Combinatorial Optimization: Week 1

Joonhyung Shin

May 5th, 2020

Introduction

- Welcome to the Combinatorial Optimization Study Group!
- Group members (6 people)
  - joon (me)
  - 16silver
  - gray
  - hyeongmin.byun
  - 만레라부싫사일레라부즐사단
  - TAMREF
- March 5th - July 21st (tentative)
- Textbook: *A First Course in Combinatorial Optimization* by Jon Lee
- Goal: Chapter 0 - Chapter 5 $(+\alpha)$

Introduction

- Schedule (tentative)
  - Week 1 (5/5) - Introduction, Chapter 0 (Host: joon)
  - Week 2 (5/12) - Chapter 0, 1 (Host: TAMREF)
  - Week 3 (5/19) - Chapter 1 (Host: hyeongmin.byun)
  - Week 4 (5/26) - Chapter 3 (Host: 만레라부싫사일레라부즐사단)
  - Week 5 (6/2) - Chapter 2 (Host: 16silver)
  - Week 6 (6/9) - Chapter 4 (Host: gray)
  - Week 7 (6/16) - Chapter 4 (Host: 만레라부싫사일레라부즐사단)
  - Week 8 (6/23) - Chapter 5 (Host: hyeongmin.byun)
  - Week 9 (6/30) - TBD (Host: 16silver)
  - Week 10 (7/7) - TBD (Host: TAMREF)
  - Week 11 (7/14) - TBD (Host: gray)
  - Week 12 (7/21) - TBD (Host: joon)
- Feedback is always welcome! (Regarding schedule, topics, ...)

Introduction

- Q. Is it okay to cover more or less topics?
- A. Of course!
- Q. Is it okay to exchange my seminar schedule with another member?
- A. Of course!
- Q. Is it okay to quote some examples or demonstrations from some other source?
- A. Of course!
- Q. Is it okay to run away?
- A. NO.

Introduction

What is *combinatorial optimization*?

Introduction

What is *optimization*?

Introduction

What is *optimization*?
Maximizing $c : \mathcal{S} \to \mathbb{R}$!

Introduction

Example (Quadratic function)

If $f(x) = -x^2 + 2x + 5$, then f attains its maximum 6 at $x = 1$.

Example

Let D be a closed unit disk. If $f : D \to \mathbb{R}$ is defined by $f(x, y) = 3x + 4y$, then f attains its maximum 5 at $(3/5, 4/5)$.

Introduction

Example (Maximum matching)

Let $G = (V, E)$ be a graph, and let $\mathcal{M}_G$ be the set of matchings of G. If $c : \mathcal{M}_G \to \mathbb{Z}_{\geq 0}$ is the counting function, what is the maximum of f?

Example (Machine learning)

Let $X \subseteq \mathbb{R}^n$ be the set of choices of n parameters of a machine learning model. If $L : X \to \mathbb{R}$ is a loss function, what choice of parameters minimizes the loss?

Introduction

Solving an *optimization problem* is maximizing $c : \mathcal{S} \to \mathbb{R}$.

Definition

An element of $\mathcal{S}$ is called a **feasible solution**.

Definition (Discrete optimization)

If $\mathcal{S}$ is finite, then the optimization problem is called a **discrete optimization problem**.

Definition (Combinatorial optimization)

If $\mathcal{S} \subseteq 2^E$ for some finite set E, then the optimization problem is called a **combinatorial optimization problem**.

Example (Maximum matching)

In the maximum matching problem, $\mathcal{S} = \mathcal{M}_G$ is a subset of $2^{E(G)}$.

Example (Minimum spanning tree)

In the minimum spanning tree problem, $\mathcal{S} \subseteq 2^{E(G)}$ is the set of spanning trees, and $c : \mathcal{S} \to \mathbb{R}$ is the weight function.

Example (Maximum clique)

In the maximum clique problem, $\mathcal{S} \subseteq 2^{V(G)}$ is the set of cliques, and $c : \mathcal{S} \to \mathbb{Z}_{\geq 0}$ is the counting function.

Introduction

- Sometimes, objective function $c : \mathcal{S} \subseteq 2^E \to \mathbb{R}$ can be expressed as

$$c(X) = \sum_{x \in X} c(\{x\})$$

for each $X \in \mathcal{S}$.

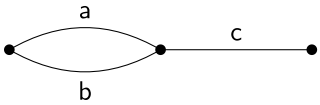
Example (Minimum spanning tree)

In the minimum spanning tree problem, $c(\{e\})$ is the weight of the edge $e \in E(G)$.

- In such case, the optimization problem can be partially handled with **linear programming** (LP).

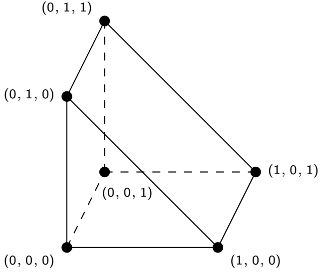
Introduction

- Toy example: *maximum* spanning tree problem



- The weights are: $w(a) = 5$, $w(b) = 4$, $w(c) = 1$.
- Express each spanning tree as an element of $2^{\{a,b,c\}} = \{0,1\}^3 \subseteq \mathbb{R}^3$. (e.g. $\{a,c\} \to (1,0,1)$, $\{b\} \to (0,1,0)$)
- The objective function is $c(x_1, x_2, x_3) = 5x_1 + 4x_2 + x_3$.
- Embed these points in $\mathbb{R}^3$ and depict the **convex hull** $\mathcal{P}$.

Introduction

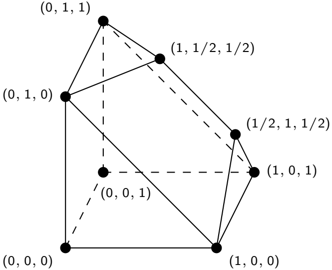


- A $0/1$ point $x \in \mathcal{P}$ if and only if it represents a spanning tree.
- The polytope $\mathcal{P}$ can be expressed in terms of the linear inequalities

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad 0 \leq x_3 \leq 1, \qquad x_1 + x_2 \leq 1.$$

- We are maximizing a linear function $5x_1 + 4x_2 + x_3$... a LP!
- The solution to this LP is $(1, 0, 1)$, representing a spanning tree $\{a, c\}$.

Introduction



- This polytope can be expressed in terms of the linear inequalities

$$
\begin{aligned}
0 \le x_1 \le 1, \qquad & 0 \le x_2 \le 1, \qquad & 0 \le x_3 \le 1,\\
x_1 + x_2 - x_3 \le 1, \qquad & x_1 + x_2 + x_3 \le 2. &
\end{aligned}
$$

- A 0/1 point x is in this polytope if and only if it represents a spanning tree.
- However, the solution to this LP is $(1, 1/2, 1/2)$, which is invalid.

Introduction

- Lessons learned: continuous modeling (LP in the example) should be done very carefully when variables represent discrete choices.
- Even though continuous modeling is not a panacea, it can be useful solving a combinatorial optimization problem.
- Chapter 0 is about linear programming.

- From linear algebra, you might be familiar with a system of linear equalities:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i, \quad \text{for } i = 1, 2, \cdots, m$$

  or equivalently, $A\mathbf{x} = \mathbf{b}$.
- In linear programming, we are interested in a system of linear **inequalities**:

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i, \quad \text{for } i = 1, 2, \cdots, m$$

  or equivalently, $A\mathbf{x} \leq \mathbf{b}$. Here, inequality is elementwise.
- This can be viewed as a more general version, as an equality is basically two inequalities.

- The system is complicated since it has too many variables; we want to eliminate some variables!
- For linear equalities, we do Gauss-Jordan Elimination.
- For linear inequalities, we do **Fourier-Motzkin Elimination**.

Proposition (Fourier-Motzkin Elimination)

Let $A\mathbf{x} \leq \mathbf{b}$ be a system of linear inequalities, and fix $1 \leq k \leq n$. Define $S_+ := \{i : a_{ik} > 0\}$, $S_- := \{i : a_{ik} > 0\}$, and $S_0 := \{i : a_{ik} = 0\}$. Then the new system

$$\sum_{j=1}^{n} a_{ij}x_j \leq b_i \text{ for } i \in S_0,$$

$$-a_{\ell k}\left(\sum_{j=1}^{n} a_{ij}x_j \leq b_i\right) + a_i k\left(\sum_{j=1}^{n} a_{\ell j}x_j \leq b_\ell\right) \text{ for } i \in S_+, \ell \in S_-$$

satisfies the following:

- *the new system does not involve x_k;*
- *$(x_1, \cdots x_{k-1}, x_{k+1}, \cdots x_n)$ solves the new system if and only if there exists x_k such that $(x_1, \cdots, x_{k-1}, x_k, x_{k+1}, \cdots x_n)$ solves the original system.*

Example

Consider the following system:

$$5x_1 - 2x_2 + x_3 \leq 4,$$
$$-x_1 + 2x_3 \leq -1,$$
$$3x_1 + x_2 \leq 6.$$

We want to eliminate x_2. Then we have $S_+ = \{3\}$, $S_- = \{1\}$, $S_0 = \{2\}$. Applying the Fourier-Motzkin elimination, we construct a new system

$$-x_1 + 2x_3 \leq -1,$$
$$1 \cdot (5x_1 - 2x_2 + x_3 \leq 4) + 2 \cdot (3x_1 + x_2 \leq 6),$$

or equivalently,

$$-x_1 + 2x_3 \leq -1, \quad 11x_1 + x_3 \leq 16.$$

Remark

- Each inequality of the new system is a linear combination of inequalities of the original system.
- If $S_0 \cup S_+$ or $S_0 \cup S_-$ is empty, the new system has no inequalities.

Theorem (Theorem of the Alternative for Linear Inequalities)

The system $A\mathbf{x} \leq \mathbf{b}$ has a solution if and only if the system

$$\begin{aligned} A^T\mathbf{y} &= \mathbf{0}, \\ \mathbf{y} &\geq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} &< 0 \end{aligned}$$

has no solution.

- This theorem makes "unsolvability" condition as useful as "solvability" condition.
- Proof idea: what happens if we eliminate all variables $\mathbf{x} = (x_1, \cdots, x_n)$?

Theorem (Farkas Lemma)

The system

$$A\mathbf{x} = \mathbf{b}, \\ \mathbf{x} \geq \mathbf{0}$$

has a solution if and only if the system

$$A^T\mathbf{y} \geq \mathbf{0}, \\ \mathbf{b}^T\mathbf{y} < 0$$

has no solution.

- Intuitively, this means that **b** is inside the cone of the column vectors of A if and only if no hyperplane (represented by a normal vector **y**) separates **b** from the cone.

Definition

The following optimization problem is called a **linear programming** (LP):

$$\max \sum_{j=1}^{n} c_j x_j$$

subject to:

$$\sum_{j=1}^{n} a_{ij} x_j \leq b_i \quad \text{for } i = 1, \cdots, m,$$

or equivalently,

$$\max \mathbf{c}^T \mathbf{x}$$

subject to:

$$A\mathbf{x} \leq \mathbf{b}.$$

Example

A farmer has a land of size L, F units of fertilizers, and P units of pesticides. Every unit land of wheat requires F_1 units of fertilizers and P_1 units of pesticides, whereas every unit land of barley requires F_2 units of fertilizers and P_2 units of pesticides. Each unit land of wheat is worth S_1 dollars, and each unit land of barley is worth S_2 dollars. What is the maximum profit?
This problem can be expressed with the following LP:

$$\max(S_1x_1 + S_2x_2)$$

subject to:

$$\begin{aligned} x_1 + x_2 &\leq L \\ F_1x_1 + F_2x_2 &\leq F \\ P_1x_1 + P_2x_2 &\leq P \\ x_1 \geq 0,\ x_2 &\geq 0 \end{aligned}$$

- A linear programming is an optimization problem where the objective and the constraints are all linear.
- Standard form (according to the author...) of LP:

$$\begin{aligned} \max \mathbf{c}^T\mathbf{x} \\ \text{subject to:} \\ A\mathbf{x} &= \mathbf{b}, \\ \mathbf{x} &\geq 0. \end{aligned}$$

- Yet another standard form of LP:

$$\begin{aligned} \max \mathbf{c}^T\mathbf{x} \\ \text{subject to:} \\ A\mathbf{x} &\leq \mathbf{b}, \\ \mathbf{x} &\geq 0. \end{aligned}$$

Definition (Asymmetric dual)

The **dual program** of the **primal program**

(P)
$$\begin{aligned} &\max \mathbf{c}^T\mathbf{x} \\ &\text{subject to:} \\ &A\mathbf{x} \le \mathbf{b} \end{aligned}$$

is the linear program

(D)
$$\begin{aligned} &\min \mathbf{b}^T\mathbf{y} \\ &\text{subject to:} \\ &A^T\mathbf{y} = \mathbf{c}, \\ &\mathbf{y} \ge \mathbf{0}. \end{aligned}$$

- The contraints of the primal program becomes the objective in the dual program, and vice versa.

Definition (Symmetric dual)

The dual program of the primal program

$$\text{(P)} \qquad \begin{aligned} \max \mathbf{c}^T\mathbf{x} \\ \text{subject to:} \\ A\mathbf{x} &\leq \mathbf{b} \\ \mathbf{x} &\geq 0 \end{aligned}$$

is the linear program

$$\text{(D)} \qquad \begin{aligned} \min \mathbf{b}^T\mathbf{y} \\ \text{subject to:} \\ A^T\mathbf{y} &\geq \mathbf{c}, \\ \mathbf{y} &\geq \mathbf{0}. \end{aligned}$$

Definition

The dual program of the primal program

$$\text{(P)} \qquad \begin{aligned} \max \mathbf{c}^T\mathbf{x} \\ \text{subject to:} \\ A\mathbf{x} = \mathbf{b} \\ \mathbf{x} \geq 0 \end{aligned}$$

is the linear program

$$\text{(D)} \qquad \begin{aligned} \min \mathbf{b}^T\mathbf{y} \\ \text{subject to:} \\ A^T\mathbf{y} \geq \mathbf{c}. \end{aligned}$$

Theorem (Weak Duality Theorem)

If $\mathbf{x}$ *is feasible to* P *and* $\mathbf{y}$ *is feasible to* D*, then* $\mathbf{c}^T\mathbf{x} \leq \mathbf{b}^T\mathbf{y}$*. Hence, (1) if* $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$*, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are optimal, and (2) if either program is unbounded, then the other is infeasible.*

Theorem (Strong Duality Theorem)

If P *and* D *have feasible solutions, then they have optimal solutions* $\mathbf{x}$, $\mathbf{y}$ *with* $\mathbf{c}^T\mathbf{x} = \mathbf{b}^T\mathbf{y}$*. If either is infeasible, then the other is either infeasible or unbounded.*

Example (Farmer revisited)

A farmer has a land of size L, F units of fertilizers, and P units of pesticides. Every unit land of wheat requires F_1 units of fertilizers and P_1 units of pesticides, whereas every unit land of barley requires F_2 units of fertilizers and P_2 units of pesticides. This time, we want to assign an *economic value* to each of the resource types, so that a unit land of wheat is worth at least S_1 dollars, and a unit land of barley is worth at least S_2 dollars. What is the minimum value of the farmer's assets?

This is the dual problem of the previous example, and it can be expressed with:

$$\begin{aligned} &\min(Ly_L + Fy_F + Py_P) \\ &\quad \text{subject to:} \\ &\qquad y_L + F_1 y_F + P_1 y_P \geq S_1, \\ &\qquad y_L + F_2 y_F + P_2 y_P \geq S_2, \\ &\qquad y_L \geq 0, y_F \geq 0, y_P \geq 0. \end{aligned}$$

- In the farmer example, the primal problem is a *resource allocation* problem, and the dual problem is a *resource valuation* problem.
- By the Strong Duality Theorem, the minimum value of the farmer's assets is the maximum profit the farmer can gain.
- Some other famous examples of the LP duality are:
  - Max-flow min-cut theorem (maximum flow is equal to the minimum cut)
  - Menger's theorem (graph connectivity)
  - Von Neumann's minimax theorem (zero-sum games)
  - König's theorem (maximum matching is minimum vertex cover)
  - ... and much more!

Definition

Points $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{y} \in \mathbb{R}^m$ are **complementary** with respect to P and D, if

$$y_i \left(b_i - \sum_{j=1}^{n} a_{ij} x_j \right) = 0 \quad \text{for } i = 1, \cdots, m.$$

Theorem (Weak Complementary-Slackness Theorem)

If feasible solutions $\mathbf{x}$ *and* $\mathbf{y}$ *are complementary, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are optimal solutions.*

Theorem (Strong Complementary-Slackness Theorem)

If $\mathbf{x}$ *and* $\mathbf{y}$ *are optimal solutions to* P *and* D*, respectively, then* $\mathbf{x}$ *and* $\mathbf{y}$ *are complementary.*

- The Complementary-Slackness Theorem gives a necessary and sufficient condition for $\mathbf{x}$ and $\mathbf{y}$ being optimal solutions.
- In the farmer example, the theorem can be interpreted as follows:
  - If the farmer (who is trying to maximize his profit) did not fully exploit a certain resource, the resource can have no value.
  - If a resource should have a positive value, the farmer must fully exploit the resource in order to maximize his profit.

- Consider a (standard) linear program

$$(\mathrm{P}') \qquad \begin{aligned} \max\ & \mathbf{c}^T\mathbf{x} \\ \text{subject to:}\ & \\ A\mathbf{x} &= \mathbf{b} \\ \mathbf{x} &\geq 0 \end{aligned}$$

  where the row vectors of A are assumed to be linear independent. (Otherwise, we can always reduce it!) In other words, $m \leq n$ and $\mathrm{rank}(A) = m$ where $A \in \mathbb{R}^{m \times n}$.
- We can choose column index set $\{\beta_1, \cdots, \beta_m\} \subseteq \{1, \cdots, n\}$ so that the column vectors $A_{\beta_1}, \cdots, A_{\beta_m}$ are linearly independent. Let $\{\eta_1, \cdots, \eta_{n-m}\}$ be the remaining index set.
- Let $\mathbf{x}^* \in \mathbb{R}^n$ be a point satisfying $\mathbf{x}_\beta = A_\beta^{-1}\mathbf{b}$ and $\mathbf{x}_\eta = \mathbf{0}$.

Definition

The index set β is called a **basis**, and η is called a **nonbasis**. The point $\mathbf{x}^*$ is called a **basic solution** associated with β and η. If $\mathbf{x}^* \geq 0$, then $\mathbf{x}$ is feasible, and when this is the case, $\mathbf{x}^*$ is said to be **primal feasible**. Furthermore, if $\mathbf{x}^*$ is optimal, it is said to be **primal optimal**.

- Now consider the dual program of P':

$$(D') \qquad \begin{aligned} &\min \mathbf{b}^T\mathbf{y} \\ &\text{subject to:} \\ &A^T\mathbf{y} \geq \mathbf{c}. \end{aligned}$$

- Associated with β we have a solution $\mathbf{y}^* = (A_\beta^{-1})^T\mathbf{c}_\beta$.

Definition

If $\bar{\mathbf{c}}_\eta := \mathbf{c}_\eta - A_\eta^T\mathbf{y}^* \leq \mathbf{0}$, then $\mathbf{y}^*$ is a feasible solution, and in such case we say that β is **dual feasible**. If $\mathbf{y}^*$ is optimal, then β is said to be **dual optimal**. We say that β is **optimal** if it is both primal optimal and dual optimal.

Theorem (Weak Optimal-Basis Theorem)

If the basis β is both primal feasible and dual feasible, then β is optimal.

Proof.

If $\mathbf{x}^*$ and $\mathbf{y}^*$ are the basic solutions associated with β, we have

$$\mathbf{c}^T\mathbf{x}^* = \mathbf{c}_\beta^T\mathbf{x}_\beta^* = \mathbf{c}_\beta^T A_\beta^{-1}\mathbf{b} = \mathbf{b}^T\mathbf{y}^*.$$

□

Theorem (Strong Optimal-Basis Theorem)

If P' and D' are feasible, then there is a basis β that is both primal feasible and dual feasible, hence optimal.

Example (Farmer revisited)

We slightly modify the original LP by adding auxiliary variables x_L, x_F, and x_P; they represent the amount *residue* for each resource. Then the primal program becomes

$$\begin{aligned}
&\max(S_1x_1 + S_2x_2)\\
&\quad\text{subject to:}\\
&\qquad \begin{array}{rcccl}
x_1 + x_2 & +x_L & & & = L,\\
F_1x_1 + F_2x_2 & & +x_F & & = F,\\
P_1x_1 + P_2x_2 & & & +x_P & = P\\
x_1 \geq 0,\ x_2 \geq 0, & \multicolumn{2}{c}{x_L \geq 0,\ x_F \geq 0,} & x_P \geq 0. &
\end{array}
\end{aligned}$$

The dual program remains unchanged.

Example (Farmer revisited)

The matrix form of the equality part of the primal program is

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ F_1 & F_2 & 0 & 1 & 0 \\ P_1 & P_2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_L \\ x_F \\ x_P \end{bmatrix} = \begin{bmatrix} L \\ F \\ P \end{bmatrix}$$

and the matrix form of the main part of the dual program is

$$\begin{bmatrix} 1 & F_1 & P_1 \\ 1 & F_2 & P_2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_L \\ y_F \\ y_P \end{bmatrix} \geq \begin{bmatrix} S_1 \\ S_2 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

- The Strong Optimal-Basis Theorem can be proved by a constructive algorithm; a so-called *Primal Simplex Method*.
    1. Start with any primal feasible basis. (Such basis can be found with a variant of the Primal Simplex Method itself)
    2. If the basis violates dual feasibility condition at some η_j, exchange it with some *good* β_i. It is guaranteed that the primal feasibility is maintained and the dual infeasibility gap always decreases after this step.
    3. Repeat until the basis is dual feasible.
- Please read the text for the full description of the algorithm!

- Though the detail of the Primal Simplex Method is quite complicated, it has an interesting gemoetric interpretation.

Definition

An **extreme point** of a convex set C is a point $x \in C$ such that $x_1, x_2 \in C$, $0 < \lambda < 1$, and $x = \lambda x_1 + (1-\lambda)x_2$ implies $x = x_1 = x_2$.

- An extreme point can be viewed as a vertex of a polytope.

Theorem

The set of feasible basic solutions of P' is the same as the set of extreme points of $\mathcal{P}'$.

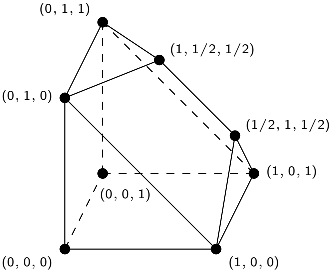
- Therefore, the process of finding an optimal solution can be described as moving from vertex to vertex in $\mathcal{P}$! This is why it is called the Primal *Simplex* Method.

Example (Farmer revisited)

The matrix form of the equality part of the primal program is

$$\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ F_1 & F_2 & 0 & 1 & 0 \\ P_1 & P_2 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_L \\ x_F \\ x_P \end{bmatrix} = \begin{bmatrix} L \\ F \\ P \end{bmatrix}.$$

Try to visualize the feasible basic solutions in $\mathbb{R}^2$. Think of an imaginary pointer moving from vertex to vertex.



Definition

A **polytope** is the convex hull of a finite set $X \subseteq \mathbb{R}^n$.

Theorem (Weyl)

If $\mathcal{P}$ is a polytope, then $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$ for some $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.

Proof.

Since $\mathcal{P} = \mathrm{conv}(X)$ for some $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_k\} \subseteq \mathbb{R}^n$ by definition, $\mathcal{P}$ is precisely the solution set of the linear inequality system

$$x_i - \sum_{1 \leq j \leq k} \lambda_j x_{ij} = 0 \quad \text{for } i = 1, \cdots, n,$$

$$\sum_{1 \leq j \leq k} \lambda_j = 1, \quad \lambda_j \geq 0 \quad \text{for } j = 1, \cdots, k.$$

where $\mathbf{x} = (x_1, \cdots, x_n)$ and λ_j are variables. Now apply Fourier-Motzkin elimination to eliminate the variables λ_j. $\square$

Theorem (Minkowski)

If $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \le \mathbf{b}\}$ and $\mathcal{P}$ is bounded, then $\mathcal{P}$ is a polytope.

Proof.

(Sketch) Let X be the set of extreme points of $\mathcal{P}$. (Recall the definition of extreme points.) We prove that $\mathcal{P} = \text{conv}(X)$. Assume otherwise, so that we have $\tilde{\mathbf{x}} \in \mathcal{P} \setminus \text{conv}(X)$. Use the Theorem of the Alternative for Linear Inequalities, and use the fact that if a linear program has an optimal solution, at least one of the extreme points is feasible optimal. $\square$

Definition

The **dimension** of a polytope $\mathcal{P}$, denoted by $\dim(\mathcal{P})$, is the maximum number of affinely independent points in $\mathcal{P}$ minus one. The polytope $\mathcal{P} \subset \mathbb{R}^n$ is **full dimensional** if $\dim(\mathcal{P}) = n$.

Definition

An inequality $\alpha^T \mathbf{x} \leq \beta$ is **valid** for the polytope $\mathcal{P}$ if every point in $\mathcal{P}$ satisfies the inequality. Given a valid inequality, a **face** of $\mathcal{P}$ is a set

$$\mathcal{F} := \mathcal{P} \cap \left\{ \mathbf{x} \in \mathbb{R}^n : \alpha^T \mathbf{x} = \beta \right\}.$$

- If the face is nonempty, it is the set of points that maximizes $\alpha^T \mathbf{x}$.
- A nonempty face can be viewed as the set of optimal solutions of a linear program with constraints described by the polytope.

Proposition

Let $\mathcal{P} = \text{conv}(X)$ be a polytope.

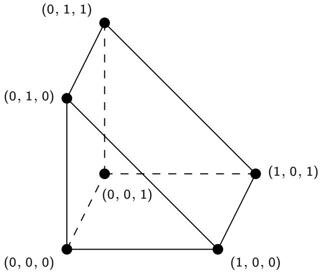
- *Faces of $\mathcal{P}$ are polytopes.*
- *The empty set ϕ and $\mathcal{P}$ itself are trivial faces of $\mathcal{P}$.*
- *The set of extreme points of $\mathcal{P}$ is equal to the set of faces of dimension 0.*

Proof.

(Sketch) For a valid inequality $\alpha^T \mathbf{x} \leq \beta$ describing a face $\mathcal{F}$, let $Y \subseteq X$ be the set of points satisfying the equality. Then $\mathcal{F} = \text{conv}(Y)$, hence $\mathcal{F}$ is a polytope. The third fact comes from Minkowski's theorem. □

Definition

Faces of dimension $\dim(\mathcal{P}) - 1$ are called **facets**.



Example

In the above polytope, each vertex is an extreme point, and each edge is a 1-dimensional face. The triangle with vertices $(0,0,0),(1,0,0),(0,1,0)$ is a 2-dimensional face, hence a facet. The dimension of the polytope is 3, thus is full dimensional.

Theorem

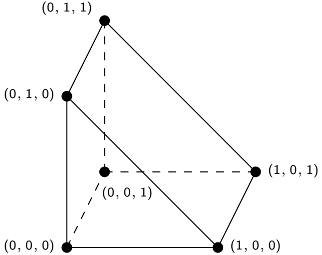
Let $\mathcal{P} = \{\mathbf{x} \in \mathbb{R}^n : A\mathbf{x} \leq \mathbf{b}\}$ be a polytope. If a (valid) inequality $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$ for some i describes a face with dimension less than $\dim(\mathcal{P}) - 1$, it is redundant. Conversely, for each facet $\mathcal{F}$ of $\mathcal{P}$, there exists i such that $\sum_{j=1}^{n} a_{ij}x_j \leq b_i$ describes $\mathcal{F}$.

- The forward direction of the theorem is called *Redundancy Theorem* in the book.
- This theorem implies that the inequalities in a "minimal" description of a polytope should describe distinct facets of the polytope.
- Now we are concerned about some kind of uniqueness of such description!

Theorem (Unique Description Theorem)

If $\mathcal{P}$ is a full-dimensional polytope, then each valid inequality that describes a facet of $\mathcal{P}$ is unique, up to multiplication by a positive scalar. Conversely, if a face is described by a unique inequality up to multiplication by positive scalar, it is a facet of $\mathcal{P}$.

- Therefore, it is more convenient to work with full-dimensional polytopes.

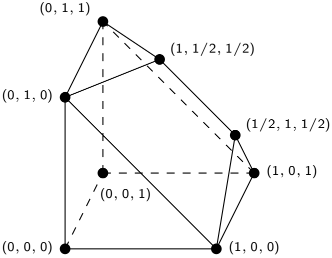


Example

The above polytope can be described by the five linear inequalities

$$x_1 \geq 0, \qquad x_2 \geq 0, \qquad 0 \leq x_3 \leq 1, \qquad x_1 + x_2 \leq 1.$$

These inequalities describe each facet, and by the Unique Description Theorem, this description is unique up to multiplcation by positive scalar.



Example

The above polytope can be described by the eight linear inequalities

$$0 \le x_1 \le 1, \qquad 0 \le x_2 \le 1, \qquad 0 \le x_3 \le 1,$$

$$x_1 + x_2 - x_3 \le 1, \qquad x_1 + x_2 + x_3 \le 2.$$

Homeworks

- P6 Exercise (Comparing relaxations)
- P14 Problem (Farkas Lemma)
- P17 Problem (Theorem of the Alternative for Linear Inequalities)