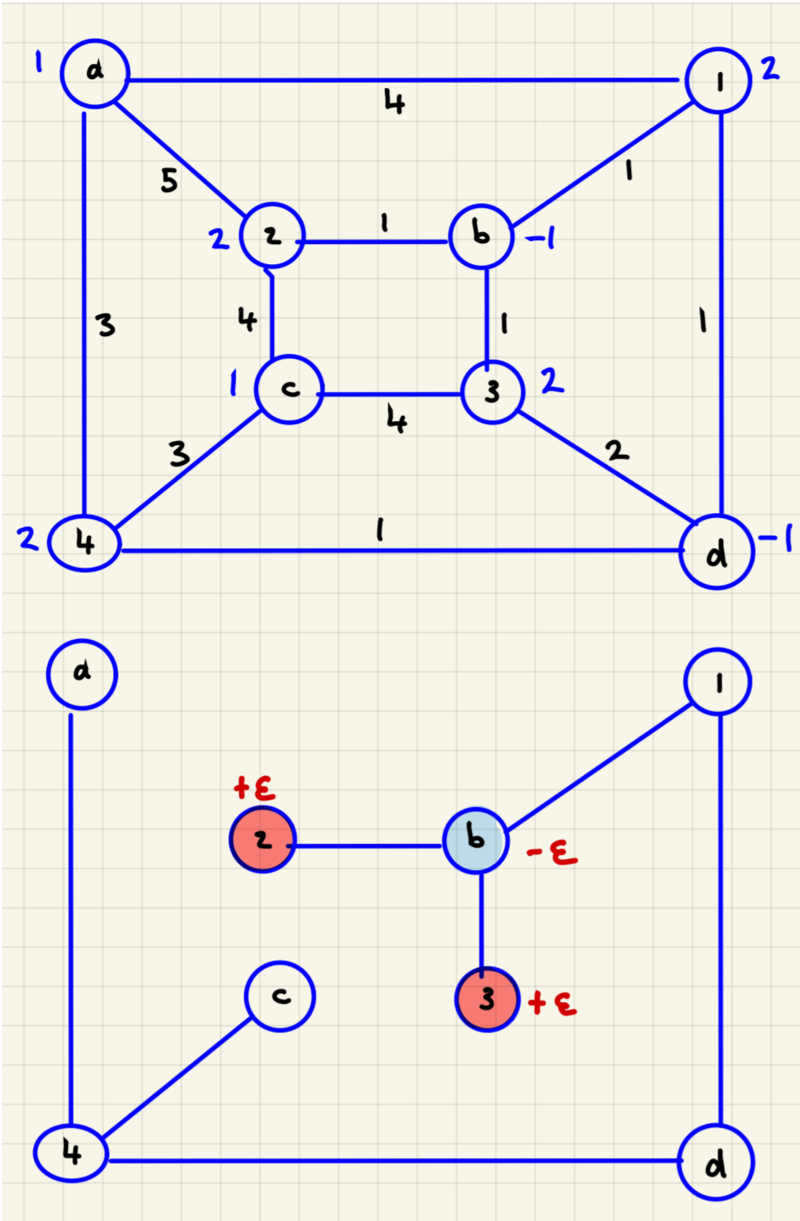
assign $\hat{y} \mapsto \bar{y}$ and repeat from **STEP 1**

■ **Example 6.4 — Existence of feasible solution $\bar{y}$ to (D).** Set $y_v = 0$ for each $v \in X$ and set $y_w = \min(C(vw) : vw \in E)$ for each $w \in Y$.

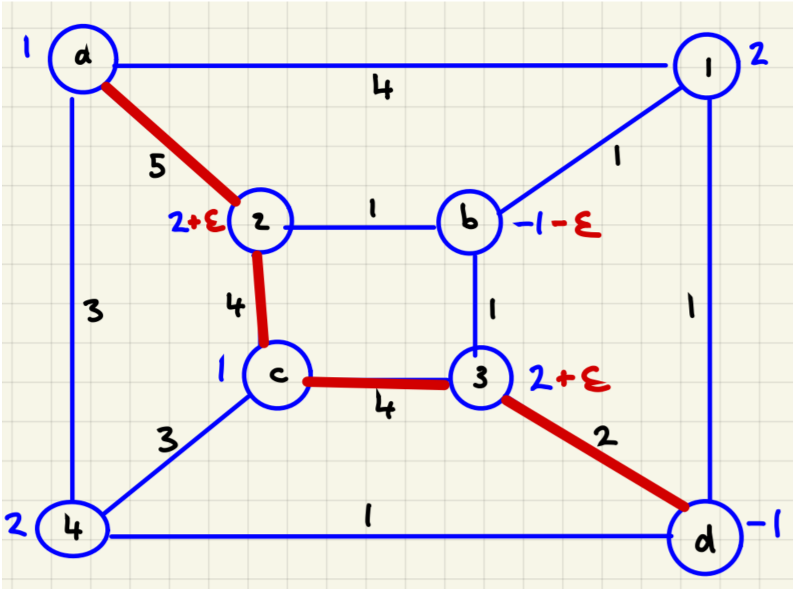


■

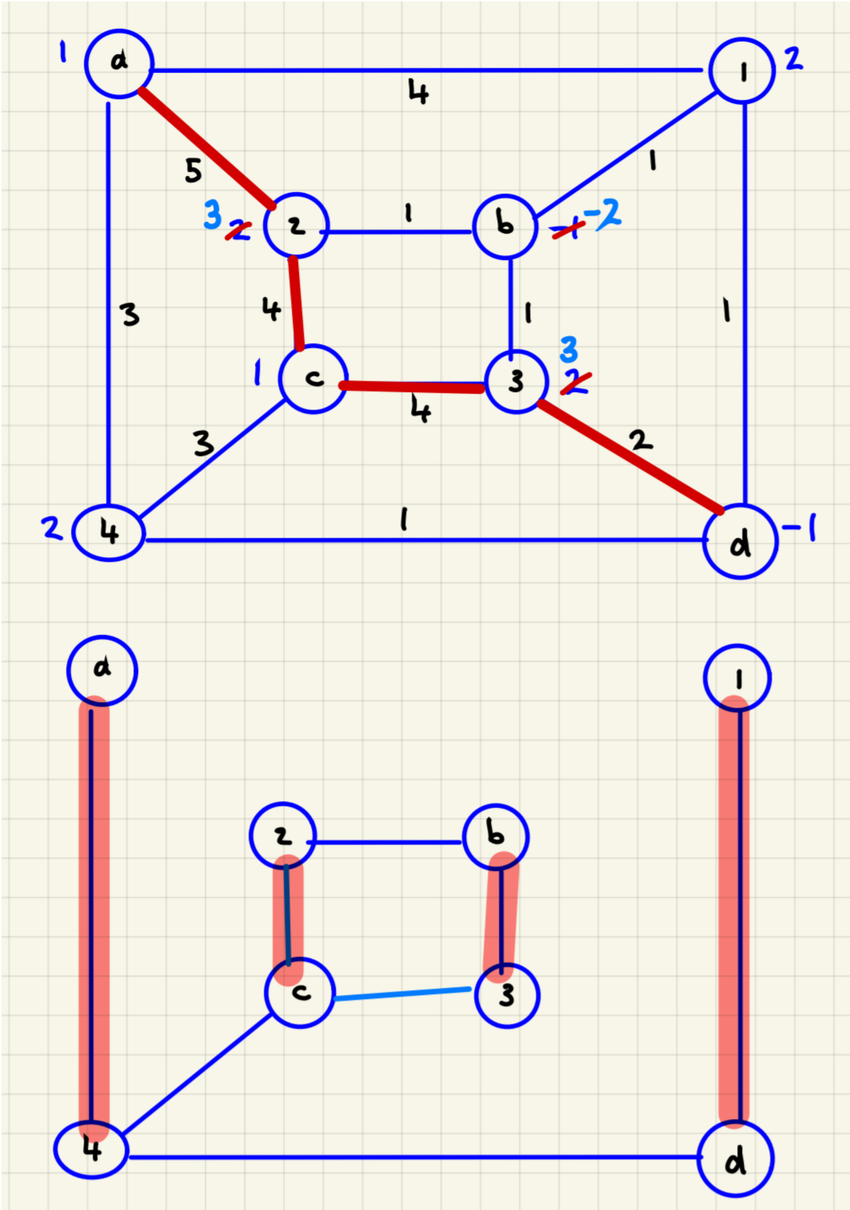
■ **Example 6.5 — Perturbation.** We start with a feasible solution $\bar{y}(V) = 8$ and consider its equality subgraph $G^{=}(\bar{y})$.



By Hall's Theorem, $|N(\{2,3\})| = 1$ means there is no local perfect matching. We perturb $\{2,3,b\}$.



We check the complementary slackness conditions for the perturbed case.



The new equality subgraph contains a perfect matching and it will be a minimum cost perfect matching. ■

Algorithm 6.2 — Detailed Minimum-cost perfect matching search.

1. **Step 0: Find a feasible solution $\bar{y}$ to (D)**

$$\bar{y}_v = \begin{cases} 0 & v \in X \\ \min(C(uv) : uv \in E) & v \in Y \end{cases}$$

2. **Step 1: Test for Optimality** If $G^{=}(\bar{y})$ has a perfect matching $\bar{M}$, then STOP and return $\bar{M}$ as the min-cost perfect matching. Otherwise, find $A \subseteq X$ with $|A| > \left|N_{G^{=}(\bar{y})}(A)\right|$
3. **Step 2: Check feasibility** If $|A| > |N_G(A)|$, STOP and output "infeasible"
4. **Step 3: Update** Let ε be the minimum of $C(e) - \bar{y}_u - \bar{y}_v$ for all edges $e = uv$ with $u \in A$

and $v \in N_G(A) \backslash N_{G^=(\bar{y})}(A)$. Let

$$\bar{y}_v = \begin{cases} \bar{y}_v + \varepsilon & v \in A \\ \bar{y}_v - \varepsilon & v \in N_{G^=(\bar{y})}(A) \\ \bar{y}_v \end{cases}$$

Repeat from **STEP 1**

1. If the algorithm terminates, then the output is correct
2. If C is integer-valued, then the algorithm terminates.

R

1. There is a way to choose A such that the number of iterations is at most $|V(G)|^2$ (see CO450)
2. The min-cost perfect matching problem can be solved in polynomial-time (Edmonds, see CO450)

6.7.2 Matrix Rounding

How should one round the entries of a matrix?

1.3	2.1	0.4	0.3	4.1
2.8	0.6	0.7	1.6	5.7
4.1	2.7	1.1	1.9	9.8

Table 6.7.1: *Sum Table*

1	2	0	0	4
3	0	0	2	6
4	3	1	2	10

Table 6.7.2: *Excel Solution (WTF?)*

We can make all row and column sums to zero as we just want to know which one to round up or down.

1.3	2.1	0.4	0.3	-4.1
2.8	0.6	0.7	1.6	-5.7
-4.1	-2.7	-1.1	-1.9	9.8

Table 6.7.3: *Sum Table*

The following theorem can finish the problem.

Theorem 40 If $A \in \mathbb{R}^{m \times n}$ has row and column sums zero, then there is a matrix $\bar{A} \in \mathbb{Z}^{m \times n}$ with all row and column sums zero such that $\bar{a}_{ij} \in \{\lceil a_{ij} \rceil, \lfloor a_{ij} \rfloor\}$ for each $i \in \{1, \cdots, m\}$ and $j \in \{1, \cdots, n\}$.

6.7.3 Network Flow (CO351)

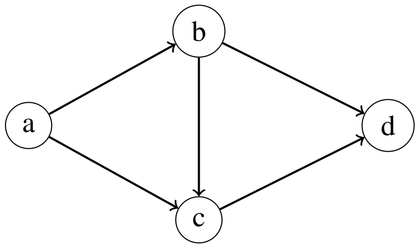


Figure 6.7.8: *Directed graph*

Definition 6.7.9 — Directed Graph. The **signed incidence matrix** is

$$\begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{ccccc} ab & bc & ac & bd & cd \end{array} \\ \begin{pmatrix} -1 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 \\ 0 & 1 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \end{array}$$

Corollary 6.7.3 The signed incidence matrix of a directed graph is totally unimodular.

Minimum-cost Flow Problems

Consider the problem:

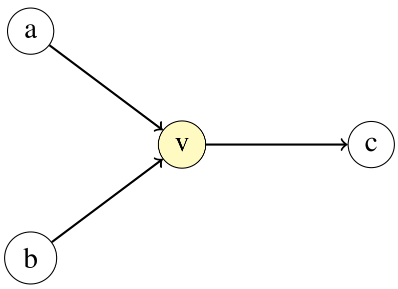
$$(P) \qquad \min(c^\top x : Ax = b, l \le x \le u)$$

where A is the signed incidence matrix of a directed graph $G = (V,E), b \in \mathbb{Z}^V, l, u \in \mathbb{Z}^E$ and $c \in \mathbb{R}^E$. Note that since A is TU, if there is an optimal solution, then there exists an optimal solution that is integer-valued.

Let x_{vw} denote the edge flow on from vertex v to w. Then, we can write (P) as

$$\min \sum_{vw \in A} c_{vw} x_{vw}$$

where c_{vw} is the unit cost for flow from v to w.

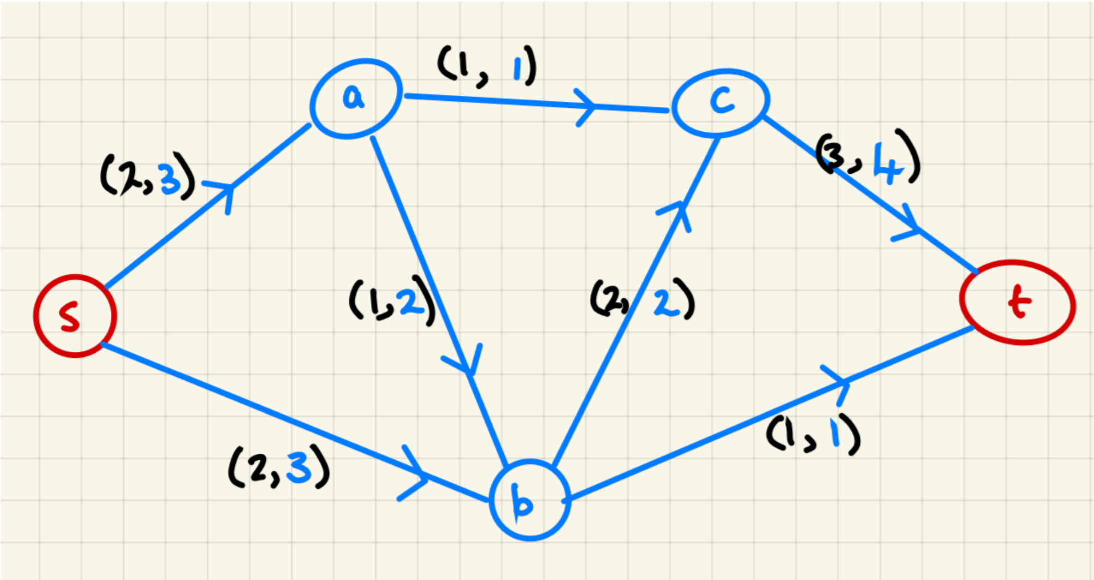


The net flow into v is $b_v = x_{av} + x_{bv} - x_{vc}$.

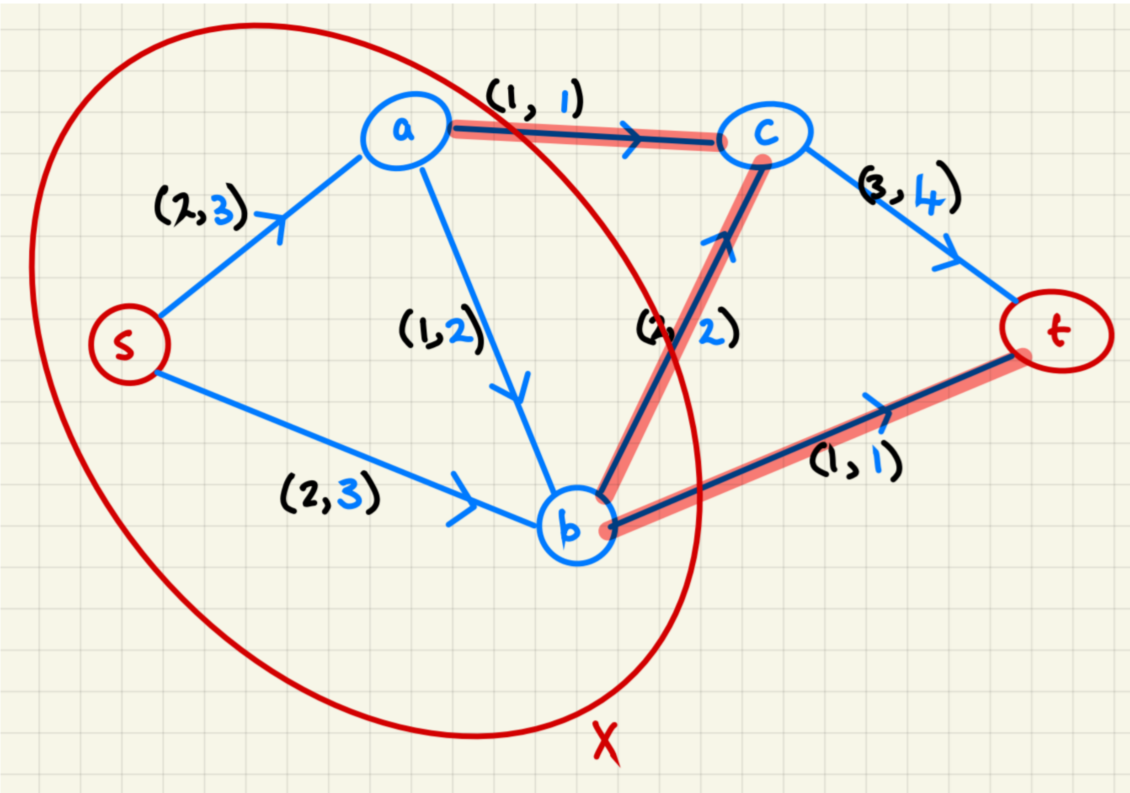
Maximum (s,t)−flow

A directed graph $G = (V,E)$ vertices $s,t \in V$, and edge capacities $u \in \mathbb{R}^E$ with $u \ge 0$. We need to find a feasible (s,t)−flow with maximum value.

Definition 6.7.10 — $(s,t)-$flow. $x \in \mathbb{R}^E$ is an $(s,t)-$flow if each vertex in $V \setminus \{s,t\}$ has net-flow zero. An $(s,t)-$flow x is feasible if $0 \le x \le u$. The value of an $(s,t)-$flow x is the net-flow into t.



Is this a maximum $(s,t)-$flow? We can consider the following (s,t)-cut with capacity 4.



For $X \subseteq V$, we define $OUT(X) = \{vw \in E : v \in X, w \notin X\}$. If $s \in X$ and $t \notin X$, then we call $OUT(X)$ an $(s,t)-$cut. The capacity of the cut is

$$u(OUT(X)) := \sum (u_{vw} : vw \in OUT(X))$$

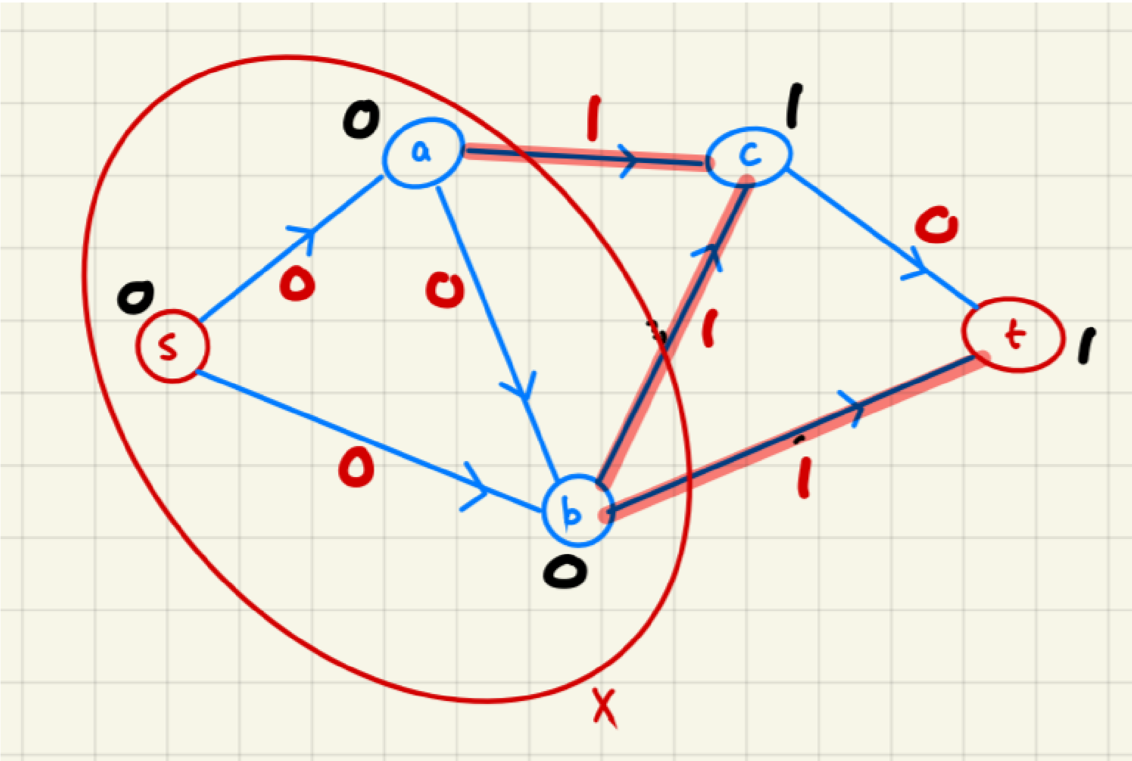
Theorem 41 — Max-flow Min-cut Theorem. The maximum value of an $(s,t)-$flow is equal to the minimum capacity of an $(s,t)-$cut.

Formulation for min-capacity (s,t)-cut For $u \geq 0$,

$$(IP)\begin{cases} \min & \sum_{e\in E} u_e z_e \\ s.t. & z_{vw} \geq y_w - y_v \qquad vw \in E \\ & y_t = 1 \\ & y_s = 0 \\ & 0 \leq y \leq 1 \\ & z \geq 0 \\ & y \in \mathbb{Z}^V, z \in \mathbb{Z}^E \end{cases}$$

that is

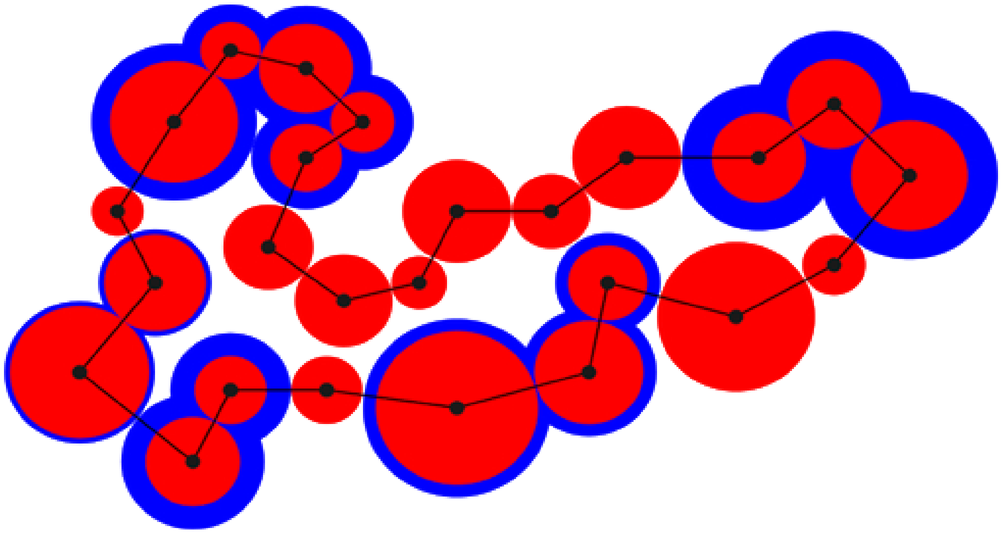
$$\min(u^\top z : z \geq A^\top y, 0 \leq y \leq 1, z \geq 0, y_s = 0, y_t = 1, y \in \mathbb{Z}^V, z \in \mathbb{Z}^E)$$



The linear relaxation is

$$(LP) \qquad \min\left(u^\top z : z \geq A^\top y, 0 \leq y \leq 1, z \geq 0, y_s = 0, y_t = 1\right)$$

since A is TU, we have $\mathbf{OPT}(IP) = \mathbf{OPT}(LP)$.



7. Convex Optimization

Convex optimization involves solving

$$(P) \qquad \min(c^\top x : x \in S)$$

where $S \subseteq \mathbb{R}^n$ is a closed convex set and $c \in \mathbb{R}^n$.
We have seen that

1. Any optimization problem can be reduced to this form
2. These problems are NP-hard

Our goals are

1. find optimally conditions
2. give an "efficient" algorithm

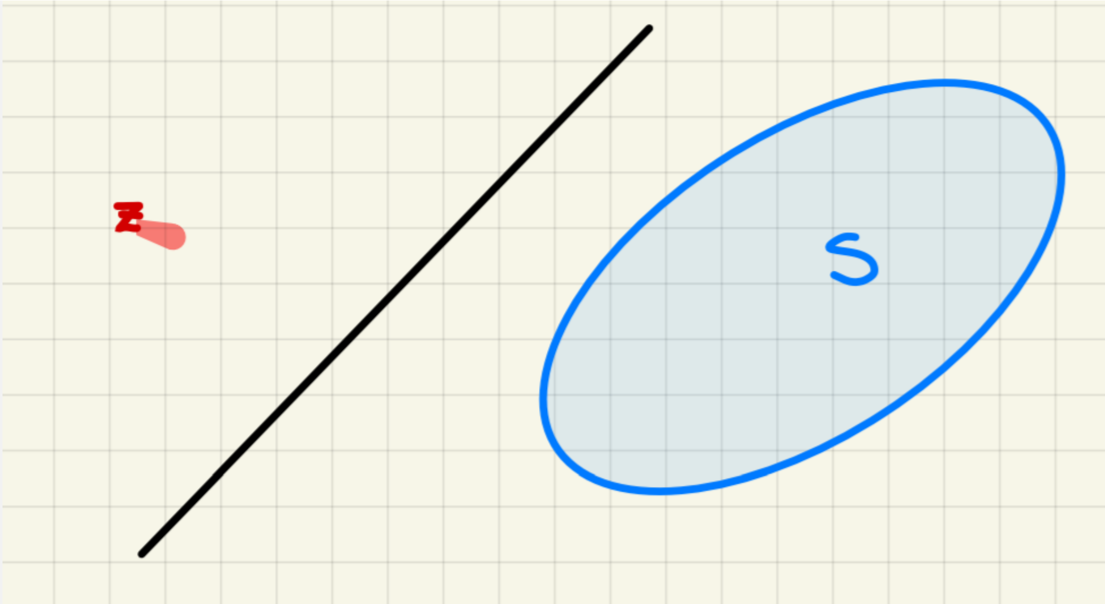
7.1 Convex Topology

Definition 7.1.1 — Some topologies. A set $S \subseteq \mathbb{R}^n$ is closed if each convergent sequence in S converges to an element of S. S is bounded if $S \subseteq [-N, N]^n$ for some $N \subseteq \mathbb{R}$. A set that is closed and bounded is compact.

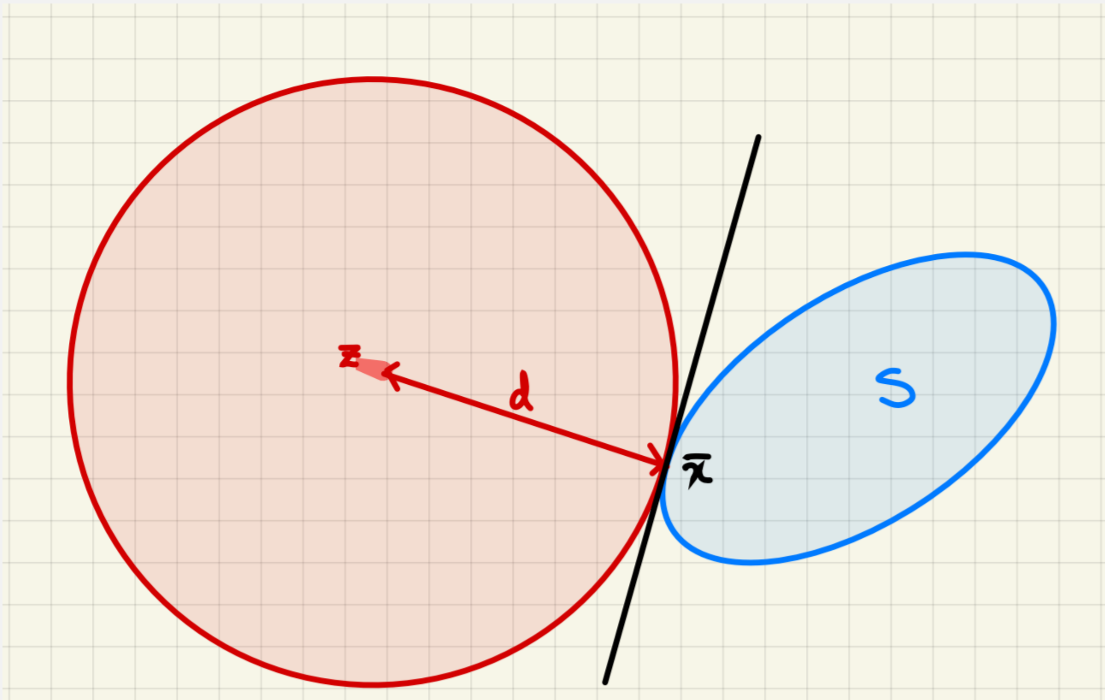
Theorem 42 — Bolzano-Weierstrass Theorem. Every bounded infinite sequence in $\mathbb{R}^n$ has a convergent subsequence.

Corollary 7.1.1 — Weierstrass Theorem. If $S \subseteq \mathbb{R}^n$ is non-empty and compact and $f : S \to \mathbb{R}$ is continuous, then there exists $\bar{x} \in S$ minimizing $\{f(x) : x \in S\}$.

Theorem 43 — Separating Hyperplane Theorem. Let $S \subseteq \mathbb{R}^n$ be a closed convex set and $z \in \mathbb{R}^n$. Then, $z \notin S$ if and only if there exist $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $a^\top x \leq b$ for all $x \in S$ and $a^\top z > b$.



Proof. We may assume that $S \neq \emptyset$ and $z \notin S$. By Weierstrass's Theorem, there exists a nearest point $\bar{x} \in S$ to z. We can do this since $B(z, \|\bar{x} - z\|_2) \cap S$ is a compact set for any $\bar{x} \in S$. Let $a = z - \bar{x}$ and $b = a^\top \bar{x}$ will work.



$\blacksquare$

Definition 7.1.2 — Hyperplane. A hyperplane H is a non-empty set of the form $H = \{x \in \mathbb{R}^n : a^\top x = b\}$ where $a \in \mathbb{R}^n \backslash \{0\}$ and $b \in \mathbb{R}$. H separates z from S if the half space $\{x \in \mathbb{R}^n : a^\top x \leq b\}$ contains S but not z.

(R) The separating hyperplane theorem is equivalent to

Every closed convex set is the intersection of closed half-spaces.

In particular, polyhedra are the intersection of finitely many closed half-spaces.

Definition 7.1.3 — More topologies. A point $x \in \mathbb{R}^n$ is in the interior of a set $S \subseteq \mathbb{R}^n$, if there exists $\varepsilon > 0$ such that $B(x,r) \subseteq S$. The interior of S is denoted $int(S)$. The boundary of a closed set S is $\partial S = S \backslash int(S)$.
A hyperplane $H = \left\{x \in \mathbb{R}^n : a^\top x = b\right\}$ is a supporting hyperplane for S if $H \cap S \neq \emptyset$ and S is contained in either $\left\{x \in \mathbb{R}^n : a^\top x \leq b\right\}$ or $\left\{x \in \mathbb{R}^n : a^\top x \geq b\right\}$.

Theorem 44 Each point in the boundary of a closed convex set is contained in a supporting hyperplane.

Proof. If $\bar{x}$ is in the boundary, then there is a sequence $(x_n)_n$ of points in $\mathbb{R}^n \backslash S$ that converges to $\bar{x}$. ■

R This is equivalent to

each boundary point is optimal for some non-zero, linear objective function.

Back to the original convex optimization problem, **how is S given?**

1. **Separation Oracle:** given $\bar{x} \in \mathbb{R}^n$, the oracle either tells that $\bar{x} \notin S$ or returns a separating hyperplane.
2. **Intersection of "nice" convex sets:**

$$S = \bigcap_{\lambda \in \Lambda} S_\lambda$$

where $\{S_\lambda\}_{\lambda \in \Lambda}$ are "nice" closed convex sets.

Linear programs have both of these properties inherently.

7.2 Optimality Conditions

Given $\bar{x} \in S$, is $\bar{x}$ optimal

Definition 7.2.1 — Tangent Cone. Let S be a closed convex set and $\bar{x} \in S$. $T(S,\bar{x}) := \overline{\mathbf{cone}(S - \{\bar{x}\})}$ is the tangent cone of S at $\bar{x}$.

Theorem 45 $\bar{x}$ is optimal for (P) if and only if 0 is optimal for $\min(c^\top x : x \in T(S,\bar{x}))$

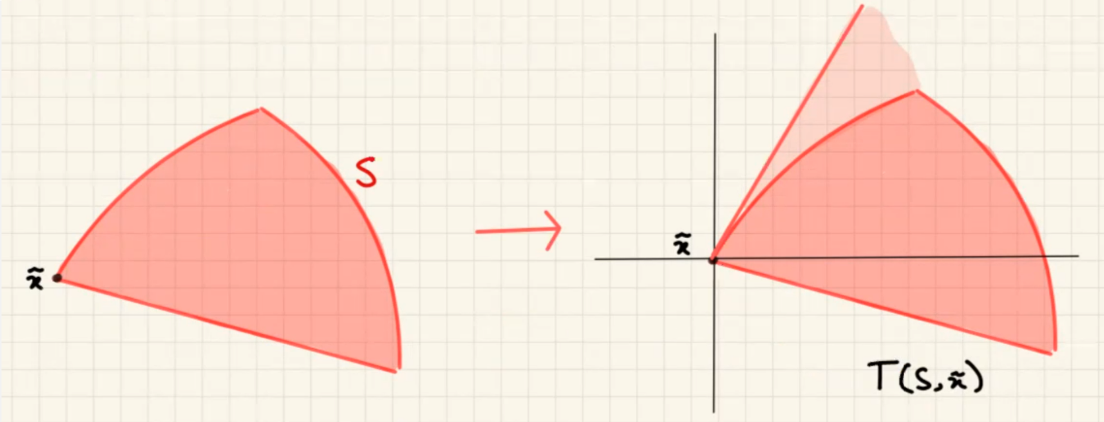


Figure 7.2.1: *Tangent Cone*

■ **Example 7.1 — LP Case.** For the problem (P) $\min(c^\top x : Ax \geq b)$. Let $\bar{x}$ be a feasible solution, and let $A^= x \geq b^=$ be the equality subsystem.

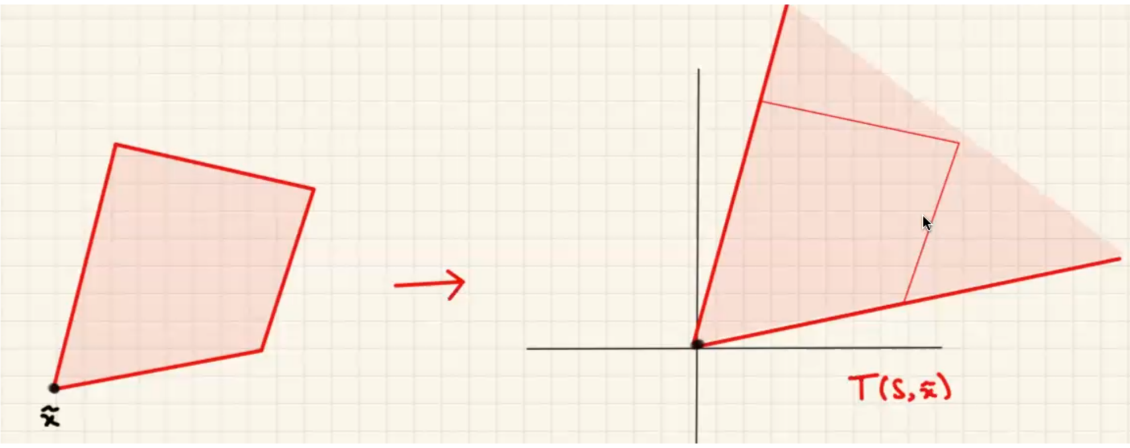


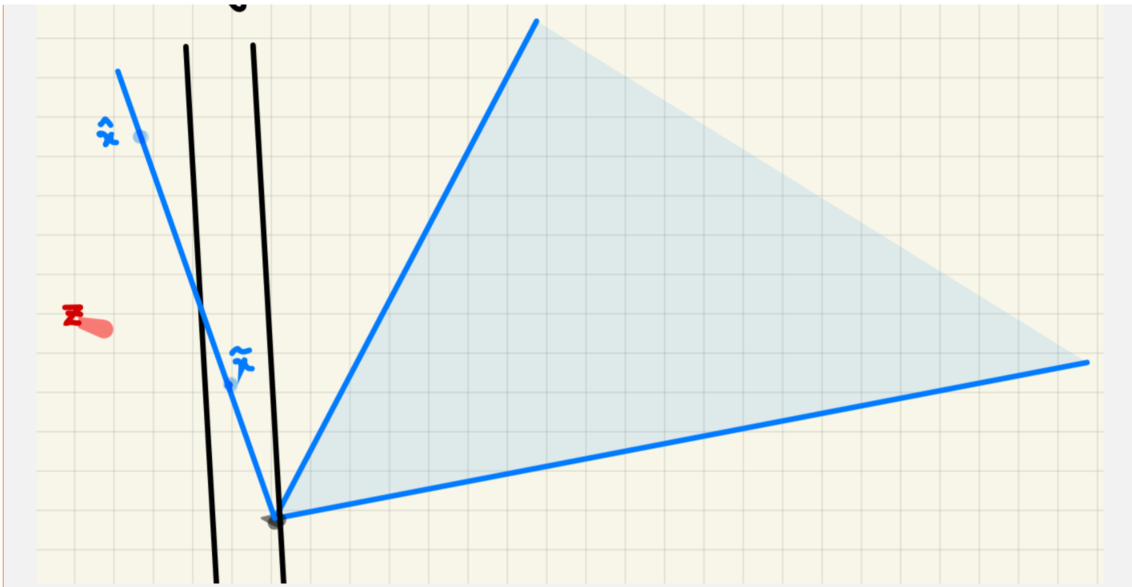
Figure 7.2.2: *LP Tangent Cone*

$$T(S, \bar{x}) = \{x \in \mathbb{R}^n : Ax \geq 0\}$$

■

Theorem 46 — Separating Hyperplane Theorem for Cones. Let $K \subseteq \mathbb{R}^n$ be a closed cone and $z \in \mathbb{R}^n$. Then, $z \notin K$ if and only if there exists $a \in \mathbb{R}^n$ such that $a^\top x \leq 0$ for all $x \in K$ and $a^\top z > 0$.

Proof. We may assume that $z \notin K$. By the Separating hyperplane theorem, there exists $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $a^\top x \geq b$ for each $x \in K$ and $a^\top z < b$. Since $0 \in K, b \leq 0$. Now, $a^\top z < b \leq 0$. It suffices to prove that $a^\top x \geq 0$ for each $z \in K$. Suppose that there exists $\bar{x} \in K$ such that $a^\top \bar{x} < 0$. Then, we can scale to get $\hat{x} \in K$ with $a^\top \hat{x} < b$. This yields a contradiction.



Equivalently: every closed cone is the intersection of closed half-spaces containing 0 in the boundary.

7.2.1 Duality for Cones

For $S \subseteq \mathbb{R}^n$, define $S^* := \{c \in \mathbb{R}^n : c^\top x \geq 0, \forall x \in S\}$

> **R** If $0 \in S$, then s^* is the set of all $c \in \mathbb{R}^n$ such that 0 minimizes $(c^\top x : x \in S)$.

Example 7.2

1. $\{x \in \mathbb{R}^n : x \geq 0\}^* = \{x \in \mathbb{R}^n : x \geq 0\}$
2. $\{x \in \mathbb{R}^n : c^\top \geq 0\}^* = \mathbf{cone}(\{c\})$
3. $B([1,0]^\top, 1)^* = \{[\lambda, 0]^\top : \lambda \geq 0\}$
4. For $A \in \mathbb{R}^{m \times n}$, $\{x \in \mathbb{R}^n : Ax \geq 0\}^* = \{A^\top y : y \in \mathbb{R}^m, y \geq 0\}$

For $a \in \mathbb{R}^n$, let $H(a) = \{x \in \mathbb{R}^n : a^\top x \geq 0\}$. By definition,

$$S^* = \bigcap (H(x) : x \in S)$$

Theorem 47 S^* is a closed cone.

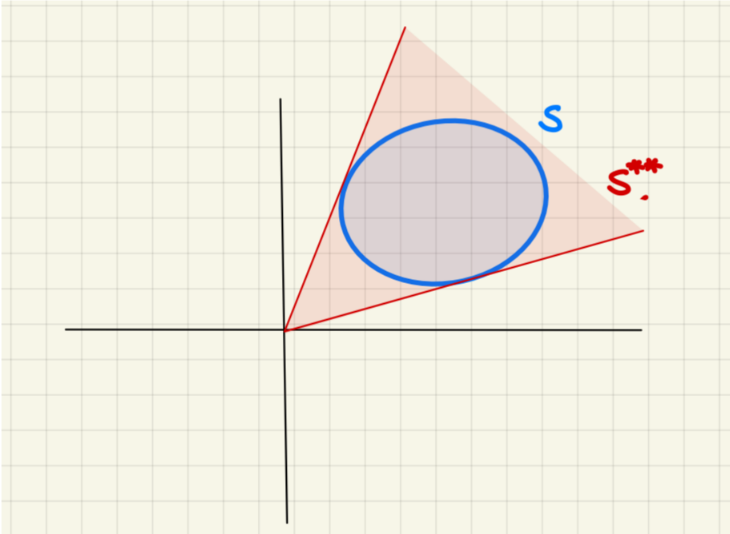
Proof. S^* is the intersection of closed half-spaces. ■

Note that $S^* = \{a \in \mathbb{R}^n : S \subseteq H(a)\}$.

Lemma 7.3 For any $S \subseteq \mathbb{R}^n$, $S^{**} = \cap(H(a) : a \in \mathbb{R}^n, S \subseteq H(a))$.

Proof.

$$S^{**} = \cap(H(a) : a \in S^*) = \cap(H(a) : a \in \mathbb{R}^n, S \subseteq H(a))$$



■

Theorem 48 — Duality Theorem for Cones. If K is a closed cone, then $K^{**} = K$.

Proof. By the lemma,

$$K^{**} = \cap(H(a) : a \in \mathbb{R}^n, K \subseteq H(a))$$

By the separating hyperplane theorem for cones, we have

$$K = \cap(H(a) : a \in \mathbb{R}^n, K \subseteq H(a))$$

■

Definition 7.3.1 — Dual Cone. If $K \subseteq \mathbb{R}^n$ is a closed cone, then K^* is the dual of K.

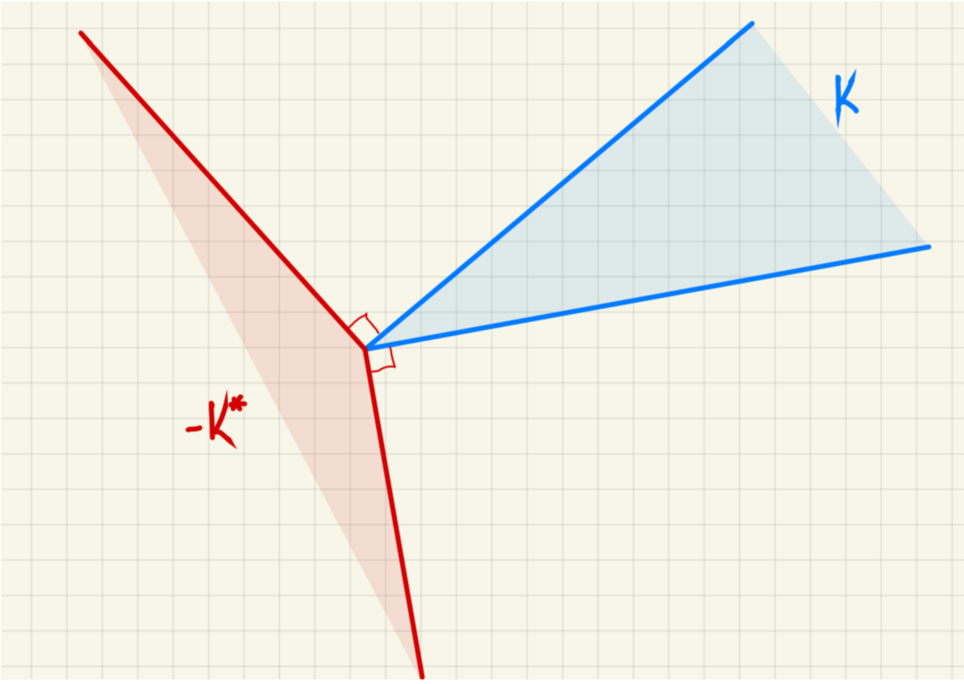


Figure 7.3.1: $-K^*$ *and* K

Theorem 49 For any $S \subseteq \mathbb{R}^n$, $S^{**} = \overline{\textbf{cone}(S)}$.

Proof.

$$
\begin{aligned}
S^{**} &= \cap(H(a) : a \in \mathbb{R}^n, S \subseteq H(a)) \\
&= \cap(H(a) : a \in \mathbb{R}^n, \overline{\textbf{cone}(S)} \subseteq H(a)) \\
&= \overline{\textbf{cone}(S)}^{**} \\
&= \overline{\textbf{cone}(S)}
\end{aligned}
$$

■

■ **Example 7.3**

$$K = \{x \in \mathbb{R}^n : x \geq 0\}$$

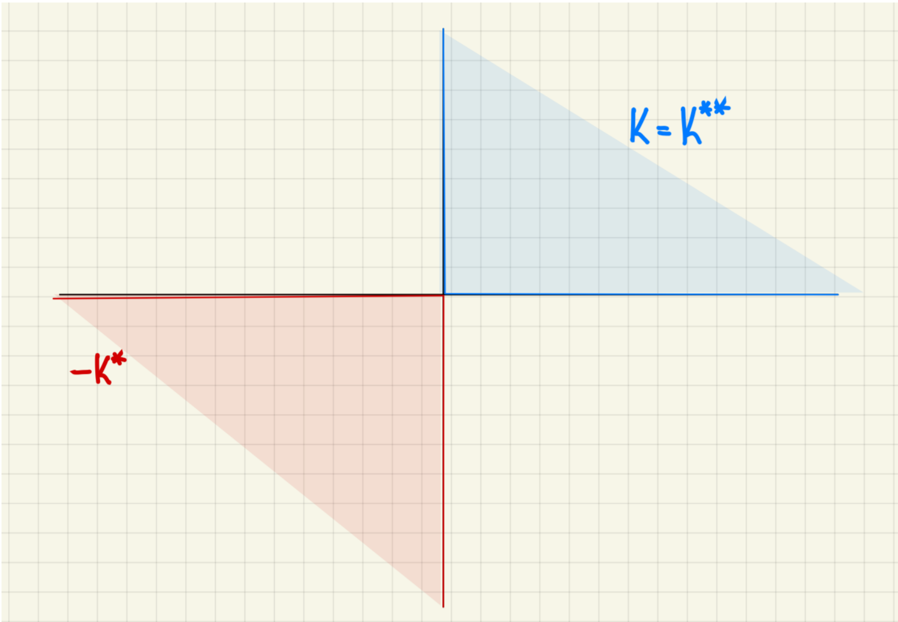


Figure 7.3.2: $K = K^*$

■

■ **Example 7.4**

$$K = H(a)$$

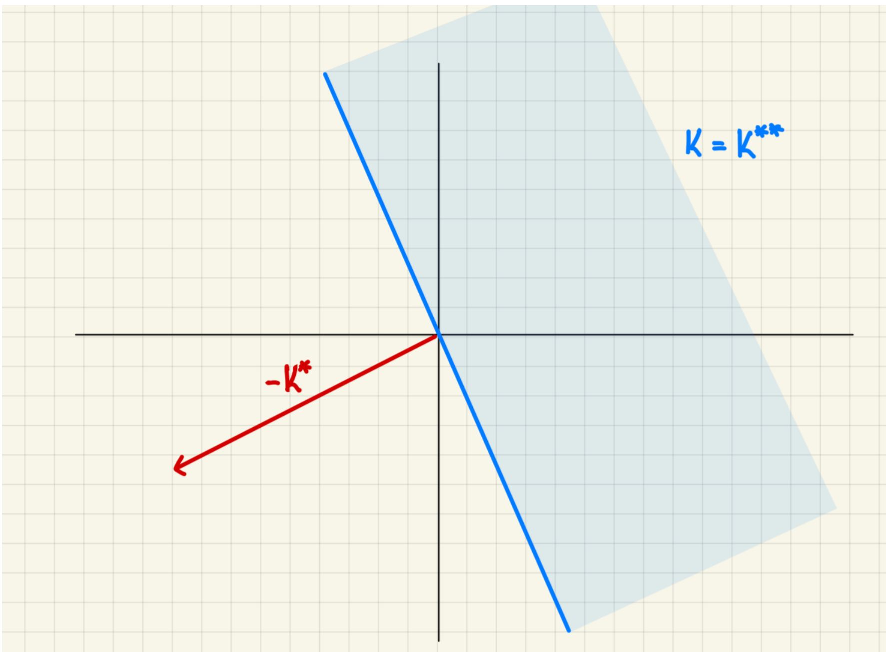


Figure 7.3.3: $K^* = \boldsymbol{cone}(\{a\})$

■

■ **Example 7.5**

$$S = B\left([1,0]^\top, 1\right)$$

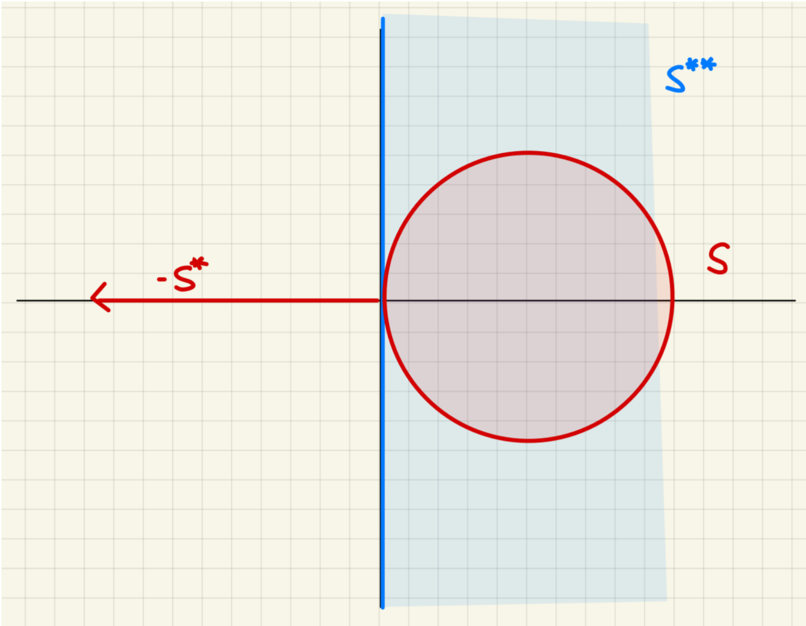
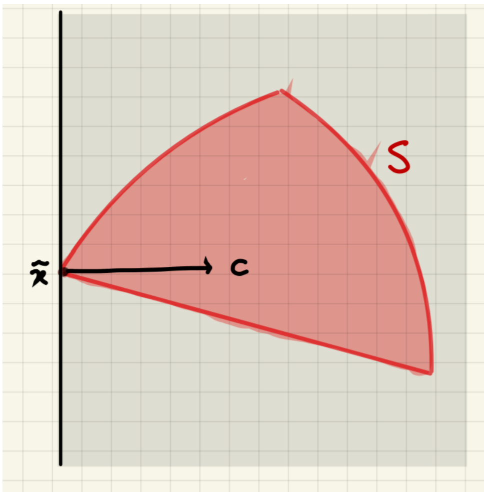


Figure 7.3.4: $S^* = \textbf{\textit{cone}}(\{[1,0]^\top\})$

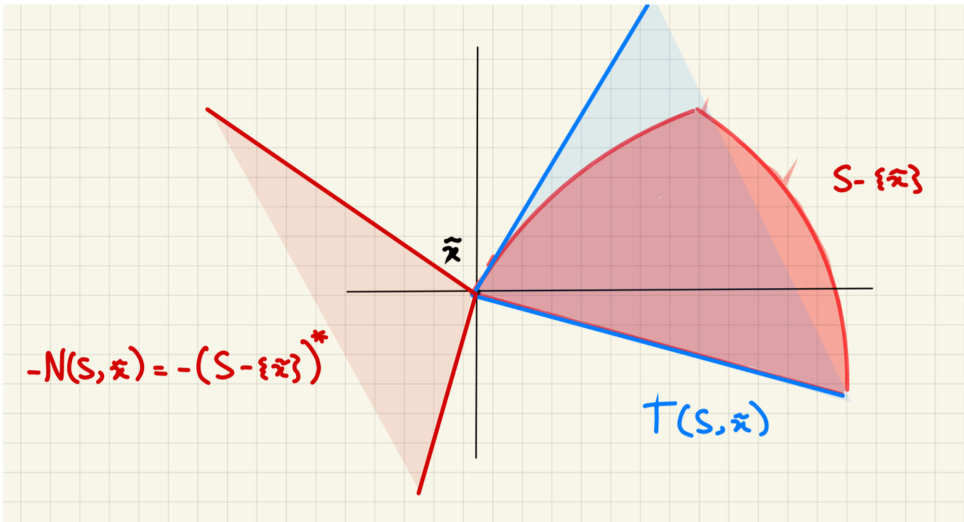
■

Definition 7.3.2 — Normal Cone. The normal cone to S at $\bar{x}$, denoted by $N(S,\bar{x})$, is the set of $c \in \mathbb{R}^n$ such that $\bar{x}$ minimizes $(c^\top x : x \in S$.



That is,

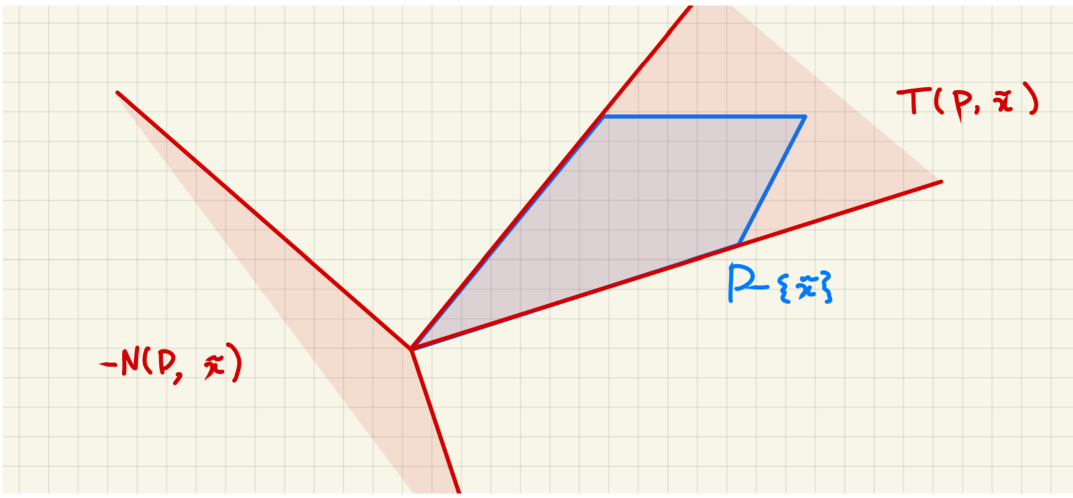
$$\begin{aligned} N(S,\bar{x}) &= \left\{c \in \mathbb{R}^n : c^\top x \geq c^\top \bar{x}, \forall x \in S\right\} \\ &= \left\{c \in \mathbb{R}^n : c^\top x \geq 0, \forall x \in S - \{\bar{x}\}\right\} \\ &= (S - \{\bar{x}\})^* \end{aligned}$$



We note that

1. $N(S,\bar{x})$ is a closed cone
2. $N(S,\bar{x}) = T(S,\bar{x})^*$ and $T(S,\bar{x}) = N(S,\bar{x})^*$

■ **Example 7.6 — LP Case.** Let $P = \{x \in \mathbb{R}^n, Ax \geq b\}$ and $\bar{x} \in P$. Let $A^= x \geq b^=$ be the equality subsystem for $\bar{x}$.



We note that

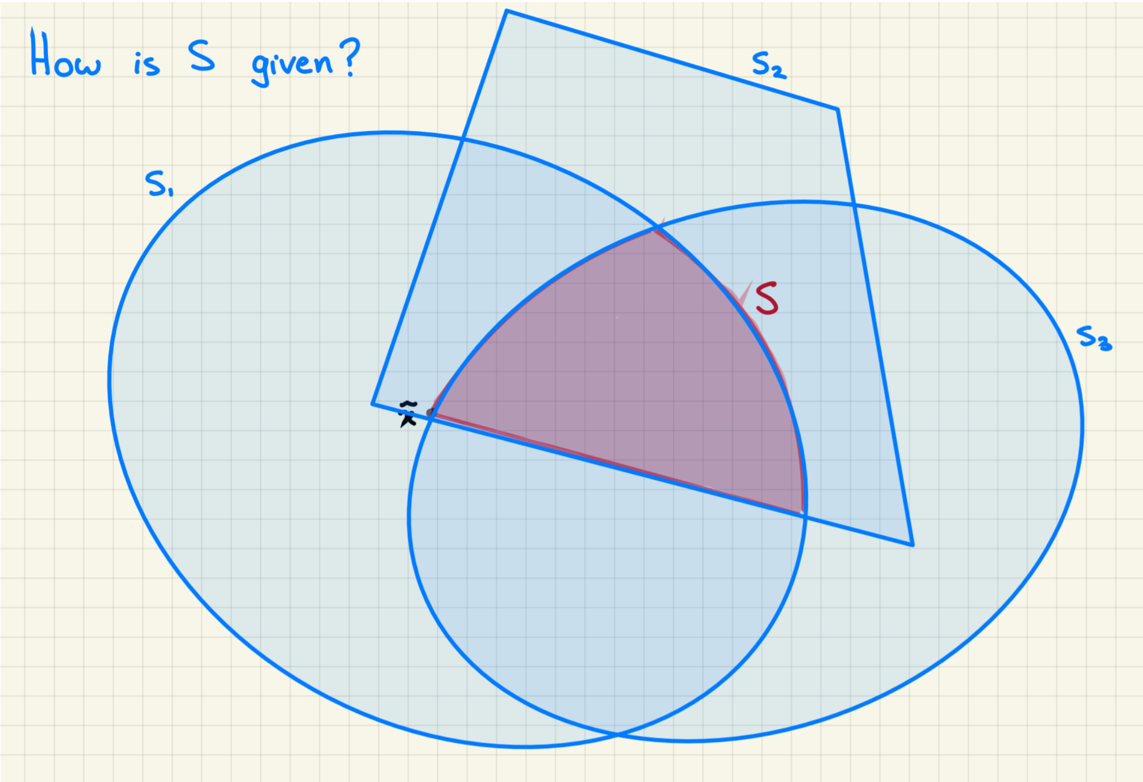
1. $T(P,\bar{x}) = \{x \in \mathbb{R}^n : A^= x \geq 0\}$
2. $N(P,\bar{x}) = \left\{(A^=)^\top y : y \in \mathbb{R}^n, y \geq 0\right\}$: complementary slackness theorem

■

So, how do we certify optimality conditions?

$$\bar{x} \text{ is optimal} \iff c \in N(S,\bar{x})$$

How is S given?



Suppose that $S = S_1 \cap \cdots \cap S_m$ where $S_1, \cdots, S_m$ are "nice" closed convex sets. Here nice means that we already know $N(S, \bar{x})$.

■ **Example 7.7 — Closed half-spaces are nice.** $H = \left\{x \in \mathbb{R}^n : a^\top x \geq b\right\}$ where $a \in \mathbb{R}^n \backslash \{0\}$ and $b \in \mathbb{R}$. For $\bar{x} \in H$, we have

$$N(H, \bar{x}) = \begin{cases} \{0\} & a^\top \bar{x} > b \\ \mathbf{cone}(\{a\}) & a^\top \bar{x} = b \end{cases}$$

■

■ **Example 7.8 — Balls are nice.** $B = B(c, r)$ where $c \in \mathbb{R}^n, r \geq 0$. Let $\bar{x} \in B$,

$$N(B, \bar{x}) = \begin{cases} \{0\} & \|\bar{x} - c\| < r \\ \mathbf{cone}(\{c - \bar{x}\}) & \|\bar{x} - c\| = r \end{cases}$$

■

Recall

$$(P) \qquad \min(c^\top x : x \in S_1 \cap \cdots \cap S_m)$$

where $c \in \mathbb{R}^n$ and $S_1, \cdots, S_m \subseteq \mathbb{R}^n$.

Theorem 50 — Weak cost-splitting theorem. Let $\bar{x} \in S_1 \cap \cdots \cap S_m$. If there exist $c_1, \cdots, c_m \in \mathbb{R}^n$ such that $c = c_1 + \cdots + c_m$ and $\bar{x}$ minimizes $(c_i^\top x : x \in S_i)$, for each $i \in \{1, \cdots, m\}$, then $\bar{x}$ minimizes (P).

Equivalently: if $S = S_1 \cap \cdots \cap S_m$ and $\bar{x} \in S$, then

$$N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x}) \subseteq N(S, \bar{x})$$

However, do we always get

$$N(S_1,\bar{x})+\cdots+N(S_m,\bar{x})=N(S,\bar{x})$$

NO!

■ **Example 7.9** Let $B_1=B([-1,0]^\top,1), B_2=B([1,0]^\top,1)$ and $\bar{x}=0\in B_1\cap B_2$. We know that

$$N(B_1,\bar{x})=\mathbf{cone}(\left\{[-1,0]^\top\right\}) \qquad N(B_1,\bar{x})=\mathbf{cone}(\left\{[1,0]^\top\right\})$$

then,

$$N(B_1,\bar{x})+N(B_2,\bar{x})=\left\{[a,0]^\top : a\in\mathbb{R}\right\}$$

However,

$$N(B_1\cap B_2,\bar{x})=\mathbb{R}^2\neq N(B_1,\bar{x})+N(B_2,\bar{x})$$

■

For us to certify optimality of $\bar{x}$, we need

$$N(S_1,\bar{x})+\cdots+N(S_m,\bar{x})=N(S,\bar{x})$$

Lemma 7.4 For $S_1,\cdots,S_m\subseteq\mathbb{R}^n$, then

$$S_1\cap\cdots\cap S_m\subseteq(S_1^*+\cdots+S_m^*)^*$$

Proof. Let $x\in S_1\cap\cdots\cap S_m$ and $y\in S_1^*+\cdots+S_m^*$. Now, $y=y_1+\cdots+y_m$ where $y_i\in S_i^*$. Since $x\in S_i$, we have $x^\top y_i\geq 0$. Thus,

$$x^\top y=x^\top y_1+\cdots+x^\top y_m\geq 0$$

Therefore, $x\in(S_1^*+\cdots+S_m^*)^*$. ■

Lemma 7.5 If $K_1,\cdots,K_m\subseteq\mathbb{R}^n$ are closed cones, then $K_1\cap\cdots\cap K_m=(K_1^*+\cdots+K_m^*)^*$

Proof. By the previous lemma, we have $K_1\cap\cdots\cap K_m\subseteq(K_1^*+\cdots+K_m^*)^*$. It suffices to show that, if $\bar{x}\in\mathbb{R}^n\backslash(K_1\cap\cdots\cap K_m)$, then $\bar{x}\notin(K_1^*+\cdots+K_m^*)^*$. Since $\bar{x}\notin K_1\cap\cdots\cap K_m$, there exists $i\in\{1,\cdots,m\}$ such that $\bar{x}\notin K_i$. By the Separating Hyperplane theorem for cones, there exists $a\in\mathbb{R}^n\backslash\{0\}$ such that $a^\top\bar{x}<0$ but $a^\top x\geq 0$ for each $x\in K_i$. By definition, $a\in K_i^*$. Then since each cone K_j^* contains 0, $a\in K_1^*+\cdots+K_m^*$. Then, since $\bar{x}^\top a<0$, thus, $\bar{x}\notin(K_1^*+\cdots+K_m^*)^*$ as required. ■

Theorem 51 If $K_1,\cdots,K_m\subseteq\mathbb{R}^n$ are closed cones, then

$$(K_1\cap\cdots\cap K_m)^*=\overline{K_1^*+\cdots+K_m^*}$$

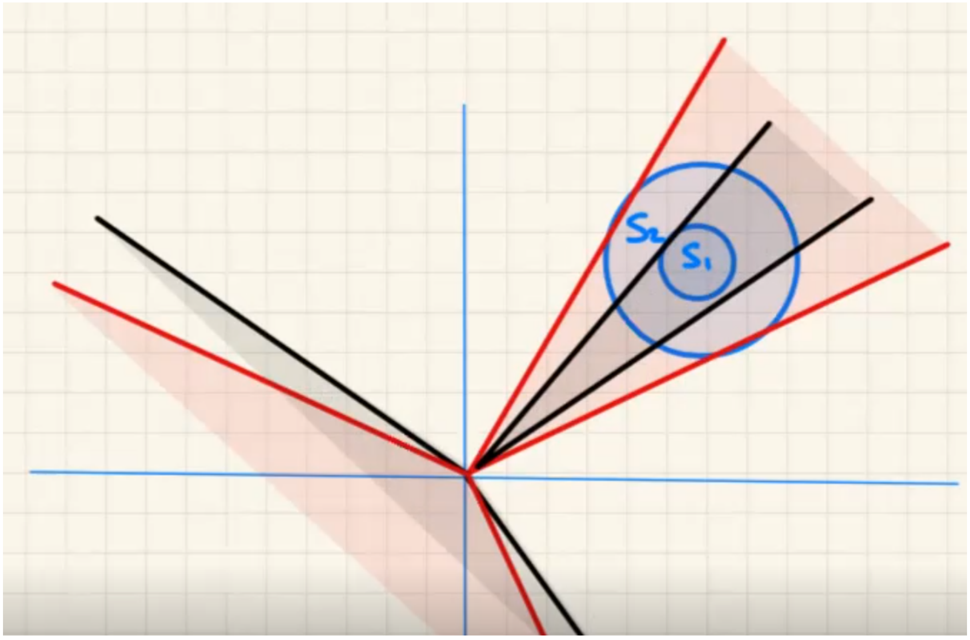
Proof. By the second lemma,

$$\begin{aligned}(K_1\cap\cdots\cap K_m)^* &= (K_1^*+\cdots+K_m^*)^{**}\\ &= \overline{\mathbf{cone}(K_1^*+\cdots+K_m^*)}\\ &= \overline{K_1^*+\cdots+K_m^*}\end{aligned}$$

■

Lemma 7.6 If $S_1\subseteq S_2\subseteq\mathbb{R}^n$, then $S_2^*\subseteq S_1^*$.

Proof. Consider $a \in S_2^*$. Since $S_1 \subseteq S_2$, we have $a^\top x \geq 0$ for all $x \in S_1$. Thus, $a \in S_1^*$.



■

Necessary and Sufficient Conditions for Optimality

Theorem 52 If $S = S_1 \cap \cdots \cap S_m$ where $S_1, \cdots, S_m \subseteq \mathbb{R}^n$ and $\bar{x} \in S$, then $N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x}) \subseteq N(S, \bar{x})$. Moreover equality holds if and only if

1. $T(S, \bar{x}) = T(S_1, \bar{x}) \cap \cdots \cap T(S_m, \bar{x})$, and
2. $N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x})$ is closed.

Proof.

$$\begin{aligned} N(S, \bar{x}) &= T(S_1 \cap \cdots \cap S_m, \bar{x})^* \\ &\supseteq (T(S_1, \bar{x}) \cap \cdots \cap T(S_m, \bar{x}))^* \\ &= \overline{T(S_1, \bar{x})^* + \cdots + T(S_m, \bar{x})^*} \\ &= \overline{N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x})} \\ &\supseteq N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x}) \end{aligned}$$

■

Relation to Linear Optimiation

Let $P = H_1 \cap \cdots \cap H_m$ where $H_i = \{x \in \mathbb{R}^n : a_i^\top x \geq b_i\}$ where $a_i \in \mathbb{R}^n$ and $b_i \in \mathbb{R}$ for $i \in \{1, \cdots, m\}$. Let $\bar{x} \in P$ and let $D = \{i \in \{1, \cdots, m\} : a_i^\top x = b_i\}$. Then,

$$T(H_i, \bar{x}) = \begin{cases} \mathbb{R}^n & i \notin D \\ H_i - \{\bar{x}\} & i \in D \end{cases}$$

we have

$$T(P, \{\bar{x}\}) = T(H_1, \{\bar{x}\}) \cap \cdots \cap T(H_m, \bar{x}) \tag{1}$$

also,

$$N(H_i, \bar{x}) = \begin{cases} \{0\} & i \notin D \\ \mathbf{cone}(\{a_i\}) & i \in D \end{cases}$$

then,

$$N(H_1,\bar{x})+\cdots+N(H_m,\bar{x}) = \mathbf{cone}(\{a_i, i \in D\})$$

, so

$$(2) \qquad N(H_1,\bar{x})+\cdots+N(H_m,\bar{x})$$

is closed. (The cone of a finite set is a polyhedron). Then,

$$N(P,\bar{x}) = N(H_1,\bar{x})+\cdots+N(H_m,\bar{x})$$

which is **Strong Duality Theorem**.

■ **Example 7.10 — $T(S,\bar{x}) = T(S_1,\bar{x})\cap\cdots\cap T(S_m,\bar{x})$ does not hold.** Let $B_1 = B([-1,0]^\top,1), B_2 = B([1,0]^\top,1)$ and $\bar{x} = 0 \in B_1 \cap B_2$. We know that

$$T(B,\bar{x})\cap T(B_2,\bar{x}) = \{x\in\mathbb{R}^2 : x_1 \le 0\}\cap\{x\in\mathbb{R}^2 : x_1 \ge 0\} = \{x\in\mathbb{R}^2 : x_1 = 0\}$$

while

$$T(B_1\cap B_2,\bar{x}) = \{0\} \neq T(B,\bar{x})\cap T(B_2,\bar{x})$$

■

■ **Example 7.11 — Sum of closed sets need not to be closed.** Consider

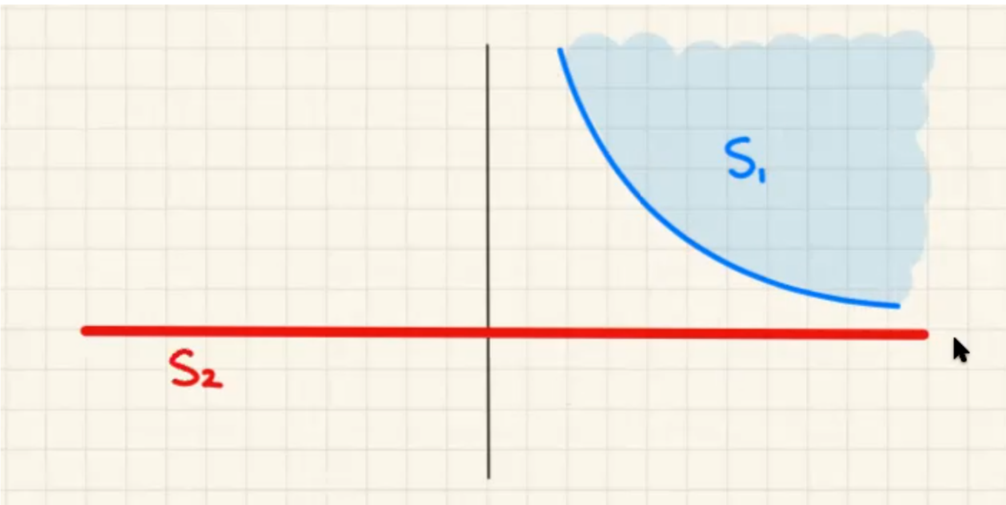
$$S_1 = \left\{x\in\mathbb{R}^2 : x_1 > 0, x_2 \ge \frac{1}{x_1}\right\}$$

$$S_2 = \{x\in\mathbb{R}^2 : x_2 = 0\}$$

then,

$$S_1 + S_2 = \{x\in\mathbb{R}^2 : X_2 > 0\}$$

S_1, S_2 are closed, but S_1+S_2 is not.



■

Exercise 7.1 Give examples of closed cones $K_1, K_2 \subseteq \mathbb{R}^n$ such that K_1, K_2 are not closed. ■

Exercise 7.2 Prove that if $S_1, S_2 \subseteq \mathbb{R}^n$ are compact sets, then $S_1 + S_2$ is compact. ■

Theorem 53 — Main Theorem. Let $S = S_1 \cap \cdots \cap S_m$ where $S_1, \cdots, S_m \subseteq \mathbb{R}^n$ are closed convex sets and let $\bar{x} \in S$. If the interior of S is non-empty, then

$$N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x}) = N(S, \bar{x})$$

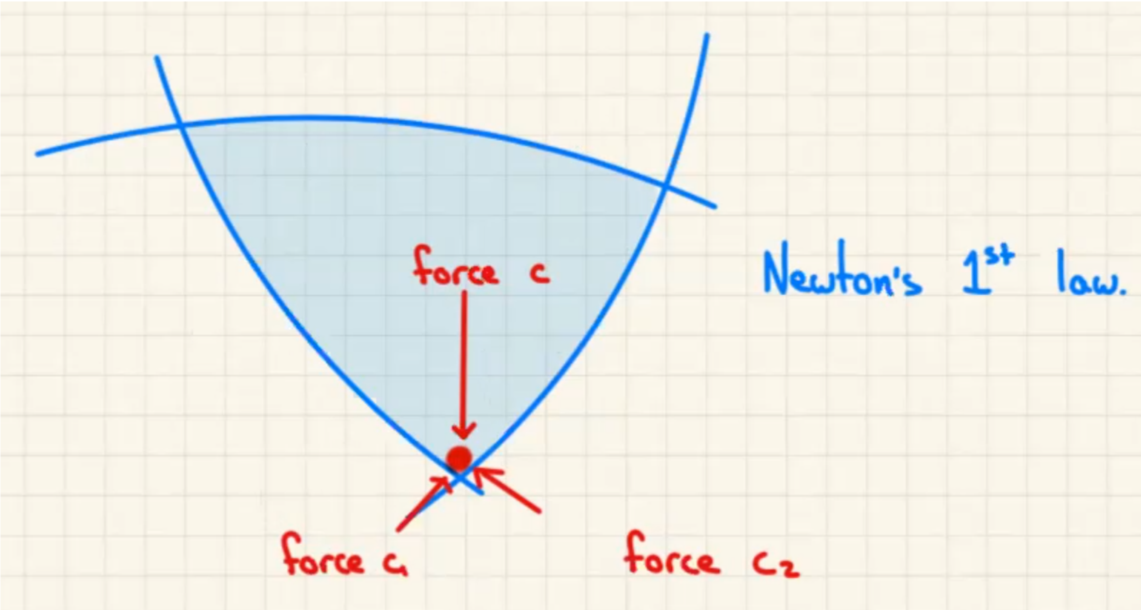


Figure 7.6.1: *Physical Interpretation*

Recall that if $int(S) = \emptyset$, then there is a hyperplane in $\mathbb{R}^n$ that contains S.

R For convex sets $S_1, S_2 \subseteq \mathbb{R}$, $\overline{S_1 \cap S_2}$ can be different from $\overline{S_1} \cap \overline{S_2}$.

■ **Example 7.12** $S_1 = \{x \in \mathbb{R}^2 : x_1 \geq 0\}$ and $S_2 = \{x \in \mathbb{R}^2 : x_1 < 0\}$. Then, $\overline{S_1 \cap S_2} = S_1 \cap S_2 = \emptyset$. But $\overline{S_1} \cap \overline{S_2} = \{x \in \mathbb{R}^n : x_1 = 0\}$. ■

Lemma 7.7 If $S_1, S_2 \subseteq \mathbb{R}^n$ are convex sets and $int(S_1 \cap S_2) \neq \emptyset$, then $\overline{S_1 \cap S_2} = \overline{S_1} \cap \overline{S_2}$.

Recall that $T(\bar{x}, S) = \overline{\mathbf{cone}(S - \{\bar{x}\})}$.

Theorem 54 Let $S = S_1 \cap \cdots \cap S_m$ where $S_1, \cdots, S_m \subseteq \mathbb{R}^n$ are closed convex sets and let $\bar{x} \in S$. If the interior of S is non-empty, then

$$T(S, \bar{x}) = T(S_1, \bar{x}) \cap \cdots \cap T(S_m, \bar{x})$$

Proof. We shall do the case for $m = 2$ and let induction do the work. We can translate so that $\bar{x} = 0$. Note that, since S_1, S_2 are convex and $0 \in S_1 \cap S_2$. Then, $\mathbf{cone}(S_1 \cap S_2) = \mathbf{cone}(S_1) \cap \mathbf{cone}(S_2)$.

Since $int(S_1 \cap S_2) \neq \emptyset$, $int(\mathbf{cone}(S_1) \cap \mathbf{cone}(S_2)) \neq \emptyset$. So, by the lemma,

$$\begin{aligned}
T(\bar{x}, S_1 \cap S_2) &= \overline{\mathbf{cone}(S_1 \cap S_2)} \\
&= \overline{\mathbf{cone}(S_1) \cap \mathbf{cone}(S_2)} \\
&= closedcone(S_1) \cap closedcone(S_2) \\
&= T(\bar{x}, S_1) \cap T(\bar{x}, S_2)
\end{aligned}$$

■

Recall that the sum of closed sets need not be closed.

■ **Example 7.13** $S_1 = \left\{x \in \mathbb{R}^n : x > 0, x_2 \geq \frac{1}{x_1}\right\}$, $S_2 = \left\{x \in \mathbb{R}^2 : x_2 = 0\right\}$, and $S_1 + S_2 = \left\{x \in \mathbb{R}^2 : x_2 > 0\right\}$. S_1 and S_2 are closed, but $S_1 + S_2$ is not. ■

Lemma 7.8 If $S_1, S_2 \subseteq \mathbb{R}^n$ are compact, then $S_1 + S_2$ is compact.

Lemma 7.9 Let $K \subseteq \mathbb{R}^n$ be a closed cone, let $a \in int(K^*)$, and let $H = \left\{x \in \mathbb{R}^n : a^\top x = 0\right\}$. Then, $H \cap K = \{0\}$.

Lemma 7.10 Let $K \subseteq \mathbb{R}^n$ be a closed cone, let $a \in int(K^*)$, and let $S = \left\{x \in \mathbb{R}^n : a^\top x \leq 1\right\}$. Then, $K \cap S$ is compact and $K = \mathbf{cone}(K \cap S)$.

Proof. Since $0 \in int(S)$ and K is a cone, $K = \mathbf{cone}(K \cap S)$. Since K and S are closed convex sets, $K \cap S$ is closed and convex. It remains to show that $K \cap S$ is bounded; suppose otherwise. Then, there exists $\mathbf{ray}(0, d) \subseteq K \cap S$. Since $a \in K^*$, we have $\mathbf{cone}(\{d\}) \subseteq \left\{x \in \mathbb{R}^n : 0 \leq a^\top x \leq 1\right\}$. Therefore, $a^\top d = 0$. However, that contradicts Lemma 7.9. ■

Lemma 7.11 If $K_1, K_2 \subseteq \mathbb{R}^n$ are closed cones and $int(K_1^* \cap K_2^*) \neq \emptyset$, then $K_1 + K_2$ is closed.

Proof. Let $a \in int(K_1^* \cap K_2^*)$ and let $s = \left\{x \in \mathbb{R}^n : a^\top x \leq 1\right\}$. By lemma 7.10, $S \cap K_1, S \cap K_2$ are compact. Note that

$$(K_1 + K_2) \cap S = ((K_1 \cap S) + (K_2 \cap S)) \cap S$$

Thus, $(K_1 + K_2) \cap S$ is compact. Now, $K_1 + K_2$ is a cone and $0 \in int(S)$, so $K_1 + K_2$ is the closure of $(K_1 + K_2) \cap S$ under non-negative scaling. Since $(K_1 + K_2) \cap S$ is closed, so is $K_1 + K_2$. ■

Theorem 55 Let $S = S_1 \cap \cdots \cap S_m$ where $S_1, \cdots, S_m \subseteq \mathbb{R}^n$ are closed convex sets and let $\bar{x} \in S$. If the interior of S is non-empty, then

$$N(S_1, \bar{x}) + \cdots + N(S_m, \bar{x})$$

is closed.

Proof. We shall do the case for $m = 2$ and let induction do the rest. By duality,

$$\begin{aligned}
int(N(\bar{x}, S_1)^* \cap N(\bar{x}, S_2)^*) &= int(T(\bar{x}, S_1) \cap T(\bar{x}, S_2)) \\
&= int((S_1 - \{\bar{x}\}) \cap (S_2 - \{\bar{x}\})) \\
&= int(S_1 \cap S_2) - \{\bar{x}\} \neq \{\emptyset\}
\end{aligned}$$

Thus, by Lemma 7.11, we have $N(\bar{x}, S_1) + N(\bar{x}, S_2)$ is closed. ■

7.12 Convex Functions

Recall that

Definition 7.12.1 — Convex Function. $f : S \to \mathbb{R}$ is convex if for each $x, y \in S$ and $\lambda \in [0,1]$, we have

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

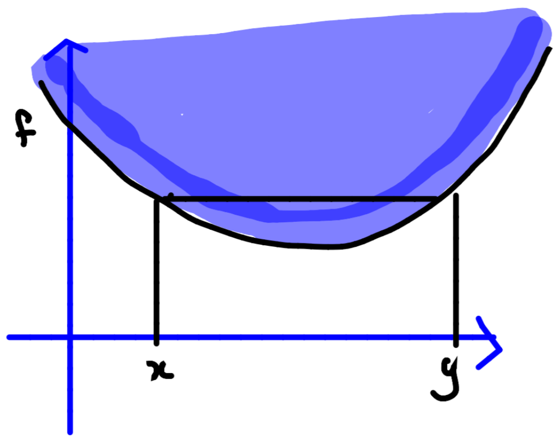


Figure 7.12.1: *Convex Function*

Definition 7.12.2 — Epigraphs of Convex Functions. The epigraph of a function $f : S_0 \to \mathbb{R}$, where $S_0 \subseteq \mathbb{R}^n$, is the subset of $\mathbb{R}^{n+1}$ defined by

$$\mathbf{epi}(f) := \left\{ \begin{bmatrix} x \\ z \end{bmatrix} \in \mathbb{R}^{n+1} : z \geq f(x), x \in S_0 \right\}$$

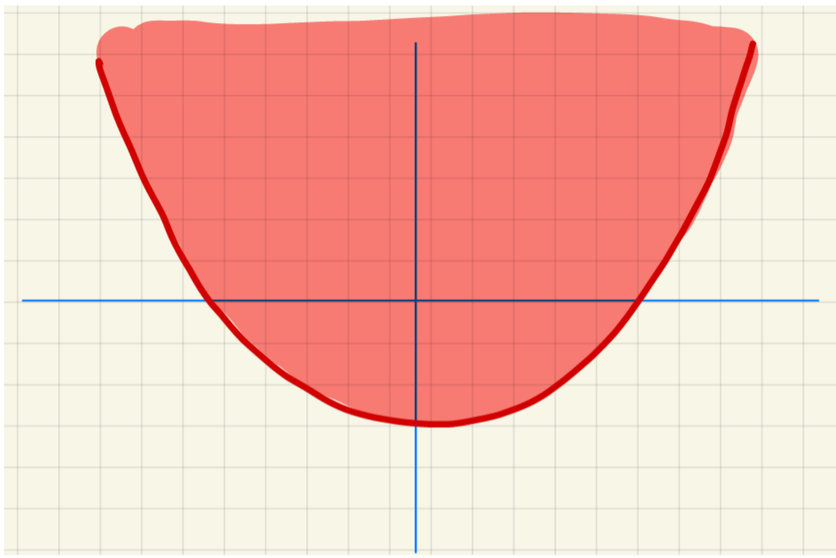


Figure 7.12.2: *Epigraph of f*

Lemma 7.13 $f : S_0 \to \mathbb{R}$ is convex if and only if $\mathbf{epi}(f)$ is convex.

Proof. 1. Suppose f is convex so S_0 is a convex set. Let $\begin{bmatrix} x_1 \\ z_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ z_2 \end{bmatrix} \in \mathbf{epi}(f)$ where $f(z_i) \geq$

$x_i \in S_0$ and let $\lambda \in [0,1]$. We check

$$\lambda \begin{bmatrix} x_1 \\ z_1 \end{bmatrix} + (1-\lambda) \begin{bmatrix} x_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} \lambda \lambda x_1 + (1-\lambda)x_2 \\ \lambda z_1 + (1-\lambda) z_2 \end{bmatrix} \in \mathbf{epi}(f)$$

where $\lambda x_1 + (1-\lambda)x_2 \in S_0$. And

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \leq \lambda z_1 + (1-\lambda) z_2$$

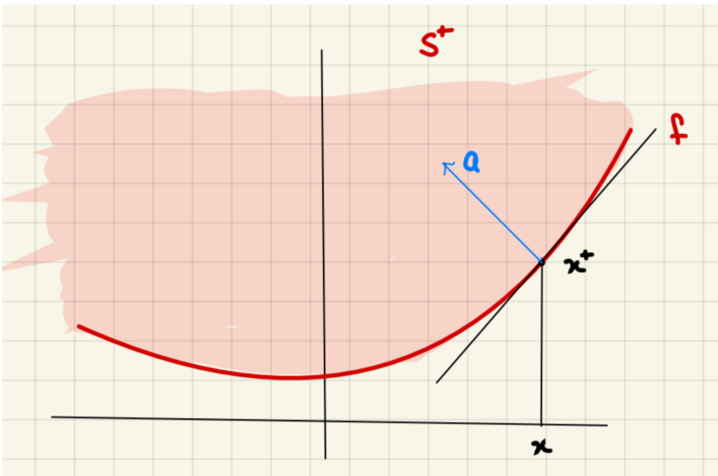
2. The converse is left as an exercise.

■

Lemma 7.14 Suppose that $S_0 \subseteq \mathbb{R}^n$ is an open set. A function $f : S_0 \to \mathbb{R}$ is convex if and only if it is convex on each line segment in S_0.

7.14.1 Functions in One Variable

Let $S_0 \subseteq \mathbb{R}$ be an open interval and let $f : S_0 \to \mathbb{R}$ be a convex function. Now let S^+ be the epigraph of f.



For a point $x \in S_0$, we let $x^+ = \begin{bmatrix} f(x) \\ x \end{bmatrix}$. Note that:

1. f is differentiable at x if and only if $N(S^+, x^+) = \mathbf{cone}(\{a\})$ for some $a \in \mathbb{R}^n$.
2. If f is differentiable, then $N(S^+, x^+) = \mathbf{cone}\left(\left\{\begin{bmatrix} 1 \\ -f'(x) \end{bmatrix}\right\}\right)$

Let $S_0 \subseteq \mathbb{R}$ be an open interval and let $f : S_0 \to \mathbb{R}$ be a differentiable function.

Lemma 7.15 If f is convex and $x, y \in S_0$, then $f(y) \geq f(x) + f'(x)(y-x)$.

Proof. Suppose that $y > x$. Consider $z \in (x,y)$. By convexity,

$$\frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}$$

Then,

$$f'(x) = \lim_{z \to x} \frac{f(z) - f(x)}{z - x} \leq \frac{f(y) - f(x)}{y - x}$$

Then,

$$f(y) \geq f(x) + (y-x) f'(x)$$

■

Theorem 56 Let $S_0 \subseteq \mathbb{R}$ be an open interval and let $f : S_0 \to \mathbb{R}$ be a differentiable function. TFAE:

1. f is convex
2. $f(y) \geq f(x) + f'(x)(y-x)$ for each $x, y \in S_0$
3. $(f'(y) - f'(x))(y-x) \geq 0$ for each $x, y \in S_0$
4. f' is non-decreasing on S_0

Theorem 57 Suppose that f is twice differentiable. Then, f is convex if and only if $f''(x) \geq 0, \forall x \in S_0$.

■ **Example 7.14** Let $S_0 = (0, \infty)$ and $f(x) = -\log(x)$. Now, $f'(x) = -\frac{1}{x}$ and $f''(x) = \frac{1}{x^2} > 0, \forall x \in S_0$. So, f is convex. ■

7.15.1 Functions in n Variables

Let $S_0 \subseteq \mathbb{R}^n$ be an open convex set and let $f : S_0 \to \mathbb{R}$ be a convex function. Let S^+ be the epigraph of f. For a point in $x \in S_0$, let $x^+ = \begin{bmatrix} f(x) \\ x \end{bmatrix}$. Note that:

1. f is differentiable at x if and only if $N(S^+, x^+) = \mathbf{cone}(\{a\})$ for some $a \in \mathbb{R}^n$
2. If f is differentiable, then $N(S^+, x^+) = \mathbf{cone}\left(\left\{\begin{bmatrix} 1 \\ -\nabla f(x) \end{bmatrix}\right\}\right)$

Note that $\nabla f(\bar{x}) = \left[\frac{\partial f(\bar{x})}{\partial x_1}, \cdots, \frac{\partial f(\bar{x})}{\partial x_n}\right]^\top$.

Proof. Consider the restriction of f to the line $\bar{x} + x_i e_i$ where $e_i = [0, \cdots, 1, \cdots, 0]^\top$. Recall that f is convex if and only if it is convex on each line segment in S_0. Consider distinct $\hat{x}, \hat{y} \in S_0$ and let $I = \{\alpha \in \mathbb{R} : (1-\alpha)\hat{x} + \alpha\hat{y} \in S_0\}$. Now define $g : I \to \mathbb{R}$ by $g(\alpha) = f((1-\alpha)\hat{x} + \alpha\hat{y})$. Therefore, $g(0) = f(\hat{x})$ and $g(1) = f(\hat{y})$. Note that $g'(0) = \nabla f(\hat{x}) \cdot (\hat{y} - \hat{x})$ ■

Theorem 58 TFAE:

1. f is convex
2. $f(y) \geq f(x) + \nabla f(x)^\top (y-x)$ for $x, y \in S_0$
3. $(\nabla f(y) - \nabla f(x))^\top (y-x) \geq 0$

R Recall that if f is C^2, the Hessian Hf is defined. Let $g : I \to \mathbb{R}$ be defined as above, then

$$g''(0) = (\hat{y} - \hat{x})^\top Hf(\hat{x})(\hat{y} - \hat{x})$$

Theorem 59 Let $S_0 \subseteq \mathbb{R}^n$ be an open convex set and let $f : S_0 \to \mathbb{R}$ be twice continuously differentiable. Then f is convex if and only if Hf is positive semidefinite for all $x \in S_0$.

■ **Example 7.15** Consider a quadratic function $f(x) = x^\top Ax + b^\top x$ where A is a symmetric $n \times n$ matrix. Then, $\nabla f(x) = 2Ax + b$ and $Hf(x) = 2A$. So f is convex if and only if A is positive semidefinite. ■

7.16 Level Sets of Convex Functions

Definition 7.16.1 — Level Sets. If $f : S_0 \to \mathbb{R}$ is a convex function and $\alpha \in \mathbb{R}$, then the set $S = \{x \in S_0 : f(x) \leq \alpha\}$ is convex. We call this S a level set of S.

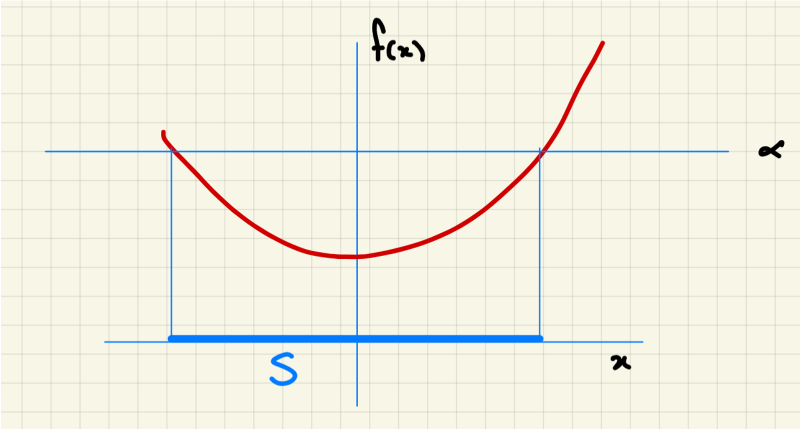


Figure 7.16.1: *Level Set*

R Level sets of non-convex functions can be convex. For example, $f(x) = \sqrt{|x|}$.

7.16.1 Normal Cones and Level Sets

Let $S_0 \subseteq \mathbb{R}^n$ be an open convex set, let $f : S_0 \to \mathbb{R}$ be a differentiable convex function, let $\alpha \in \mathbb{R}$, and $S = \{x \in S_0 : f(x) \leq \alpha\}$. Note that, if $\hat{x} \in S$ and $f(\hat{x}) < \alpha$, then $\hat{x}$ is in the interior of S and, hence, $N(\hat{x}, S) = \{0\}$.

Lemma 7.17 If $\hat{x} \in S$ with $f(\hat{x}) = \alpha$, then $-\nabla f(\hat{x}) \in N(\hat{x}, S)$.

Proof. Let $c = -\nabla f(\hat{x})$. Recall for each $x \in S$,

$$f(x) \geq f(\hat{x}) + \nabla f(\hat{x})^\top (x - \hat{x})$$

Thus,

$$\begin{aligned} c^\top x - c^\top \hat{x} &= -\nabla f(\hat{x})(x - \hat{x}) \\ &\geq f(\hat{x}) - f(x) \\ &= \alpha - f(x) \\ &\geq 0 \end{aligned}$$

hence, $c \in N(\hat{x}, S)$. ■

Consider an optimization problem:

$$(P) \qquad \min \left(c^\top x : f(x) \leq \alpha, x \in S_0 \right)$$

where $S_0 \subseteq \mathbb{R}^n$ is an open set, $f : S_0 \to \mathbb{R}$ is convex and differentiable and $\alpha \in \mathbb{R}$.

Lemma 7.18 For $\bar{x} \in S_0$ with $f(\bar{x}) = \alpha$, if $c = -\lambda \nabla f(\bar{x})$ for some $\lambda \geq 0$, then $\bar{x}$ is optimal for (P).

Proof. Let $S = \{x \in S_0 : f(x) \leq \alpha\}$. Since $-\nabla f(\bar{x}) \in N(\bar{x}, S)$ and $\lambda \geq 0$, we have $c \in N(\bar{x}, S)$ and hence $\bar{x}$ is optimal ■

■ **Example 7.16 — Arithmetic Mean-Geometric Mean Inequality.** For real numbers $x_1, \cdots, x_n > 0$,

$$\frac{x_1 + \cdots + x_n}{n} \geq \sqrt[n]{x_1 \cdots x_n}$$

Proof. By recaling, we can assume that $\sqrt[n]{x_1 \cdots x_n} = 1$. Then, consider the problem

$$(P) \begin{cases} \min & \frac{1}{n}x_1 + \cdots + \frac{1}{n}x_n \\ s.t. & \sqrt[n]{x_1 \cdots x_n} \geq 1 \\ & x_1, x_2, \cdots, x_n > 0 \end{cases}$$

It suffices to prove that $\mathbf{OPT}(P) \geq 1$, or equivalently, that $\bar{x}[1, \cdots, 1]^\top$ is optimal. We note that TFAE:

1. $\sqrt[n]{x_1 \cdots x_n} \geq 1$
2. $x_1, \cdots, x_n \geq 1$
3. $-\ln(x_1) - \cdots - \ln(x_n) \leq 0$

We know that $-\ln(x)$ is convex on $\{x \in \mathbb{R} : x > 0\}$, is $f(x) = -\ln(x_1) - \cdots - \ln(x_n)$ is convex on $S_0 = \{x \in \mathbb{R}^n : x > 0\}$. We rewrite (P) as

$$(P') \qquad \min(c^\top x : f(x) \leq 0, x \in S_0)$$

where $c = \left[\frac{1}{n}, \cdots, \frac{1}{n}\right]^\top$, $S_0 = \{x \in \mathbb{R}^n : x > 0\}$, and for $x \in S_0$. The feasible region of (P') is the level set of a convex function. Note that $\bar{=}[1, \cdots, 1]^\top$ is feasible for (P') and $f(\bar{x}) = 0$. Now,

$$-\nabla f(\bar{x}) = \left[\frac{1}{\bar{x}_1}, \cdots, \frac{1}{\bar{x}_n}\right]^\top = [1, \cdots, 1]^\top = nc^\top$$

Thus, $\bar{x}$ is optimal for (P') as required. ■

■

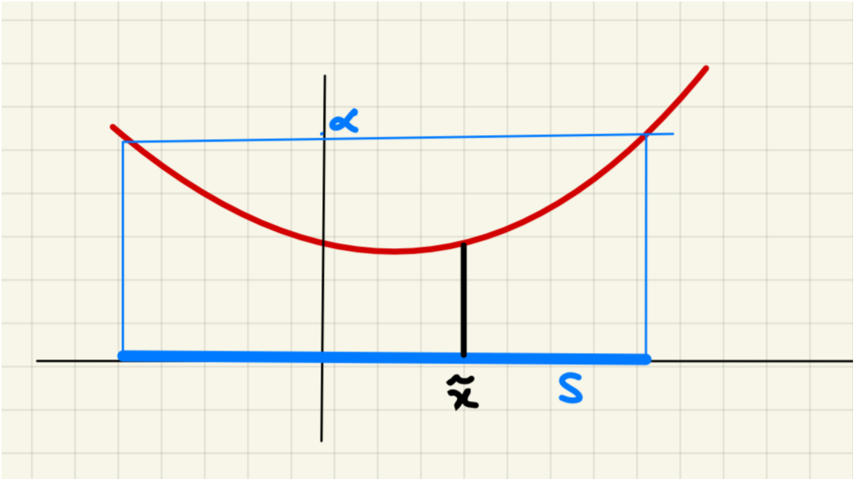
We have seen that if $\bar{x} \in S$ with $f(\bar{x}) = \alpha$, then $-\nabla f(\bar{x}) \in N(\bar{x}, S)$. But is $N(\bar{x}, S) = \mathbf{cone}(\{-\nabla f(\bar{x})\})$? **No!**

■ **Example 7.17 — Counterexample.** For $x \in \mathbb{R}$, let

$$f(x) = \begin{cases} x^2 & x \geq 0 \\ 0 & x \leq 0 \end{cases}$$

let $S = \{x \in \mathbb{R} : f(x) \leq 0\} = \{x \in \mathbb{R} : x \leq 0\}$. Note that f is convex and differentiable and $f(0) = 0$. Now, $\nabla f(0) = 0$. But $N(0, S) = \{\lambda \in \mathbb{R} : \lambda \leq 0\} \neq \mathbf{cone}(\{-\nabla f(0)\})$ ■

Definition 7.18.1 — Slater Point. Let $S_0 \subseteq \mathbb{R}^n$ be an open convex set, let $f : S_0 \to \mathbb{R}$ be a differentiable convex function, let $\alpha \in \mathbb{R}$, and $S = \{x \in S_0 : f(x) \leq \alpha\}$. A point $x_0 \in S$ with $f(x_0) < \alpha$ is called a slater point.



Theorem 60 Let $\bar{x} \in S$ with $f(\bar{x}) = \alpha$. If there exists a slater point, then $N(\bar{x}, S) = \mathbf{cone}(\{-\nabla f(\bar{x})\})$