CO 255 with Ricardo Fukasawa

Eason Li

2024 F

Contents

1 Optimization Problem

Definition 1.1: Optimization Problem

Given $S \subset \mathbb{R}^n$ and a function $f : \mathbb{R}^n \to \mathbb{R}$, and **optimization problem** is a problem of the form:

$$\max f(x) \qquad \text{s.t.} \qquad x \in S \qquad (opt)$$

where

1. S is called **feasible region**;
2. A point $\bar{x} \in S$ is called a **feasible solution**;
3. $f(\bar{x})$ is **objective function valued at** $\bar{x}$.

Remark: (opt) stands for

$$\text{Find } x^* \in S \text{ such that } f(x) \leq f(x^*) \text{ for all } x \in S$$

If x^* is found, it is called an **optimal solution** and $f(x^*)$ is **optimal value**.

Discovery 1.1: Alternative ways of writing (opt)

$$\max\{f(x) : x \in S\} \qquad \text{or} \qquad \max_{x \in S} f(x)$$

Definition 1.2:

Analogous definitions hold for minimization as well.

Result 1.1

Notice that we have

$$\big[\max f(x) \ s.t. \ x \in S\big] = \big[(-1) \cdot \min(-f(x)) \ s.t. \ x \in S\big]$$

This tells that

$$x^* \text{ is opt for max problem} \iff x^* \text{ is opt for min problem}$$

and $f(x^*) = -1 \cdot (-f(x^*))$.

1.1 Problems encountered

We may find problems when solving optimization problems: Optimal solution may not exist:

1. $S = \varnothing$ (problem is **infeasible**);
2. (opt) may be **unbounded**, i.e. $\forall\, \alpha \in \mathbb{R}$, $\exists\, x \in S$ such that $f(x) > \alpha$;
3. Optimal solution does not exist because of limits.

Example 1.1

An example would be

$$\min e^x \quad \text{s.t.} \quad x \in \mathbb{R}$$

Definition 1.3: Supremum

We have

$$\sup\{f(x) : x \in S\} = \begin{cases} -\infty & \text{if } S = \varnothing \\ +\infty & \text{if (opt) is unbounded} \\ \min\{\zeta \in \mathbb{R} : \zeta \geq f(x)\ \forall\, x \in S\} & \text{otherwise} \end{cases}$$

Now **supremum** always exists and is well-defined.

Definition 1.4: Infimum

In terms of **infimum**, we have

$$\inf_{x \in S} f(x) = -\sup_{x \in S}(-f(x))$$

2 Linear Programs (LP)

Definition 2.1: Linear Program

In LPs,

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

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$ and $f(x) = c^T x$ for some $c \in \mathbb{R}^n$. i.e. an LP has the form

$$\max c^T x \ \ s.t. \ \ Ax \leq b$$

Strick inequalities are NOT allowed.

Definition 2.2: Inequality in $\mathbb{R}^m$

For $u, v \in \mathbb{R}^m$,

$$u \leq v \iff u_i \leq v_i,\ \forall\, i = 1, \ldots, m$$

Discovery 2.1

Notice that for $u, v \in \mathbb{R}^m$,

$$u \not\leq v \text{ is not the same as } u > v$$

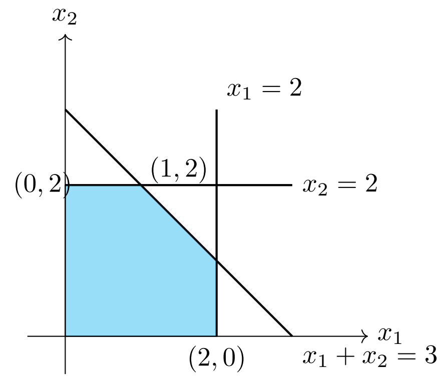
Example 2.1

As an example, consider the LP:

$$\max\ 2x_1 + 0.5x_2 \ \ s.t. \quad \begin{aligned} &x_1 \leq 2 \\ &x_2 \leq 2 \\ &x_1 + x_2 \leq 3 \\ &x \geq 0 \end{aligned}$$

Solution: Thus we have

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 2 & 2 & 3 & 0 & 0 \end{pmatrix}$$



Therefore the optimal solution in this case would be the point $(2, 1)$. $\square$

Lecture 2 - Tuesday, September 10

Definition 2.3: Halfspace, Hyperplane, and Polyhedron

Given $f \in \mathbb{R}^n$, $d \in \mathbb{R}$, the set

$$\{x \in \mathbb{R}^n : f^T x \leq d\}$$

is called a **halfspace**. The set

$$\{x \in \mathbb{R}^n : f^T x = d\}$$

is called a **hyperplane**.

Discovery 2.2

A hyperplane is a generalization of the plane in d-dimensional space, it divides the space into two halfspaces.

The set

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

is called a **polyhedron**.

Discovery 2.3

Notice that a polyhedron is the intersection of numerous (finitely many) halfspaces.

2.1 Determining Feasibility

The question is: Is

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

Notice that for $n = 1$, the problem is easy to solve. For the case when $n = 0$, we define Ax to be the zero vector, thus all that we need to do is to check if b has all its components non-negative. For the general case, the idea is: We wish to reduce the problem to a lower dimension (inductively), so we can solve problems with higher dimensions.

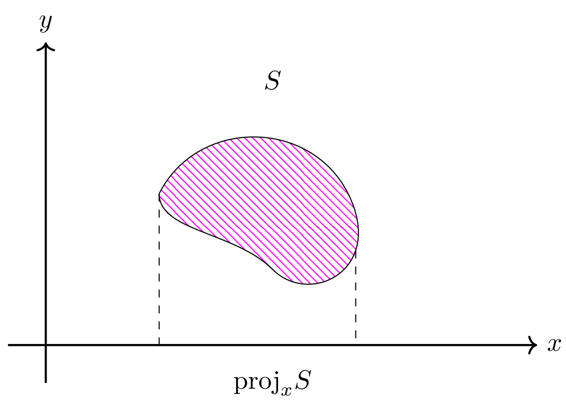
Definition 2.4:

Let $S = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p : Ax + Gy \leq b\}$, then

$$\text{proj}_x(S) := \{x \in \mathbb{R}^n : \ \exists\, y \in \mathbb{R}^p \text{ s.t. } (x, y) \in S\}$$

i.e., the (orthogonal) projection of S onto x.

Example 2.2



Discovery 2.4

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

$$P \neq \varnothing \quad \Longleftrightarrow \quad \text{proj}_{x_1,\ldots,x_{n-1}}(P) \neq \varnothing$$

In other words, the non-emptiness of the polyhedron is equivalent to the the non-emptiness of all the projections.

1. Let a_{ij}'s to be the entries of A;

2. Let

   - $M = \{1, \dots, m\}$;
   - $M^+ = \{i \in M : a_{in} > 0\}$;
   - $M^- = \{i \in M : a_{in} < 0\}$;
   - $M^0 = \{i \in M : a_{in} = 0\}$;

Notice that

$$\begin{aligned} Ax \le b \quad &\iff \quad \sum_{j=1}^{n} a_{ij}x_j \le b_i \;\; \forall\, i = 1, \dots, m \\ &\iff \quad \sum_{j=1}^{n-1} a_{ij}x_j + a_{in}x_n \le b_i \;\; \forall\, i = 1, \dots, m \\ &\iff \quad \begin{cases} \displaystyle\sum_{j=1}^{n-1} \frac{a_{ij}}{a_{in}} x_j + x_n \le \frac{b_i}{a_{in}} & \forall\, i = M^+ \quad (1) \\ \displaystyle\sum_{j=1}^{n-1} \frac{a_{ij}}{a_{in}} x_j + x_n \ge \frac{b_i}{a_{in}} & \forall\, i = M^- \quad (2) \\ \displaystyle\sum_{j=1}^{n-1} a_{ij}x_j \le b_i & \forall\, i = M^0 \quad (3) \end{cases} \end{aligned}$$

By combining (1) and (2) to cancel the like terms, we define

$$\sum_{j=1}^{n-1} \left(\frac{a_{ij}}{a_{in}} - \frac{a_{kj}}{a_{kn}} \right) x_j \le \left(\frac{b_i}{a_{in}} - \frac{b_k}{a_{kn}} \right), \quad \forall\, i \in M^+, k \in M^- \tag{4}$$

Lemma 2.1

We have

$$\underbrace{\operatorname{proj}_{x_1,\dots,x_{n-1}}(P)}_{*} = \underbrace{\{x \in \mathbb{R}^{n-1} : 3 \text{ and } 4\}}_{**}$$

In other words,

$$\exists\, x_n : (x_1, \dots, x_n) \in P \quad \iff \quad \{x \in \mathbb{R}^{n-1} : 3 \text{ and } 4\}$$

Proof. Let $x \in \mathbb{R}^{n-1}$ satisfy $(*)$, then there exists x_n such that $(x_1, \dots, x_n) \in P$, and this implies that (3) holds. But also (1) and (2) hold, and since (4) was obtained from (1) and (2), we conclude that (4) also holds. This tells us that

$$\operatorname{proj}_{x_1,\dots,x_{n-1}}(P) \subseteq \{x \in \mathbb{R}^{n-1} : 3 \text{ and } 4\}$$

Now let $x \in \mathbb{R}^{n-1}$ satisfy $(**)$. Then (because it satisfies (4)), we have

$$\sum_{j=1}^{n-1} -\frac{a_{kj}}{a_{kn}} x_j + \frac{b_k}{a_{kn}} \leq \sum_{j=1}^{n-1} -\frac{a_{ij}}{a_{in}} x_j + \frac{b_i}{a_{in}} \quad \forall\, i \in M^+, k \in M^-$$

Pick $x_n = \max_{k \in M^-}\{LHS\}$, because we have both

$$\sum_{j=1}^{n-1} -\frac{a_{kj}}{a_{kn}} x_j + \frac{b_k}{a_{kn}} \leq -x_n \qquad \text{and} \qquad -x_n \leq \sum_{j=1}^{n-1} -\frac{a_{kj}}{a_{kn}} x_j + \frac{b_k}{a_{kn}}$$

so $(x_1, \ldots, x_n) \in P$. $\square$

2.2 Fourier-Motzkin Elimination

Let $A^n = A$, and $b^n = b$. Given A^i, b^i, obtain A^{i-1} ($i-1$ columns) and b^{i-1},

$$P_i = \{x \in \mathbb{R}^i : A^i x \leq b^i\} \neq \varnothing \quad \Longleftrightarrow \quad P_{i-1} = \{x \in \mathbb{R}^{i-1} : A^{i-1} x \leq b^{i-1}\} \neq \varnothing$$

Definition 2.5:

We denote

$$P_i^n = P_i \times \mathbb{R}^{n-i} \subseteq \mathbb{R}^n$$

Discovery 2.5

Therefore we have

$$P_i = \varnothing \quad \Longleftrightarrow \quad P_i^n = \varnothing$$

and

$$P_i^n = P_i \times \mathbb{R}^{n-i} = \{x \in \mathbb{R}^n : (A^i, 0)x \leq b_i\}$$

Lecture 3 - Thusday, September 12

Example 2.3

We start with

$$P_2 = P_2^2 = \left\{ x \in \mathbb{R}^2 : \begin{array}{l} x_1 + 2x_2 \leq 1 \\ -x_1 \leq 0 \\ -x_2 \leq -2 \\ -3x_1 - 3x_2 \leq -6 \end{array} \right\}$$

Thus we have

$$A^2 = \begin{pmatrix} 1 & 2 \\ -1 & 0 \\ 0 & -1 \\ -3 & -3 \end{pmatrix} \quad \text{and} \quad b^2 = \begin{pmatrix} 1 \\ 0 \\ -2 \\ -6 \end{pmatrix}$$

Moreover, we have

1. $M^+ = \{1\}$;
2. $M^- = \{3,4\}$;
3. $M^0 = \{2\}$.

Using the method introduced above, after scaling each inequality so that the coefficients of x_2 is either 1 or -1, we get

$$\begin{array}{lcl} x_1 + 2x_2 \leq 1 & & \frac{1}{2}x_1 + x_2 \leq \frac{1}{2} \\ -x_1 \leq 0 & \Rightarrow & -x_1 \leq 0 \\ -x_2 \leq -2 & & -x_2 \leq -2 \\ -3x_1 - 3x_2 \leq -6 & & -x_1 - x_2 \leq -2 \end{array}$$

Combining equation 1 and 3, 1 and 4 yields us

$$P_1 = \left\{ x \in \mathbb{R}^1 : \begin{array}{l} -x_1 \leq 0 \\ \frac{1}{2}x_1 \leq -\frac{3}{2} \\ -\frac{1}{2}x_1 \leq -\frac{3}{2} \end{array} \right\}$$

Thus we have

$$A^1 = \begin{pmatrix} -1 \\ \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \quad \text{and} \quad b^1 = \begin{pmatrix} 0 \\ -\frac{3}{2} \\ -\frac{3}{2} \end{pmatrix}$$

Now we have

$$P_1^2 = \left\{ x \in \mathbb{R}^2 : \begin{array}{l} -x_1 \leq 0 \\ \frac{1}{2}x_1 \leq -\frac{3}{2} \\ -\frac{1}{2}x_1 \leq -\frac{3}{2} \end{array} \right\} \Rightarrow \begin{array}{l} -x_1 \leq 0 \\ x_1 \leq -3 \\ -x_1 \leq -3 \end{array}$$

Now we have

1. $M^+ = \{2\}$;
2. $M^- = \{1,2\}$;
3. $M^0 = \varnothing$.

Since

$$P_0^2 = \left\{ x \in \mathbb{R}^2 : \begin{array}{l} 0 \leq -3 \\ 0 \leq -6 \end{array} \right\}$$

We now know that $P_2 = \varnothing$.

Discovery 2.6: Important

1. If A^i, b^i are rational, then A^{i-1}, b^{i-1} are also rational.
2. All inequalities in P_i^n are non-negative combinations of $Ax \leq b$.
3. If $b = 0$, then $b^i = 0$ for all i.

2.3 Farkas' Lemma

Theorem 2.1: Farkas' Lemma

We have
$$P = \{x \in \mathbb{R}^n : Ax \leq b\} = \varnothing \quad \iff \quad \exists\, u \in \mathbb{R}^m \text{ s.t. } u^T A = 0, u^T b < 0, u \geq 0$$

Proof. `For the backward direction`: Suppose $\bar{x} \in P$, then $A\bar{x} \leq b$. Because we know that $u \geq 0$, so $u^T A\bar{x} \leq u^T b$. Therefore, $0 \leq u^T b < 0$ yields a contradiction.
`For the forward direction`: Suppose we know that

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

By Fourier-Motzkin, we have $P_0^n = \varnothing$. This implies that there exists i such that $b_i^0 < 0$. Since $(0)x \leq b^0$ is a non-negative combination of $Ax \leq b$. The constraint corresponding to b_i^0 can be obtained as:

$$u^T Ax \leq u^T b, \qquad \text{for some } u \geq 0$$

Have $u^T A = 0$ and $u^T b = b_i^0 < 0$. □

Definition 2.6: Certificate of Infeasibility

u is called a **certificate of infeasibility**.

Corollary 2.1: Farkas' Lemma Equivalent Statement

Exactly one of the following has a solution:

1. $\exists\, x \in \mathbb{R}^n$ s.t. $Ax \leq b$;
2. $\exists\, u \in \mathbb{R}^m$ s.t. $u^T A = 0, u^T b < 0, u \geq 0$.

2.3.1 Farkas' Lemma (Alternative Form)

Theorem 2.2

Exactly one of the following has a solution:

1. $Ax = b$, $x \geq 0$;
2. $u^T b < 0$, $u^T A \geq 0$.

Proof. **Exercise**. □

2.4 Fundamental Theorem of LP

Theorem 2.3: Fundamental Theorem of LP

Recall that an LP has the form

$$\max c^T x \;\text{ s.t. }\; Ax \le b \tag{LP}$$

then (LP) always has exactly one of the three possible outcomes:

1. infeasible;
2. unbounded;
3. there exists an optimal solution.

Proof. Suppose (1) and (2) are not true. If $n = 1$, then the problem has an optimal solution. Otherwise, we define

$$\max z \;\text{ s.t. }\; z - c^T x \le 0 \;\wedge\; Ax \le b \tag{LP$'$}$$

It is clear that (LP') is neither infeasible or unbounded. Also, if (x^*, z^*) is optimal solution for (LP'), then x^* is optimal solution for (LP). Apply Fourier-Motzkin to

$$\left\{ (x, z) : \begin{array}{l} z - c^T x \le 0 \\ Ax \le b \end{array} \right\}$$

until we get

$$\{ z \in \mathbb{R} : A^1 z \le b^1 \}$$

Then the problem becomes

$$\max z \;\text{ s.t. }\; A^1 z \le b^1$$

which is obviously solvable. Solving this problem to obtain the optimal solution z^*. Reconstruct using Fourier-Motzkin the optimal solution (x^*, z^*) to (LP'). $\square$

Lecture 4 - Thursday, September 19

Discovery 2.7

The Fundamental Theorem of LP applies to *ANY* LP.

3 Determining Optimality

Using an example for illustration:

Example 3.1

$$\max\ 2x_1 + x_2 \quad \text{s.t.} \quad \begin{aligned} x_1 + 2x_2 &\leq 2 \\ x_1 + x_2 &\leq 2 \\ x_1 - x_2 &\leq 0.5 \end{aligned}$$

Easy to notice that $\bar{x} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$ is a feasible solution, but the question is whether or not $\bar{x}$ is optimal. It is easy to observe that the answer is no, since we have $x^* = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$ is a better solution. However, here comes another question. Is x^* optimal?
Notice that any feasible solution satisfies:

$$\begin{aligned} x_1 + 2x_2 &\leq 2 \qquad (\times \tfrac{1}{3}) \\ x_1 + x_2 &\leq 2 \qquad (\times 1) \qquad \Rightarrow_{sum} \quad 2x_1 + x_2 \leq 3 \\ x_1 - x_2 &\leq 0.5 \qquad (\times \tfrac{2}{3}) \end{aligned}$$

Notice that alternatively, we may also have

$$\begin{aligned} x_1 + 2x_2 &\leq 2 \qquad (\times 1) \\ x_1 + x_2 &\leq 2 \qquad (\times 0) \qquad \Rightarrow_{sum} \quad 2x_1 + x_2 \leq 2.5 \\ x_1 - x_2 &\leq 0.5 \qquad (\times 1) \end{aligned}$$

This verifies that $x^* = \begin{pmatrix} 1 \\ 0.5 \end{pmatrix}$ is our optimal solution in this particular example.

Discovery 3.1

In general,

$$\begin{aligned} x_1 + 2x_2 &\leq 2 \qquad (\times y_1) \\ x_1 + x_2 &\leq 2 \qquad (\times y_2) \qquad \Rightarrow_{sum} \quad (y_1 + y_2 + y_3)x_1 + (2y_1 + y_2 - y_3)x_2 \leq 2y_1 + 2y_2 + 0.5y_3 \\ x_1 - x_2 &\leq 0.5 \qquad (\times y_3) \end{aligned}$$

As long as $y_i \geq 0$ for all i, $y_1 + y_2 + y_3 = 2$, and $2y_1 + y_2 - y_3 = 1$, we have

$$c^T x \leq 2y_1 + 2y_2 + 0.5y_3$$

for any feasible solution x.

Result 3.1

To get the best possible upper bound, we should solve:

$$\min\ 2y_1 + 2y_2 + 0.5y_3 \quad \text{s.t.} \quad \begin{aligned} y_1 + y_2 + y_3 &= 2 \\ 2y_1 + y_2 - y_3 &= 1 \\ y_1, y_2, y_3 &\geq 0 \end{aligned}$$

Definition 3.1: Dual

The above LP is called a **dual LP**.

In general, we have

Definition 3.2: Primal

$$\max\ c^T x \quad \text{s.t.} \quad Ax \leq b \tag{P}$$

$$\min\ b^T y \quad \text{s.t.} \quad \begin{aligned} A^T y &= c \\ y &\geq 0 \end{aligned} \tag{D}$$

3.1 Weak Duality Theorem

Theorem 3.1: Weak Duality

If $\bar{x}$ is feasible for (P), $\bar{y}$ is feasible for (D), then

$$c^T x \leq b^T y$$

Proof. If $\bar{x}$ is feasible for (P), then

$$A\bar{x} \leq b$$

But since $\bar{y}$ is feasible for (D) (particularly that $\bar{y} \geq 0$):

$$c^T \bar{x} = (A^T \bar{y})^T = \bar{y}^T A\bar{x} \leq \bar{y}^T b = b^T \bar{y}$$

as desired. □

Discovery 3.2

1. There is one variable in (D) for every constraint in (P). In other words, the size of y is the same as the size of b.
2. There is one constraint in (D) for every variable in (P) (plus the non-negativity).

Corollary 3.1

If (P) is unbounded, then (D) is infeasible.

Corollary 3.2

If (D) is unbounded, then (P) is infeasible.

Exercise: There are examples where (P) and (D) are both infeasible.

Corollary 3.3

If x^* is feasible for (P), y^* is feasible for (D), and $c^T x^* = b^T y^*$, then x^* is the optimal solution for (P) and y^* is the optimal solution for (D).

Result 3.2: Possible Outcomes

(D) \ (P)	unbounded	infeasible	optimal
unbounded	NO	YES	NO
infeasible	YES	YES	?
optimal	NO	?	?

3.2 Strong Duality Theorem

Theorem 3.2: Strong Duality

If (P) has an optimal solution, denoted as x^*, then (D) also has an optimal solution y^* such that

$$c^T x^* = b^T y^*$$

Proof. `Part 1`: We first prove that (D) is feasible.
Suppose (D) is infeasible, i.e.,

$$\left\{ y \in \mathbb{R}^m : \begin{array}{l} A^T y = c \\ y \geq 0 \end{array} \right\} = \varnothing$$

By the alternative form of Farkas' Lemma, we know that

$$\exists\, u \;\text{ s.t. } \begin{array}{l} u^T A^T \geq 0 \\ u^T c < 0 \end{array} \iff \exists\, u \;\text{ s.t. } \begin{array}{l} Au \geq 0 \\ c^T u < 0 \end{array} \iff \exists\, d \;\text{ s.t. } \begin{array}{l} Ad \leq 0 \\ c^T d > 0 \end{array}$$

But then $x^* + d$ is feasible for (P) since

$$A(x^* + d) = Ax^* + Ad \leq b + 0$$

However, notice that

$$c^T(x^* + d) = c^T x^* + c^T d > c^T x^*$$

which contradicts the fact that x^* is our optimal solution for (P), thus we must have that (D) is feasible.
`Part 2`: By the Fundamental Theorem of LP and Weak Duality Theorem, (D) has an optimal solution.
`Part 3`: denote $\gamma = b^T y^*$, and we consider

$$\Theta = \left\{ x \in \mathbb{R}^n : \begin{array}{l} Ax \leq b \\ -c^T x \leq -\gamma \end{array} \right\}$$

If $\Theta = \varnothing$, then by Farkas' Lemma,

$$\exists \begin{pmatrix} y \\ \lambda \end{pmatrix} : \begin{array}{l} \begin{pmatrix} y \\ \lambda \end{pmatrix}^T \begin{pmatrix} A \\ -c^T \end{pmatrix} = 0 \\ \begin{pmatrix} y \\ \lambda \end{pmatrix}^T \begin{pmatrix} b \\ -\gamma \end{pmatrix} < 0 \\ \begin{pmatrix} y \\ \lambda \end{pmatrix} \geq 0 \end{array} \iff \exists\, y \in \mathbb{R}^m, \lambda \in \mathbb{R}^n : \begin{array}{l} A^T y = c\lambda \\ y^T b < \lambda\gamma \\ y \geq 0, \lambda \geq 0 \end{array}$$

1. *Case 1*: $\lambda > 0$
   Hence we have
   $$\begin{aligned} &A^T (y/\lambda) = c \\ &b^T (y/\lambda) < \gamma \\ &(y/\lambda) \geq 0 \end{aligned}$$
   which contradicts the fact that y^* is the optimal solution for (D).

2. *Case 2*: $\lambda = 0$
   Hence we have
   $$\begin{aligned} &A^T y = 0 \\ &b^T y < 0 \\ &y \geq 0 \end{aligned}$$
   Now we have
   $$A^T (y^* + y) = A^T y^* + A^T y = A^T y^* = c$$
   but
   $$b^T (y^* + y) = b^T y^* + b^T y < b^T y^*$$
   which again contradicts the fact that y^* is the optimal solution for (D).

□

Lecture 5 - Tuesday, September 24

Definition 3.3: Certificate of Optimality

The y^* introduced above is a **certificate of optimality**.

Result 3.3

In general, same results hold for

$$\max c^T x \quad \text{s.t.} \quad \begin{aligned} Ax \ ? \ b \\ x \ ? \ 0 \end{aligned} \tag{P}$$

$$\min b^T y \quad \text{s.t.} \quad \begin{aligned} A^T y \ ? \ c \\ y \ ? \ 0 \end{aligned} \tag{D}$$

where the ? are replaced in regards to the chart below:

(P) max		(D) min	
	$\leq$		≥ 0
Constr	$\geq$	Var	≤ 0
	$=$		free
	≥ 0		$\geq$
Var	≤ 0	Constr	$\leq$
	free		$=$

Example 3.2

$$\max 2x_1 + 3x_2 - 4x_3 \quad \text{s.t.} \quad \begin{array}{cccl} x_1 & & +7x_3 & \leq 5 \\ & 2x_2 & -x_3 & \geq 3 \\ x_1 & & +x_3 & = 8 \\ & x_2 & & \leq 6 \\ & x_1 \geq 0, & x_2 \leq 0 & \end{array} \tag{P}$$

Then we would have the dual represented as

$$\min 5y_1 + 3y_2 + 8y_3 + 6y_4 \quad \text{s.t.} \quad \begin{array}{ccccl} y_1 & & +y_3 & & \geq 2 \\ & 2y_2 & & +y_4 & \leq 3 \\ 7y_1 & -y_2 & +y_3 & & = -4 \\ y_1 \geq 0, & y_2 \leq 0, & y_4 \geq 0 & & \end{array}$$

3.3 What if primal is minimization

Example 3.3

$$\min x_1 - x_2 \quad \text{s.t.} \quad \begin{array}{ccl} 2x_1 & +2x_2 & \leq 5 \\ x_1 & -x_2 & \geq 3 \\ x_1 & 5x_2 & = 7 \\ x_1 \geq 0, & x_2 \leq 0 & \end{array} \tag{P}$$

$$\max\ 5y_1 + 3y_2 + 7y_3 \quad \text{s.t.} \quad \begin{array}{llll} 2y_1 & +y_2 & +y_3 & \leq 1 \\ 3y_1 & -y_2 & +5y_3 & \geq -1 \\ \multicolumn{4}{l}{y_1 \leq 0, \quad y_2 \geq 0} \end{array} \tag{D}$$

Notice that here we use the "opposite" of the above chart.

Theorem 3.3: Weak Duality for Primal as a Minimalization

If $\bar{x}$ is feasible for (P) and $\bar{y}$ is feasible for (D), then

$$c^T\bar{x} \geq b^T\bar{y}$$

Theorem 3.4

Strong duality still holds for (P) in minimization form.

Discovery 3.3

If (P) is primal and (D) is dual, then the dual of (D) is (P).

3.4 Interpretations of the dual

3.4.1 Maximizing Profit

I wish to decide how much of two products to produce, and I have two resources available:

	Per unit profit	Resource Consumption A	Resource Consumption B
Product A	5	2	3
Product B	3	4	1

Suppose we have 15 units of resource A and 10 units of resource B. Therefore we wish to solve

$$\max\ 5x_1 + 3x_2 \quad \text{s.t.} \quad \begin{array}{lll} 2x_1 & +4x_2 & \leq 15 \\ 3x_1 & +x_2 & \leq 10 \end{array}$$

Suppose I am willing to sell resource A for y_1 dollars and resource B for y_2 dollars. Notice that with 2 units of resource A and 3 units of resource B, I earn at least \$5, thus

$$2y_1 + 3y_2 \geq 5$$

Similarly,

$$4y_1 + y_2 \geq 3$$

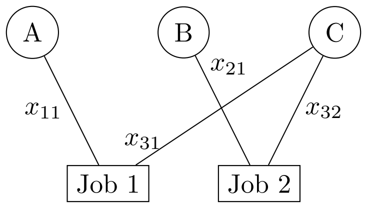
Thus I wish to solve

$$\min\ 15y_1 + 10y_2 \quad \text{s.t.} \quad \begin{array}{lll} 2y_1 & +3y_2 & \geq 5 \\ 4y_1 & +y_2 & \geq 3 \end{array}$$

to figure out the least amount I could sell the resources for.

3.4.2 Vertex Cover Problem

Suppose I have the following: Three job candidates A, B, and C applying for two job positions. A, B, and C applied for job position 1 and C has applied for job position 2.



What is the largest number of matches?

We wish to solve

$$\max x_{11}+x_{21}+x_{31}+x_{32} \quad \text{s.t.} \quad \begin{array}{ccccl} x_{11} & & & & \leq 1 \\ & x_{21} & & & \leq 1 \\ & & x_{31} & +x_{32} & \leq 1 \\ x_{11} & & +x_{31} & & \leq 1 \\ & x_{21} & & +x_{32} & \leq 1 \\ & & & x & \geq 0 \end{array}$$

Remark: the optimal solution to this is integral (not proven yet).

Notice that the dual for this problem would become

$$\min y_1+y_2+y_3+y_4+y_5 \quad \text{s.t.} \quad \begin{array}{cccccl} y_1 & & & +y_4 & & \geq 1 \\ & y_2 & & & +y_5 & \geq 1 \\ & & y_3 & +y_4 & & \geq 1 \\ & & y_3 & & +y_5 & \geq 1 \\ & & & & y & \geq 0 \end{array}$$

Definition 3.4: Vertex Cover Problem

Pick set of vertices with the minimum size such that every line (edge) touches at least one of the vertices in the picked set.

Theorem 3.5

The optimal solution to the LP above (Vertex Cover Problem) has all variable either 0 or 1.

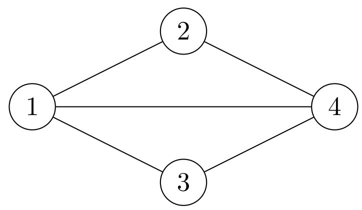
Lecture 6 - Thursday, September 26

4 Graphs

Definition 4.1: Graph

A **graph** G is denoted by $G = (V, E)$, where V is a finite set representing the **vertices** (nodes) and E is a subset of pairs of vertices, or the set of **edges**.

Example 4.1



Thus we have

$$V = \{1, 2, 3, 4\}$$
$$E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}, \{1, 3\}\}$$

Definition 4.2: Endpoint & Incident

A vertex v is an **endpoint** of an edge e if v is one of the vertices in the pair of e. In this case, we say edge e is **incident** of v.

Example 4.2

In the above example, node 2 is an endpoint of edge $\{2, 4\}$ (sometimes we write edge 24 for simplicity), and edge 24 is incident to 2, and 4.

Definition 4.3:

We define

$$\delta(v) := \{e \in E : e \texttt{ is incident to } v\}$$

Definition 4.4:

Let $S \subseteq V$, we define

$$\delta(S) := \{uv \in E : u \in S \wedge v \notin S\}$$

Definition 4.5: Adjacent

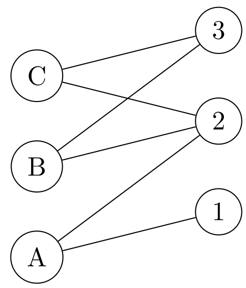
Two vertices $u, v \in V$ are **adjacent** if $\{u, v\} \in E$.

Definition 4.6: Bipartite

A graph G is called **bipartite** if $V = A \sqcup B$. Consequently, if

$$u, v \text{ are adjacent} \quad \Rightarrow \quad u \in A \wedge v \in B$$

Example 4.3



4.1 Max Cardinality Matching in Bipartite Graphs

Definition 4.7: Matching

$M \subseteq E$ is called a **matching** if

$$|\delta(v) \cap M| \leq 1, \qquad \forall\, v \in V$$

Here is the problem: Given $G = (A \sqcup B, E)$ a bipartite, we wish to find the matching M of G with largest cardinality.

Definition 4.8: Decision Variables

We define

$$x_e := \begin{cases} 1 & \text{if } e \in M \\ 0 & \text{if } e \notin M \end{cases}$$

Therefore, our problem becomes

$$\max \sum_{e \in E} x_e \qquad \text{s.t.} \qquad \begin{array}{l} \sum_{e \in \delta(v)} x_e \leq 1 \quad \forall\, v \in V \\ 0 \leq x_e \leq 1, \quad \forall\, e \in E \\ x \in \mathbb{Z}^E \end{array}$$

Recall that for the dual, we have one variable for each of the constraints, thus we may write down the dual:

$$\min \sum_{v \in V} y_v \qquad \text{s.t.} \qquad \begin{array}{l} y_u + y_v \geq 1 \quad \forall\, uv \in E \\ y_v \geq 0, \quad \forall\, v \in V \end{array}$$

Example 4.4

Here is another example. Consider $G = (V, E)$ as a given graph. Let $c_e \geq 0$ for all $e \in E$ be edge costs. Denote the decision variable as x_e for all $e \in E$.
Consider the LP:

$$\min \sum_{e \in E} c_e x_e \qquad \text{s.t.} \qquad \begin{aligned} &\sum_{e \in \delta(v)} x_e = 2, \quad \forall\, v \in V \\ &\sum_{e \in \delta(s)} x_e \geq 2, \quad \forall\, S \subsetneq V \wedge |S| > 1 \\ &x \geq 0 \end{aligned}$$

Now we have the dual:

$$\min 2 \sum_{v \in V} y_v + 2 \sum_{S \subsetneq V : |S| > 1} y_S \quad \text{s.t.} \quad \begin{aligned} &y_u + y_v + \sum_{S \subsetneq V : |S| > 1 \wedge uv \in S} y_S \leq c_{uv}, \quad \forall\, uv \in E \\ &y_S \geq 0, \quad \forall\, S \subsetneq V : |S| > 1 \end{aligned}$$

4.2 Complementary Slackness

Theorem 4.1

Let x^* be feasible for primal LP, y^* be feasible for dual LP. Then

(i). Either $x_j^* = 0$, or the corresponding dual constraint is tight at y^*, for all $j = 1, \ldots, n$;

(ii). Either $y_i^* = 0$, or the corresponding dual constraint is tight at x^*, for all $i = 1, \ldots, m$.

(inclusive or, could be both).

Example 4.5

Suppose we have primal:

$$\min x_1 - x_2 \quad \text{s.t.} \quad \begin{array}{ccl} 2x_1 & +3x_2 & \leq 5 \\ x_1 & -x_2 & \geq 3 \\ x_1 & +5x_2 & = 7 \\ x_1 \geq 0, & x_2 \leq 0 & \end{array}$$

Thus the dual would be:

$$\max 5y_1 + 3y_2 + 7y_3 \quad \text{s.t.} \quad \begin{array}{cccl} 2y_1 & +y_2 & +y_3 & \leq 1 \\ 3y_1 & -y_2 & +5y_3 & \geq -1 \\ y_1 \leq 0, & y_2 \geq 0 & & \end{array}$$

Solution: Then the C.S. tells us that

1. • $x_1^* = 0$ or $2y_1^* + y_2^* + y_3^* = 1$;
   • $x_2^* = 0$ or $3y_1^* - y_2^* + 5y_3^* = -1$;

2. • $y_1^* = 0$ or $2x_1^* + 3x_2^* = 5$;
   • $y_2^* = 0$ or $x_1^* - x_2^* = 3$;
   • $y_3^* = 0$ or $x_1^* + 5x_2^* = 7$;

as an example. □

Theorem 4.2: C.S.

Let x^* be feasible for primal and y^* be feasible for dual, TFAE:

1. x^* is optimal solution for primal and y^* is optimal solution for dual;
2. $c^T x^* = b^T y^*$;
3. x^* and y^* satisfy C.S.

Lecture 7 - Tuesday, October 01

Proof. `[1]` $\iff$ `[2]` This holds because this is equivalent to strong duality.
`[2]` $\iff$ `[3]` Assume we have

$$\max c^T x \qquad \text{s.t.} \qquad \begin{aligned} Ax &\le b \\ x &\ge 0 \end{aligned} \tag{P}$$

$$\min b^T y \qquad \text{s.t.} \qquad \begin{aligned} A^T y &\ge c \\ y &\ge 0 \end{aligned} \tag{D}$$

We know that

$$c^T x^* = \sum_{j=1}^{n} c_j x_j^*$$

and the constraint of the dual gives us that

$$c_j \le \sum_{i=1}^{m} a_{ij} y_i^*$$

Because $x \ge 0$, we have

$$c_j x_j^* \le \sum_{i=1}^{m} a_{ij} y_i^* x_j^* \tag{1}$$

hence we have

$$c^T x^* \le \sum_{i=1}^{n} \left(\sum_{j=1}^{m} a_{ij} x_j^* \right) y_i^*$$

Feasibility of (P) yields that

$$\sum_{j=1}^{n} a_{ij} x_j^* \le b_i \quad \forall\, i \quad \Rightarrow \quad \left(\sum_{j=1}^{n} a_{ij} x_j^* \right) y_i^* \le b_i y_i^* \tag{2}$$

Therefore, we have

$$c^T x^* \leq \sum_{i=1}^{m} b_i y_i^* = b^T y^*$$

where the equality holds if and only if (1) and (2) both hold for equalities if and only if C.S. holds. $\square$

4.2.1 Geometric Interpretations of C.S.

$$\max c^T x \qquad \text{s.t.} \qquad Ax \leq b \tag{P}$$

$$\min b^T y \qquad \text{s.t.} \qquad \begin{matrix} A^T y = c \\ y \geq 0 \end{matrix} \tag{D}$$

We know that A can be written in the form of

$$A = \begin{pmatrix} --- & a_1^T & --- \\ --- & a_1^T & --- \\ & \vdots & \\ --- & a_m^T & --- \end{pmatrix}$$

C.S. (ii) tells us that

$$y_i^* = 0 \qquad \text{or} \qquad a_i^T x^* = b_i$$

and we know

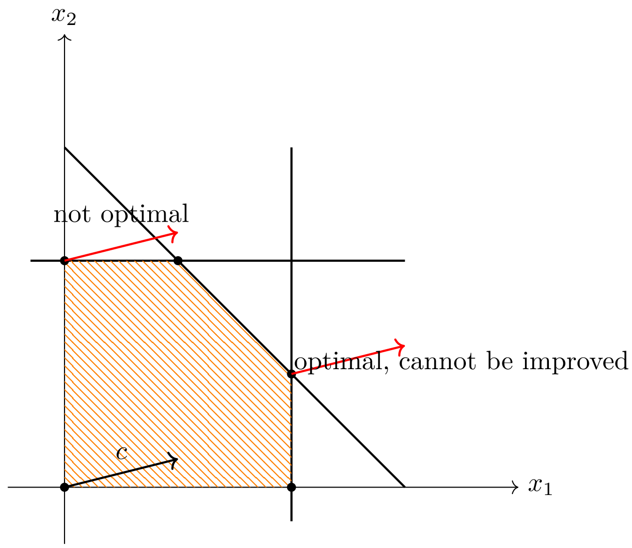
$$A^T y = c \quad \Longleftrightarrow \quad \sum_{i=1}^{m} a_i y_i = c$$

and $y_i^* = 0$ for all constraints that are not tight at x^*. Thus, c is a non-negative combination of coefficients that are tight constraints.

Example 4.6

Consider the example

$$\max 2x_1 + 0.5x_2 \;\; s.t. \quad \begin{array}{ccr} x_1 & & \leq 2 \\ & x_2 & \leq 2 \\ x_1 & +x_2 & \leq 3 \\ & & x \geq 0 \end{array}$$



Discovery 4.1

c is a non-negative combination of coefficients that are tight constraints if and only if c is in the cone of tight constraints.

4.3 Characterizing Unboundedness

Theorem 4.3

The problem

$$\max c^T x \qquad \text{s.t.} \qquad Ax \leq b$$

is unbounded if and only if it is feasible, and

$$\exists\, d : Ad \leq 0, c^T d > 0$$

Proof. [$\Longleftarrow$] Let $\bar{x}$ be such that $A\bar{x} \leq b$, then for any $\alpha \in \mathbb{R}$, consider $x^* + \beta d$ for $\beta \geq 0$, we have

$$A(x^* + \beta d) = Ax^* + \beta Ad \leq b$$

and

$$c^T(x^* + \beta d) > \alpha$$

for some choice of $\beta > 0$, since

$$c^T x^* + \beta c^T d > \alpha \qquad \text{for} \qquad \beta > \frac{\alpha - c^T x^*}{c^T d}$$

[$\Longrightarrow$] Consider the dual

$$\min b^T y \qquad \text{s.t.} \qquad \begin{aligned} A^T y &= c \\ y &\geq 0 \end{aligned}$$

We know that (P) is unbounded implies that (D) is infeasible. Thus by Farkas' lemma, we know that there exists $u \in \mathbb{R}^n$ such that $u^T A^t \geq 0$ and $u^T c < 0$. Take $d = -u$. □

4.4 Geometry of polyhedra

Definition 4.9: Line Segment

Given $\bar{x}, \bar{y} \in \mathbb{R}^n$, the line segment between $\bar{x}$ and $\bar{y}$ is the set of points

$$\{x \in \mathbb{R}^n : x = \lambda\bar{x} + (1-\lambda)\bar{y} \text{ s.t. } \lambda \in [0,1]\}$$

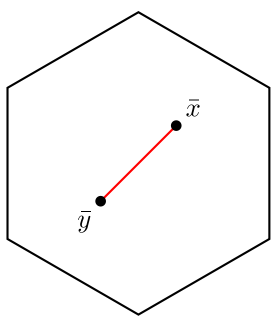
Definition 4.10: Convex

A set $S \subseteq \mathbb{R}^n$ is a convex set if for all $\bar{x}, \bar{y} \in S$, the line segment between $\bar{x}$ and $\bar{y}$ is contained in S.

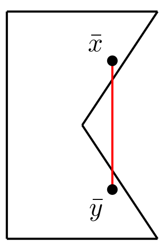
Result 4.1

$\varnothing$ is convex.

Example 4.7



Convex.



Not convex.

Discovery 4.2

Polyhedra are convex sets.

Proof. For

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

Let $\bar{x}, \bar{y} \in P$, and $\lambda \in [0,1]$. We have

$$A[\lambda\bar{x} + (1-\lambda)\bar{y}] \leq \lambda b + (1-\lambda)b$$

as desired. □

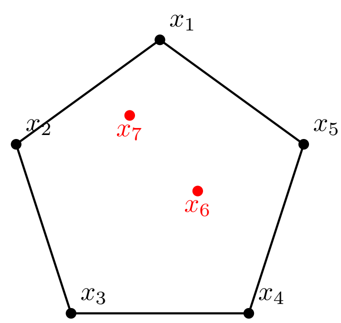
Definition 4.11: Convex Combination

Generally, for $x_1, \ldots, x_k \in \mathbb{R}^n$, we say that x is a **convex combination** of $x_1, \ldots, x_k$ if there exists $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that

$$x = \sum_{i=1}^{k} \lambda_i x_i$$

with $\sum_i \lambda_i = 1$ and $\lambda_i \geq 0$ for all $i = 1, \ldots, k$.

Example 4.8



The region enclosed by the shape to the left is the set of all possible convex combinations of $x_1, \ldots, x_7$.

Discovery 4.3

The idea is that optimal solution happens at "corners".

Lecture 8 - Thursday, October 03

4.5 Basic Definitions in Polyhedra

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

Definition 4.12: Extreme Point

$\bar{x} \in P$ is an **extreme point** of P if $\nexists\, u, v \in P \backslash \{\bar{x}\}$ and $\lambda \in [0, 1]$ such that $\bar{x} = \lambda u + (1 - \lambda)v$.

Definition 4.13: Basic Feasible Solution

$\bar{x} \in P$ is a **basic feasible solution** if there exist n linearly independent constraints $a_i^T x \leq b_i$ that tight at $\bar{x}$.

Definition 4.14: Vertex

$\bar{x} \in P$ is a vertex of P if there exists $c \in \mathbb{R}^n$ so that $\bar{x}$ is the unique optimal solution to max $c^T x$ such that $x \in P$.

Theorem 4.4

TFAE:

1. $\bar{x}$ is a vertex of P;
2. $\bar{x}$ is a basic feasible solution of P;
3. $\bar{x}$ is an extreme point of P.

Proof. `[1 ⇒ 3]`:
Suppose $\bar{x}$ is not an extreme point, so

$$\bar{x} = \lambda u + (1-\lambda)v \qquad \text{for } \ u, v \in P \backslash \{\bar{x}\}, \ \lambda \in (0,1)$$

and this gives us that $c^T\bar{x} = \lambda c^T u + (1-\lambda)c^T v$ for any $c \in \mathbb{R}^n$. However, for $\bar{x}$ to be the unique optimal solution, we would have

$$c^T x = \lambda c^T x + (1-\lambda)c^T x > \lambda c^T u + (1-\lambda)c^T v$$

which holds for strick inequality. Notice that we now obtain a contradiction, thus $\bar{x}$ is a vertex of P implies that $\bar{x}$ is an extreme point of P.
`[3 ⇒ 2]`:
Suppose $\bar{x}$ is not a basic feasible solution. Let $I \subseteq \{1, \ldots, m\}$ be the indices of constraints tight at $\bar{x}$, and A_I be the matrix obtained by deleating the rows from A that are not in I. Hence we have

$$\operatorname{rank}(A_I) < n$$

Consider $A_I d = 0$. By rank-nullity theorem, we know that there exists $d \neq 0$ such that the equality holds. Let $\varepsilon > 0$ and let $u = \bar{x} + \varepsilon d$, $v = \bar{x} - \varepsilon d$. Clearly, we have $\bar{x} = 0.5u + 0.5v$. Moreover, we have $u, v \in P$ because at the tight constraints a_i we have $a_i d = 0$ and for the remaining constraints we may choose ε small, thus this yields us a contradiction. `[2 ⇒ 1]`:
Let I be the set of indices of tight constraints at $\bar{x}$. Let $c = \sum_{i=I} a_i$. Thus for all $x \in P$, we have

$$c^T x = \sum_{i \in I} a_i^T x \leq \sum_{i \in I} b_i = \sum_{i \in I} a_i^T \bar{x} = c^T \bar{x}$$

Notice that for $c^T x = c^T \bar{x}$, we must have $a_i^T x = b_i$ for all $i \in I$, hence the solution is unique. $\square$

Theorem 4.5

Let P be a polyhedron with at least one extreme point, then if

$$\max \ c^T x \qquad \text{s.t.} \qquad x \in P$$

has an optimal solution, there exists an optimal solution which is an extreme point.

Proof. Let $\bar{x}$ be an optimal solution to

$$\max c^T x \qquad \text{s.t.} \qquad x \in P = \{x : Ax \leq b\}$$

Let I be the index set of tight constraints at $\bar{x}$. If $\text{rank}(A_n) = n$, then we are done. Otherwise, pick $d \neq 0$ such that $A_I d = 0$. For some $\varepsilon > 0$, we know that $\bar{x} \pm \varepsilon d \in P$. Because we know that $\bar{x}$ is the optimal solution, we have

$$c^T x + \varepsilon x^T d \leq c^T x \quad \Rightarrow \quad c^T d \leq 0$$

Similarly, because $c^T x - \varepsilon x^T d \leq c^T x$, we get $c^T d \geq 0$, thus we have $c^T d = 0$. This tells us that $\bar{x} \pm \varepsilon d \in P$ are also optimal solutions. Suppose $\forall\, \varepsilon > 0$, $\bar{x} \pm \varepsilon d \in P$, then P does not have an extreme point (**proof will be later**). This means that for the largest ε for which $\bar{x} \pm \varepsilon d \in P$, some constraints $a_i^T x \leq b_i, i \notin I$ will become tight, so $a_i^T d \neq 0$. Hence a_i is linearly independent from A_I. $\square$

Now the questions arises:

`when does a polyhedron attain extreme point`

Lecture 9 - Tuesday, October 08

Definition 4.15: Line

Let $x, d \in \mathbb{R}^n$ with $d \neq 0$, the set

$$\{x + \lambda d : d \in \mathbb{R}\}$$

is called a **line**.

Definition 4.16:

We say a polyhedron P **has a line** if there exists $\bar{x} \in P$ and $d \neq 0 \in P$ such that

$$\{\bar{x} + \lambda d : d \in \mathbb{R}\} \subseteq P$$

Proposition 4.1

$P = \{x \in \mathbb{R}^n : Ax \leq b\}$ has a line if and only if $P \neq \varnothing$ and there exists $d \neq 0$ such that $Ad = 0$ if and only if $P \neq \varnothing$ and $\text{rank}(A) < n$.

Proof. It is easy to see that

$$\text{there exists } d \neq 0 \text{ such that } Ad = 0 \text{ if and only if } P \neq \varnothing \text{ and } \text{rank}(A) < 0$$

`[2` $\Rightarrow$ `1]`
Since P is non-empty, we have $\bar{x} \in P$ for some $\bar{x}$, then we have

$$A(\bar{x} + \lambda d) = A\bar{x} + \lambda A d = A\bar{x} \leq b \qquad \forall\, \lambda \in \mathbb{R}$$

which implies that $\bar{x} + \lambda d \in P$ for all $\lambda \in \mathbb{R}$.
`[1 ⇒ 2]`
Since P has a line, we know that there exists $\bar{x} \in P$, hence $P \neq \varnothing$. Suppose $a_i^T d > 0$, then there exists λ such that

$$a_i^T(\bar{x} + \lambda d) > b_i$$

which implies that P does not have the line, contradiction. □

Theorem 4.6

A polyhedron P has an extreme point if and only if $P \neq \varnothing$ and P has no lines.

Proof. `Forward direction`:
We prove this using contrapositive. If $P = \varnothing$, then P has no extreme point. If P is non-empty and P has a line, then there exists $d \in \mathbb{R}^n$, $d \neq 0$ such that $Ad = 0$. For any $x \in P$, we have

$$x = \frac{1}{2}(x + d) + \frac{1}{2}(x - d)$$

where both $1/2(x + d)$ and $1/2(x - d)$ are feasible solutions.
`Backward direction`:
For any $x \in P$, let $I(x)$ be the indices of constraints tight at x. Let $\bar{x} \in P$ be a point with largest rank $\left(A_{I(\bar{x})}\right)$. If the rank is n, we are done. Otherwise, we find d with $d \neq 0$ such that $A_{I(\bar{x})}d = 0$, which implies that we must have some $i \in 1, \ldots, m \backslash I(\bar{x})$ such that

$$a_i^T d \neq 0$$

(else it would have a line). Therefore, we can find a point with more linear independent tight constraints than $\bar{x}$, which is a contradiction. □

Definition 4.17: Pointed

A non-empty polyhedron that has no lines is called **pointed**.

5 Midterm

This is the midterm cuff-off line.

6 Cones and Extreme Rays

6.1 Cone

Definition 6.1: Cone

A set $C \subseteq \mathbb{R}^n$ is a cone if $\forall\, x \in C$, $\forall\, d \in \mathbb{R}$ with $d \geq 0$, $\lambda x \in C$.

Example 6.1

A line $C = \{c(1,1) : c \in \mathbb{R}\} \subseteq \mathbb{R}^2$ is a cone.

Discovery 6.1

In our class, we do not need the requriement that for $x^1, x^2 \in C$, $x^1 + x^2 \in C$. We will instead call this a **convex cone**.

Definition 6.2: Polyhedral Cone

A cone C is called a **polyhedral cone** if C can be written in the form of

$$C = \{x \in \mathbb{R}^n : Ax \leq 0\}$$

Definition 6.3: Recession Cone

Given a polyhedron $P \neq \varnothing$, its **recession cone** is

$$\operatorname{rec}(P) = \{r \in \mathbb{R}^n : \bar{x} + \lambda r \in P,\ \ \forall\, \bar{x} \in P,\ \ \forall\, \lambda \geq 0\}$$

Theorem 6.1

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

$$\operatorname{rec}(P) = \{r \in \mathbb{R}^n : Ar \leq 0\}$$

Proof. It is easy to see that

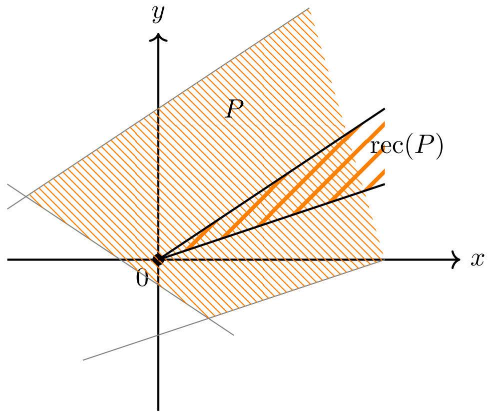
$$\operatorname{rec}(P) \supseteq \{r : Ar \leq 0\}$$

Pick $r \notin \{r : Ar \leq 0\}$, so $a_i^T r > 0$ for some $i \in \{1, \ldots, m\}$. Then there exists $\lambda > 0$ such that

$$\bar{x} + \lambda r \notin P \quad \Rightarrow \quad r \notin \operatorname{rec}(P)$$

hence we must have $\operatorname{rec}(P) \subseteq \{r : Ar \leq 0\}$. $\square$

Example 6.2



6.2 Extreme Rays

Definition 6.4: Extreme Rays

Let C be a cone, $r \in C$ is an **extreme ray** of C if $r \neq 0$ and for any $r^1, r^2 \in C$ such that r is the line segment between r^1 and r^2, then r^1 and r^2 are non-negative multiples of r.

Theorem 6.2

Let $C = \{x \in \mathbb{R}^n : Ax \leq 0\}$ be pointed. Then $r \in C$ is an extreme ray of C if and only if $r \neq 0$ and there are $n-1$ linearly independent constraints tight at r.

Lecture 10 - Thursday, October 10

Proof. Let

$$\alpha = \sum_{i=1}^{m} a_i \qquad \text{and} \qquad A = \begin{pmatrix} - & a_1^T & - \\ & \vdots & \\ - & a_m^T & - \end{pmatrix}$$

Note that if $r \neq 0$, $r \in C$. Then there exists $i : a_i^T r < 0$ since otherwise we have $Ar = 0$, which would imply $r = 0$ because $\text{rank}(A) = n$. Therefore, $\alpha^T r < 0$. Take

$$P = \left\{ x \in \mathbb{R}^n : \begin{array}{ll} Ax & \leq 0 \\ \alpha^T x & \geq -1 \end{array} \right\}$$

Recall that

$$\text{rec}(P) = \left\{ x \in \mathbb{R}^n : \begin{array}{ll} Ax & \leq 0 \\ \alpha^T x & \geq 0 \end{array} \right\}$$

But $Ax \leq 0$ implies that $\alpha^T x \leq 0$, and this tells us that if $x \in \text{rec}(P)$, then x satisfies $\alpha^T x = 0$. Hence $Ax = 0$ implies that $x = 0$. Now let $\hat{x}$ be an extreme point of P. If $\alpha^T \hat{x} > -1$, then the n linearly

independent constraints tight at $\hat{x}$ come from $Ax \leq 0 \;\; \Rightarrow \;\; \hat{x} = 0$.
Otherwise, we have $\alpha^T \hat{x} = -1$, there are $n-1$ linearly independent constraints from $Ax \leq 0$ tight at $\hat{x}$.
Now STP r is an extreme ray of C if and only if there exists $\beta > 0$ such that $r = \beta \hat{x}$ for some extreme point $\hat{x}$ of P that satisfies $\alpha^T \hat{x} = -1$.
Remark: Note that for every $r \neq 0$, $r \in C$, we have $\frac{r}{|\alpha^T r|}$ is a point in P satisfying $\alpha^T \left(\frac{r}{|\alpha^T r|}\right) = -1$.
($\Longrightarrow$) Hence we pick $\beta = |\alpha^T r|$, and let $\hat{x} = \frac{1}{\beta} r$, so $\alpha^T \hat{x} = -1$. Now suppose $\hat{x} = 0.5x^1 + 0.5x^2$ for $x^1, x^2 \in P \backslash \{x\}$. Then

$$\underbrace{\beta \hat{x}}_{r} = 0.5 \underbrace{\beta x^1}_{r^1} + 0.5 \underbrace{\beta x^2}_{r^2}$$

But since r was an extreme ray, r^1 and r^2 are multiples of r. Thus

$$r^1 = \gamma^1 r, \quad r^2 = \gamma^2 r \qquad \text{for } \gamma^1, \gamma^2 > 0$$

But x^1 and x^2 must satisfy $\alpha^T x^1 = \alpha^T x^2 = -1$. Then

$$x^1 = \frac{r^1}{|\alpha^T r^1|} = \frac{\gamma^1 r}{\gamma^1 |\alpha^T r|} = \hat{x}$$

which is a contradiction.
($\Longleftarrow$) analugous. $\square$

Definition 6.5: Ray

If $P \neq \varnothing$ and r is a ray (extreme ray) of rec(P), we say r is a ray (extreme ray) of P.

Proposition 6.1

Let $P \neq \varnothing$ and P has a ray r, then P has extreme rays if and only if P has no lines.

Proof. Skipped. Here is a sketch:
If it has a line, then you can use the line to write any extreme ray in terms of vectors that are not positive scalars of itself.
If it has no lines, then A has rank n and you can use $n-1$ linearly independent constraints to "construct" an extreme ray. (Note that we know by hypothesis that the recession cone is non-zero) $\square$

Proposition 6.2

Let $\varnothing \neq P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be pointed. Let $r^1, \ldots, r^\ell$ be its extreme rays. Then

$$\max c^T x \qquad \text{s.t.} \qquad x \in P$$

is unbounded if and only if $\exists\, j : c^T r^j > 0$.

6.3 Minkowski-Weyl

Theorem 6.3: Minkowski-Weyl

Let $P \neq \varnothing$ be a pointed polyhedron, then

$$P = \left\{ x \in \mathbb{R}^n : \begin{array}{ccc} x & = \sum_{i=1}^k \lambda_i x^i & + \sum_{j=1}^\ell \mu_j r^j \\ & \sum_{i=1}^k \lambda_i & = 1 \\ & \lambda \geq 0, & \mu \geq 0 \end{array} \right\}$$

where $x^1, \dots, x^k$ are extreme points, and $r^1, \dots, r^\ell$ are extreme rays.

Proof. We denote the big {} as Q.
`Forward direction:`
Let $x \in Q$, we have

$$\begin{aligned} Ax &= \sum_{i=1}^k \lambda_i \underbrace{Ax^i}_{\leq b} + \sum_{j=1}^\ell \underbrace{\mu_j Ar^j}_{\leq 0} \\ &\leq \sum_{i=1}^k \lambda_i b + 0 = b \end{aligned}$$

which implies that $x \in P$, thus $Q \subseteq P$.
`Backward direction:` See below. □

Corollary 6.1

The reverse of the above theorem is also true.

Result 6.1

As a result, any polyhedron has two descriptions:

1. Intersection of finitely many inequalities (out description);
2. Convex combination of points and non-negative combination of rays (inner description).

Lecture 11 - Tuesday, October 22

Proof. We finish the proof for Minkowski-Weyl Theorem. Assume there exists $w \in P \setminus Q$. Thus we know that

$$\min 0 \quad \text{s.t.} \quad \begin{array}{cccc} \sum_{i=1}^k \lambda_i x^i & + \sum_{j=1}^\ell \mu_j r^j & = w & (\alpha) \\ \sum_{i=1}^k \lambda_i & & = 1 & (\alpha_0) \\ \lambda \geq 0, & \mu \geq 0 & & \end{array} \tag{QLP}$$

is infeasible. Writing out the dual we have:

$$\max w^T \alpha + \alpha_0 \quad \text{s.t.} \quad \begin{array}{cccc} \alpha^T x^i & + \alpha_0 & \leq 0 & \forall\, i = 1, \dots, k \\ \alpha^T r^j & & \leq 0 & \forall\, j = 1, \dots, \ell \end{array} \tag{QD}$$

which we know is unbounded. Hence there exist $\overline{\alpha}$ and $\overline{\alpha_0}$ such that

$$\overline{\alpha}^T w + \overline{\alpha_0} > 0$$

Now consider

$$\max \bar{\alpha}^T x \qquad \text{s.t.} \qquad x \in P \tag{*}$$

which we know is feasible. Now we have two cases,

1. $(*)$ **`has an optimal solution`**
   In this case, $(*)$ has an extreme point optimal solution. However,

   $$\overline{\alpha}^T w + \overline{\alpha_0} > 0 \geq \overline{\alpha}^T x^i + \overline{\alpha_0} \quad \Rightarrow \quad \overline{\alpha}^T w > \overline{\alpha}^T x^i \quad \forall\, i = 1, \ldots, k$$

   and there exists i such that $\overline{\alpha}^T x^i \geq \overline{\alpha}^T w$, which is a contradiction.

2. $(*)$ **`is unbounded`**
   Then there exists $r \in \text{rec}(P)$ such that

   $$\overline{\alpha}^T r > 0$$

   In fact, there exists extreme ray r^j such that

   $$\overline{\alpha}^T r^j > 0$$

   But this contradicts the fact that $\overline{\alpha}$ and $\overline{\alpha_0}$ are feasible for (QD).

Hence we complete the proof. □

Definition 6.6: Conv, Cone

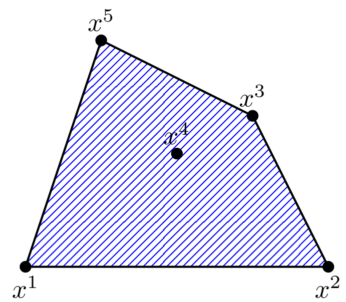
Let $x^1, \ldots, x^k \in \mathbb{R}^n$, then we define

$$\text{conv}\left(\{x^1, \ldots, x^k\}\right) := \left\{ x : \begin{array}{ll} x & = \sum_{i=1}^k \lambda_i x^i \\ 1 & = \sum_{i=1}^k \lambda_i \\ \lambda & \geq 0 \end{array} \right\}$$

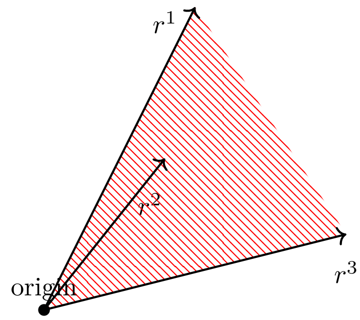
We define

$$\text{cone}\left(\{r^1, \ldots, r^k\}\right) := \left\{ r : \begin{array}{ll} r & = \sum_{i=1}^k \mu_i r^i \\ \mu & \geq 0 \end{array} \right\}$$

Example 6.3



Shaded region: $\text{conv}(\{x^1, x^2, x^3, x^5\})$



Shaded region: $\text{cone}(\{r^1, r^2, r^3\})$

6.3.1 Convex Hull

Discovery 6.2: Convex Hull

For all $S \subseteq \mathbb{R}^n$, $\text{conv}(S)$ is the smallest convex set containing S, and is called a **convex hull** . (This is the same as the previous definition for $|S|$ is finite). Moreover, $\text{cone}(S)$ is the smallest convex cone containing S, called the **cone generated by** $x^1, \ldots, x^k$. (Again, in the case of $|S|$ is finite, this coincides with the previous definition as well).

Definition 6.7: Minkowski Sum

Let $S, T \subseteq \mathbb{R}^n$, the **Minkowski Sum** of S and T is

$$S + T := \left\{ x : \begin{array}{ccc} x & = a + b & \\ & a \in S & b \in T \end{array} \right\}$$

Note 6.1

If $S = \varnothing = T$, then $S + T = \varnothing$.

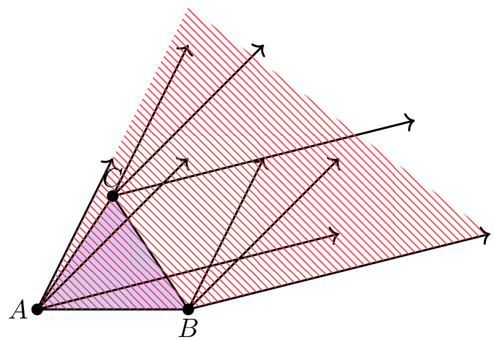
Result 6.2

Minkowski-Weyl says that if $P \neq \varnothing$ and pointed, then

$$P = \text{conv}(E) + \text{cone}(R)$$

where E is the set of extreme points, and R is the set of extreme rays.

Example 6.4



$\operatorname{conv}(E) + \operatorname{cone}(R)$

Corollary 6.2

Let $x^1, \ldots, x^k \in \mathbb{R}^n$ and $r^1, \ldots, r^\ell \in \mathbb{R}^n$, let

$$P := \operatorname{conv}(\{x^1, \ldots, x^k\}) + \operatorname{cone}(\{r^1, \ldots, r^\ell\})$$

Then P is a polyhedron, i.e., there exist A and b such that

$$P = \{x : Ax \leq b\}$$

Proof. We have

$$P = \operatorname{proj}_x \left\{ (x, \lambda, \mu) : \begin{array}{lll} x & = \sum_{i=1}^k \lambda_i x^i & + \sum_{j=1}^\ell \mu_j r^j \\ 1 & = \sum_{i=1}^k \lambda_i & \\ \lambda \geq 0, & \mu \geq 0 & \end{array} \right\}$$

as desired. $\square$

Lecture 12 - Thursday, October 24

Note 6.2

In class midterm today.

Lecture 13 - Tuesday, October 29

7 Simplex Algorithm

Definition 7.1: Standard Equality Form

An LP is said to be in **standard equality form** (SEF) if it is of the form

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Proposition 7.1

Given any LP, there is an equivalent LP in SEF.

Proof. Here we only give the idea of the proof.

- If LP is minimization of $c^T x$, it is equivalent to solving for max $-c^T x$ and multiply the result by -1.
- Suppose we have $a_i^T x \leq b_i$ as a contraint, we simply add a variable $s_i \geq 0$ and write $a_i^T x + s_i = b_i$.
- Suppose we have $a_i^T x \geq b_i$ as a contraint, we simply add a variable $s_i \geq 0$ and write $a_i^T x - s_i = b_i$.

Note 7.1

The first s_i is called the **slack variable**, while the second s_i is called a **surplus variable**.

- If we have $x_j \leq 0$, we let $x_j' = -x_j$ and replace x_j in the original LP by $-x_j'$.
- If x_k is free, we introduce x_k^+ and x_k^- such that $x_k^+, x_k^- \geq 0$ and let $x_k = x_k^+ - x_k^-$.

$\square$

Example 7.1: Free variable example

Suppose we have LP:

$$\begin{aligned} \max \quad & \begin{pmatrix} 1 & 2 & 3 \end{pmatrix} x \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 5 & 3 \\ 2 & -1 & 2 \end{pmatrix} x = \begin{pmatrix} 5 \\ 4 \end{pmatrix} \qquad x_1, x_2 \geq 0, \quad x_3 \text{ free} \end{aligned}$$

Solution: We set $x_3 := a - b$ where $a, b \geq 0$, and so the objective function has become

$$\begin{pmatrix} 1 & 2 & 3 & -3 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & a & b \end{pmatrix}^T$$

whereas the constraints has become

$$\begin{pmatrix} 1 & 5 & 3 & -3 \\ 2 & -1 & 2 & -2 \end{pmatrix} \begin{pmatrix} x_1 & x_2 & a & b \end{pmatrix}^T = \begin{pmatrix} 5 \\ 4 \end{pmatrix}$$

as desired. $\square$

Assume without loss of generality that $\text{rank}(A) = m$, because otherwise, the LP is either infeasible or there are redundant constraints.

Note 7.2

Note that the polyhedra in SEF have no lines, so if they are non-empty, they have basic feasible solution, which satisfies n linearly independent constraints out of $Ax = b$, $x \geq 0$ at equalities. We know that m of them come from $Ax = b$ (assuming $m \leq n$), hence $n - m$ of them come from $x \geq 0$.

Hence for a basic feasible solution, we will have $x_j = 0$ for $j \in N \subseteq \{1, \ldots, n\}$ with $|N| = n - m$ and $Ax = b$.

Definition 7.2:

For a matrix M with n columns, and $J \subseteq \{1, \ldots, n\}$. Matrix M_J is the matrix obtained by picking columns in J.

Example 7.2

For $M = \begin{pmatrix} 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 3 \end{pmatrix}$ and $J = \{1, 3\}$, we have $M_J = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$.

As a result, if we let $B = \{1, \ldots, n\} \backslash N$, then

$$Ax = b \quad \Longleftrightarrow \quad A_B x_B + A_N x_N = b$$

If $x_N = 0$, then we have RHS as $A_B x_B = b$.

Definition 7.3: Basis

$B \subseteq \{1, \ldots, n\}$ is a **basis** if A_B is a full rank matrix.

Definition 7.4: Basic Variable

Given a basis B, x_B are **basic variables**, and x_N are **nonbasic variables**.

Given a basis B, associated basic solution is

$$x_N = 0 \qquad \text{and} \qquad x_B = A_B^{-1} b$$

If $x_B = A_B^{-1} b \geq 0$, then it is a basic feasible solution. In this case, we say B is a **feasible basis**.

7.1 Canonical Form

Algorithm 7.1

Start with a feasible basis B, while there exists a better basis B'.
Suppose I have a starting feasible basis B, rewrite

$$\begin{array}{ll}\max & c^T x \\ \text{s.t.} & \begin{array}{r}Ax = b\\ x \geq 0\end{array}\end{array} \iff \begin{array}{ll}\max & c_B^T x_B + c_N^T x_N \\ \text{s.t.} & \begin{array}{r}A_B x_B + A_N x_N = b\\ x_B, x_N \geq 0\end{array}\end{array} \iff \begin{array}{ll}\max & c_B^T x_B + c_N^T x_N + 0 \\ \text{s.t.} & \begin{array}{r}A_B^{-1} A_B x_B + A_B^{-1} A_N x_N = A_B^{-1} b\\ x_B, x_N \geq 0\end{array}\end{array}$$

Note that

$$\begin{aligned} & x_B + A_B^{-1} A_N x_N - A_B^{-1} b = 0 \in \mathbb{R}^m \\ \Rightarrow \quad & -c_B^T \left[x_B + A_B^{-1} A_N x_N - A_B^{-1} b\right] = 0 \in \mathbb{R} \end{aligned}$$

Hence we may rewrite the objective function as

$$\begin{aligned} & \max c_B^T x_B + c_N^T x_N - c_B^T \left[x_B + A_B^{-1} A_N x_N - A_B^{-1} b\right] \\ \Rightarrow \quad & \max c_N^T x_N - c_B^T A_B^{-1} A_N x_N + c_B^T A_B^{-1} b \end{aligned}$$

Now the LP becomes

$$\begin{array}{ll}\max & \left(c_N^T - c_B^T A_B^{-1} A_N\right) x_N + c_B^T A_B^{-1} b \\ \text{s.t.} & \begin{array}{r}x_B + A_B^{-1} A_N x_N = A_B^{-1} b\\ x_B, x_N \geq 0\end{array}\end{array}$$

which is called the **canonical form** for basis B.

Discovery 7.1

Let

$$\overline{c_N}^T := c_N^T - c_B^T A_B^{-1} A_N$$

If $\overline{c_N} \leq 0$, then basis B is optimal, else there exists $j \in N$ such that $\overline{c_j} > 0$.

Example 7.3

Suppose we have

$$\max \begin{pmatrix}0 & 1 & 0 & -1\end{pmatrix} x + 240$$

and $B = \{1, 3\}$, $N = \{2, 4\}$. Then we have

$$\overline{c_N} = \begin{pmatrix}1 \\ -1\end{pmatrix} \qquad \text{and} \qquad \overline{c_4} = -1,\ \overline{c_2} = 1$$

7.2 Result of Equivalent LP's

We have shown that the following four LP's are equivalent:

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{1}$$

$$\begin{aligned} \max \quad & c_B^T x_B + c_N^T x_N \\ \text{s.t.} \quad & A_B x_B + A_N x_N = b \\ & x_B, x_N \geq 0 \end{aligned} \tag{2}$$

$$\begin{aligned} \max \quad & \left(c^T - c_B^T A_B^{-1} A\right) x + c_B^T A_B^{-1} b \\ \text{s.t.} \quad & A_B^{-1} A x = A_B^{-1} b \\ & x \geq 0 \end{aligned} \tag{3}$$

$$\begin{aligned} \max \quad & c_N^T x_N + c_B^T A_B^{-1} b \\ \text{s.t.} \quad & x_B + A_B^{-1} A_N x_N = b \\ & x_B, x_N \geq 0 \end{aligned} \tag{4}$$

7.2.1 Example of Simplex (details missing)

Example 7.4

Suppose we have the LP as

$$\begin{aligned} \max \quad & \begin{pmatrix} 2 & 1 & 1 & 0 & 0 \end{pmatrix} x \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 2 & -1 & 1 & 0 \\ 2 & -2 & -1 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{aligned} \tag{1}$$

Let $B = \{1, 4\}$, then the LP becomes

$$\begin{aligned} \max \quad & \begin{pmatrix} 2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} + \begin{pmatrix} 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 1 \\ 2 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} + \begin{pmatrix} 2 & -1 & 0 \\ -2 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ & \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} \geq 0, \qquad \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} \geq 0 \end{aligned} \tag{2}$$

and

$$\begin{aligned} \max \quad & \begin{pmatrix} 0 & 3 & 2 & 0 & -1 \end{pmatrix} x \\ \text{s.t.} \quad & \begin{pmatrix} 1 & -1 & -0.5 & 0 & 0.5 \\ 0 & 3 & -0.5 & 1 & -0.5 \end{pmatrix} x = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} \\ & x \geq 0 \end{aligned} \tag{3}$$

and

$$
\begin{aligned}
\max \quad & \begin{pmatrix} 3 & 2 & -1 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} + 3 \\
\text{s.t.} \quad & \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} + \begin{pmatrix} -1 & -0.5 & 0.5 \\ 3 & -0.5 & -0.5 \end{pmatrix} \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} \\
& \begin{pmatrix} x_1 \\ x_4 \end{pmatrix} \geq 0, \qquad \begin{pmatrix} x_2 \\ x_3 \\ x_5 \end{pmatrix} \geq 0
\end{aligned} \tag{4}
$$

Note 7.3

If $\bar{c}^T = \left(c^T - c_B^T A_B^{-1} A\right)$, is vector of **reduced costs**.

Lecture 14 - Thursday, October 31

Result 7.1

If $\bar{c} \leq 0$, then B is an optimal basis, corresponding BFS is an optimal solution. Else there exists j such that $\overleftarrow{c}_j > 0$.

Example 7.5

Continue the above example. In (3), suppose we want to increase x_2 from 0 to $\varepsilon \geq 0$ (we keep equality so that ε could take 0), while keeping $x_3 = x_5 = 0$. Thus we have

$$
\begin{pmatrix} x_1 \\ x_4 \end{pmatrix} + \begin{pmatrix} -1 \\ 3 \end{pmatrix} \varepsilon = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} \quad \Longleftrightarrow \quad \begin{aligned} x_1 &= 1.5 + \varepsilon & \geq 0 \\ x_4 &= 0.5 - 3\varepsilon & \geq 0 \end{aligned} \quad \Rightarrow \quad \varepsilon \in \left[-\frac{3}{2}, \frac{1}{6}\right]
$$

Take $\varepsilon = 1/6$, we have $x_1, x_2 \neq 0$ and $x_3 = x_4 = x_5 = 0$, which implies that we arrive at a new basis $B = \{1, 2\}$ and $N = \{3, 4, 5\}$.

7.3 Simplex Algorithm (basic idea)

Algorithm 7.2

1. Start with a feasible basis B;
2. Write LP in canonical form for B;
3. If $\bar{c} \leq 0$, we halt and conclude that B is optimal basis. Otherwise, we procees to next step;
4. Pick $j \in \mathbb{N}$ such that $\bar{c}_j > 0$ entering the basis. Figure out which variable leaves the basis;
5. The new basis B' becomes $(B \backslash \{k\}) \cup \{j\}$ and go to step 2.

7.3.1 Variable Leaving Basis

- Call $\overline{A} = A_B^{-1} A$ and $\bar{b} = A_B^{-1} b$.
- For every $i = 1, \ldots, m$, let $B(i)$ be the corresponding basic variable.
- Increasing x_j to ε will mean that $x_{B(i)}$ will
  - increase by $|\overline{a_{ij}}|\varepsilon$ if $\overline{a_{ij}} < 0$, or
  - decrease by $|\overline{a_{ij}}|\varepsilon$ if $\overline{a_{ij}} > 0$
- Compute $\min_{i:\overline{a_{ij}}>0} \dfrac{\overline{b_i}}{\overline{a_{ij}}}$

Note 7.4

Currently $x_{B(i)} = \overline{b_i}$, when we increase x_j to ε, we have

$$x_{B(i)} = \overline{b_i} - \overline{a_{ij}}\varepsilon \geq 0 \quad \Rightarrow \quad \varepsilon \leq \frac{\overline{b_i}}{\overline{a_{ij}}} \qquad \text{for } \overline{a_{ij}} > 0$$

- Let ℓ be such that $\dfrac{\overline{b_\ell}}{\overline{a_{\ell j}}} = \min_{i:\overline{a_{ij}}>0} \dfrac{\overline{b_i}}{\overline{a_{ij}}}$, we then have $k = B(l)$.

7.3.2 Questions arising (Bland's Rule)

Note 7.5

Does this converge? In other words, do we halt after a finite number of steps.
If at current BFS we have a basic variable $= 0$, we may have $\varepsilon = 0 \to$ May lead to cycling. (i.e. return to current basis in future iteration)

Proof. If problem has optimal solution AND ε is always > 0, simplex finishes as we always get stricktly better solution and ther eare only finitely many bases.
Here we introduce Bland's Rule: if there are multiple choices of entering or leaving variables, always pick lowest index variable.
Using Bland's Rule avoids cycling. □

Note 7.6

What if $\overline{a}_{ij} \leq 0$ for all $i = 1, \ldots, m$?

Proof. Then the LP is unbounded. □

Discovery 7.2

Here is a quick discovery: For the LP:

$$\begin{aligned} \max \quad & c_B^T x_B + c_N^T x_N \\ \text{s.t.} \quad & A_B x_B + A_N x_N = b \\ & x_B, x_N \geq 0 \end{aligned}$$

we consider its dual:

$$\begin{aligned} \min \quad & b^T y \\ \text{s.t.} \quad & y^T A_B \geq c_B^T \\ & y^T A_N \geq c_N^T \end{aligned}$$

For x is a basic feasible solution, we know that $x_N = 0$ and $x_B = A_B^{-1} b$. To guarantee that, we need to satisfy Complementary Slackness, namely we need to have

$$y^T A_B = c_B^T \quad \Rightarrow \quad y^T = c_B^T A_B^{-1}$$

The dual feasibility further yields us

$$y^T A_N \geq c_N^T \quad \iff \quad c_B^T A_B^{-1} A_N \geq c_N^T \quad \iff \quad c_N^T - c_B^T A_B^{-1} A_N \leq 0$$

which is exactly the point we reach an optimal solution in the algorithm.

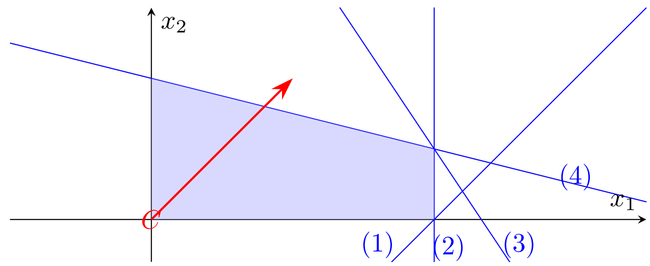
Lecture 15 - Tuesday, November 05

7.4 Geometric of the Simplex Algorithm

Here we use □ to denote a variable entering the basis, and ○ to denote a variable leaving the basis.

Consider the following LP:

$$
\begin{array}{lll}
\max & x_1 + x_2 & \\
\text{s.t.} & x_1 - x_2 \le 2 & (1) \\
& x_1 \le 2 & (2) \\
& 1.5x_1 + x_2 \le 3.5 & (3) \\
& 0.25x_1 + x_2 \le 1 & (4) \\
& x_1, x_2 \ge 0 &
\end{array}
$$



Rewriting in SEF. In canonical form for $B = \{3, 4, 5, 6\}$

$$
\begin{array}{ll}
\max & (\boxed{1}, 1, 0, 0, 0, 0)x \\
\text{s.t.} & \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1.5 & 1 & 0 & 0 & 1 & 0 \\ 0.25 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 2 \\ 3.5 \\ 1 \end{pmatrix} \\
& x \ge 0
\end{array}
$$

Now the basis become *Basis* : $\{1, 4, 5, 6\}$, and

$$
\begin{array}{ll}
\max & (0, \boxed{2}, -1, 0, 0, 0)x + 2 \\
\text{s.t.} & \begin{pmatrix} 1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 2.5 & -1.5 & 0 & 1 & 0 \\ 0 & 1.25 & -0.25 & 0 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 0 \\ 0.5 \\ 0.5 \end{pmatrix} \\
& x \ge 0
\end{array}
$$

Now the basis become *Basis* : $\{1, 2, 5, 6\}$, and

$$
\begin{array}{ll}
\max & (0, 0, \boxed{1}, -2, 0, 0)x + 2 \\
\text{s.t.} & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2.5 & 1 & 0 \\ 0 & 0 & 1 & -1.25 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 0 \\ 0.5 \\ 0.5 \end{pmatrix} \\
& x \ge 0
\end{array}
$$

Now the basis become *Basis* : $\{1, 2, 3, 6\}$, and

$$
\begin{array}{ll}
\max & (0, 0, 0, \boxed{0.5}, -1, 0)x + 2.5 \\
\text{s.t.} & \begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1.5 & 1 & 0 \\ 0 & 0 & 1 & -2.5 & 1 & 0 \\ 0 & 0 & 0 & 1.25 & -1 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 0.5 \\ 0.5 \\ 0 \end{pmatrix} \\
& x \ge 0
\end{array}
$$

Lastly, the basis become $Basis : \{1, 2, 3, 4\}$, and

$$\begin{array}{ll} \max & (0,0,0,0,-0.6,-0.4)x + 2.5 \\ \text{s.t.} & \begin{pmatrix} 1 & 0 & 0 & 0 & 0.8 & -0.8 \\ 0 & 1 & 0 & 0 & -0.2 & 1.2 \\ 0 & 0 & 1 & 0 & -1 & 2 \\ 0 & 0 & 0 & 1 & -0.8 & 0.8 \end{pmatrix} x = \begin{pmatrix} 2 \\ 0.5 \\ 0.5 \\ 0 \end{pmatrix} \\ & x \geq 0 \end{array}$$

which implies that our optimal solution is $\bar{x} = \begin{pmatrix} 2 \\ 0.5 \\ 0.5 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ and the optimal value is 2.5.

7.5 Mechanics

$$\begin{array}{ll} \max & (2,1,1,0,0)x \\ \text{s.t.} & \begin{pmatrix} 1 & 2 & -1 & 1 & 0 \\ \boxed{2} & -2 & -1 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{array}$$

Our current basis is $B = \{4, 5\}$. Notice that the smallest index j such that $c_j > 0$ is $j = 1$, thus we have x_1 entering the basis. Moreover, calculating $2/1$ and $3/2$ (b_i/a_{ki} for all k) we find $3/2 < 1/1$, thus we have x_5 leaving the basis. Now the new basis becomes $B = \{1, 4\}$. We wish to make the boxed element to be 1 and all other elements in the same column to be 0. By multiplying the boxed row by $1/2$ and subtract it from first row as well as the objective function, we obtain

$$\begin{array}{lll} & \max & (2,1,1,0,0)x + (-1)\overbrace{\left[(2,-2,-1,0,1)x - 3\right]}^{=0} \\ & \text{s.t.} & \begin{pmatrix} 0 & 3 & -0.5 & 1 & -0.5 \\ 1 & -1 & -0.5 & 0 & 0.5 \end{pmatrix} x = \begin{pmatrix} 0.5 \\ 1.5 \end{pmatrix} \\ & & x \geq 0 \\ \Longrightarrow & \max & (0,3,2,0,-1)x + 3 \\ & \text{s.t.} & \begin{pmatrix} 1 & -1 & -0.5 & 0 & 0.5 \\ 0 & \boxed{3} & -0.5 & 1 & -0.5 \end{pmatrix} x = \begin{pmatrix} 1.5 \\ 0.5 \end{pmatrix} \\ & & x \geq 0 \end{array}$$

For our second iteration, we have x_2 leaving the basis, and x_4 entering the basis. Thus the new basis becomes

$B = \{1, 2\}$. We obtain:

$$\begin{aligned} \max \quad & (0, 0, 2.5, -1, -0.5)x + 3.5 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & -2/3 & 1/3 & 1/3 \\ 0 & 1 & -1/6 & 1/3 & -1/6 \end{pmatrix} x = \begin{pmatrix} 5/3 \\ 1/6 \end{pmatrix} \\ & x \geq 0 \end{aligned}$$

Discovery 7.3

In Simplex Algorithm, we move from a BFS along a direction d defined as

$$\begin{pmatrix} d_B \\ d_N \end{pmatrix} = \begin{pmatrix} -\overline{A_j} \\ e_j \end{pmatrix}$$

where $j \in N$ enters the basis.

Note 7.7

In the example above, for $B = \{4, 5\}$, $\begin{pmatrix} d_4 \\ d_5 \\ d_1 \\ d_2 \\ d_3 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. Moving from $\overline{x}$ to $\overline{x} + \theta d$.

Also note that

$$Ad = A_B d_B + A_N d_N = A_B(-A_B^{-1} A_j) + A_j = 0$$

which is essential since we are moving in a direction that only untights one of the constraints. Now if $\overline{A_j} \leq 0$, then $d \geq 0$. And

$$\overline{c}^T d = \underbrace{\overline{c_B}^T d_B}_{=0} + \overline{c_N} d_N = \overline{c_j} > 0$$

which implies that the LP is bounded.

Lecture 16 - Thursday, November 07

7.6 Dealing with Feasibility — 2-phase Simplex Method

We wish to solve

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned} \tag{SEF}$$

Our goal is to find a feasible basis B, or to show that none exists.

7.6.1 Phase 1

Algorithm 7.3

- Assume WLOG that $b \geq 0$;
- Consider

$$\begin{aligned} \max \quad & -e^T z \\ \text{s.t.} \quad & Ax + Iz = b \\ & x, z \geq 0 \end{aligned} \tag{AUX}$$

  Notice that the objective function is equivalent to min $\sum_i z_i$. Here we have several results:

  – (AUX) has a BFS, and z is the basic variable, x is the non-basic variable;

  – (AUX) has an optimal solution, since it is bounded below by zero in regards to the equivalent form.

  – If the optimal value for (AUX) is 0, then $z_i = 0$ for all i. This case yields us a feasible basis B for (SEF) with only x variables in it.

  – If the optimal value for (AUX) < 0, then (SEF) is infeasible.

7.6.2 Phase 2

Algorithm 7.4

If Phase 1 ended with a feasible basis B for (SEF), we run simplex on (SEF) with B as our starting basis.

7.7 Brief Complexity of Algorithms (Handwavy)

Definition 7.5: Size

Given a problem instance I, the size(I) is the number of bits needed to represent it.

Example 7.6

For LPs:

$$\approx (m \times n + n + m) \begin{bmatrix} \text{max bit size} \\ \text{of } A, b, c \end{bmatrix}$$

if A, b, and c are integers.

Definition 7.6: Efficient

We say an algorithm is **efficient** if it runs in $\mathcal{O}(poly(size(I)))$ for all instances I. i.e., a polynomial time algorithm.

Result 7.2

Simplex method is NOT a polynomial time algorithm.

Theorem 7.1: Spielman & Tang (2001)

“Bad” instances are pathological, (“happens rarely”).

7.7.1 Hirsch Conjecture

Suppose we have a polytope (a bounded polyhedron):

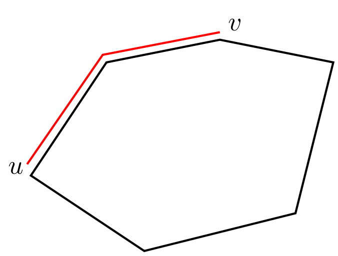
$$P = \{x \in \mathbb{R}^n : Ax \leq b\}$$

where A is $m \times n$. Let $u, v \in P$ and define $\operatorname{dist}(u, v)$ to be the shortest path from u to v going through edges of P.

Definition 7.7: Edge

Line segment between adjacent vertices.

Example 7.7



$\operatorname{dist}(u, v) = 2$

Definition 7.8: Diameter

We define the **diameter** of P, $\operatorname{diam}(P)$, to be

$$\operatorname{diam}(P) = \max_{u,v \text{ vertices } \in P} \{\operatorname{dist}(u, v)\}$$

Discovery 7.4

Then the number of simples iterations is greater or equal to $\operatorname{diam}(P)$ (worst case).

The Hirsch conjecture states that

$$\operatorname{diam}(P) \leq m - n$$

Theorem 7.2: Santos (2010)

The Hirsch conjecture is false.

7.7.2 Polynomial Hirsch Conjecture

The conjecture states that

$$\operatorname{diam}(P) \leq poly(n, m)$$

Discovery 7.5

Notice that even if this is true, the Simplex method still has issues:

→ degenerate iterations;

→ choices of entering/ leaving variables determines the edge, but we do not know the rule that allows us to guarantee $\leq \operatorname{diam}(P)$ iterations.

7.8 Polynomial Equivalence of Separation, Optimization, Feasibility

Definition 7.9: Separation Problem

Given a bounded polyhedron

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

and $\bar{x} \in \mathbb{R}^n$. The **separation problem** is to determine if $\bar{x} \in P$ or find α, α_0 such that

$$\alpha^T x \leq \alpha_0 \quad \forall\, x \in P \qquad \text{and} \qquad \alpha^T \bar{x} > \alpha_0$$

Definition 7.10: Feasibility Problem

Given a bounded polyhedron

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

Determine if $P = \varnothing$ or find $\bar{x} \in P$.

Definition 7.11: Optimization Problem

Given a bounded polyhedron

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

$c \in \mathbb{R}^n$, find if $P = \varnothing$ or find $\bar{x} \in P$ such that

$$c^T \bar{x} \geq c^T x \qquad \forall\, x \in P$$

7.8.1 Grötchel-Lovász-Schrijver

Theorem 7.3: Grötchel-Lovász-Schrijver

We have TFAE:

1. Separation problem can be solved in polynomial time
2. Feasibility problem can be solved in polynomial time
3. Optimization problem can be solved in polynomial time

Proof. `Opt` $\Rightarrow$ `Feas`: this is simple.
`Feas` $\Rightarrow$ `Opt`: the idea is binary search. We ask the question: Is there $x \in P$ such that $c^T x \geq \gamma$, and we do binary search on γ.
`Opt` $\Rightarrow$ `Sep`: Idea in HW2Q4
`Sep` $\Rightarrow$ `Feas`: GLS uses Ellipsoid Algorithm that uses separation to solve for feasibility.

Note 7.8

Ellipsoid-Karmarkar — first polynomial time algorithm for LP.

□

Result 7.3

Overall, there is no known strongly polytime algorithm for LPs (2024 Nov). i.e., $poly(n, m)$.

8 Integer Programming (IP)

Now we would encounter problems like

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \in \mathbb{Z}, \ \forall \, j \in I \end{aligned} \tag{IP}$$

where $c \in \mathbb{Q}^n$, $A \in \mathbb{Q}^{m \times n}$, $b \in \mathbb{Q}^m$ and $I \subseteq \{1, \ldots, n\}$ with $I \neq \varnothing$.

Definition 8.1: Pure IP

If $I = \{1, \ldots, n\}$, then the IP is called a **pure IP**.

Definition 8.2: Mixed IP

If $I \neq \{1, \ldots, n\}$, then the IP is called a **mixed IP**.

Definition 8.3: Binary IP

If all variables are required to be in $\{0, 1\}$, then the IP is called a **binary IP**.

Definition 8.4: Integral

We say $\bar{x} \in \mathbb{R}^n$ is **integral** if $\bar{x}_j \in \mathbb{Z}$ for all $j \in I$.

Definition 8.5: Rational Polyhedron

$P = \{x \in \mathbb{R}^n : Ax \leq b\}$ is a **rational polyhedron** if $A, b \in \mathbb{Q}$ as above.

Note 8.1

We denote

$$P_I = \{x \in P : x \text{ is integral}\}$$

8.1 If P is a rational polyhedron, then $\operatorname{conv}(P_I)$ is a rational polyhedron. If $P_I \neq \varnothing$, then $\operatorname{rec}(\operatorname{conv}(P_I)) = \operatorname{rec}(P)$

Theorem 8.1

If P is a rational polyhedron, then $\operatorname{conv}(P_I)$ is a rational polyhedron. If $P_I \neq \varnothing$, then

$$\operatorname{rec}(\operatorname{conv}(P_I)) = \operatorname{rec}(P)$$

Proof. Assume $P \neq \varnothing$ and pointed, also assume $I = \{1, \dots, n\}$ to simplify proof.
If $P_I = \varnothing$, then we may take

$$\operatorname{conv}(P_I) = \{x \in \mathbb{R}^n : 1 \leq x_1 \leq -1\}$$

which simply constructs an empty polyhedron arbitrarily. Therefore we may assume that $P_I \neq \varnothing$, thus by Minkowski-Weyl (6.3), we have

$$P = \left\{ x \in \mathbb{R}^n : \begin{array}{rll} x & = \sum_{i=1}^{p} \lambda_i x^i & + \sum_{j=1}^{q} \mu_j r^j \\ \sum_{i=1}^{p} \lambda_i & & = 1 \\ \lambda \geq 0, & & \mu \geq 0 \end{array} \right\}, \qquad \begin{array}{l} x^1, \dots, x^p \text{ extreme points} \\ r^1, \dots, r^q \text{ extreme rays} \end{array}$$

Note 8.2

Extreme points are rational. We may assume that extreme rays are rational too. Rather, we may even assume that extreme rays are integral, since we can always scale them.

Now we define a bounded polyhedron:

$$T = \left\{ x \in \mathbb{R}^n : \begin{array}{rll} x & = \sum_{i=1}^{p} \lambda_i x^i & + \sum_{j=1}^{q} \mu_j r^j \\ \sum_{i=1}^{p} \lambda_i & & = 1 \\ \lambda \geq 0, & & 0 \leq \mu_j \leq 1, \quad \forall\, j = 1, \dots, q \end{array} \right\}$$

Thus we have

$$T_I = \{x \in T : x \in \mathbb{Z}^n\} \qquad \text{and} \qquad R_I = \left\{ x = \sum_{j=1}^{q} \mu_j r^j, \quad \mu_j \in \mathbb{Z}^q_{\geq 0} \right\}$$

`Claim 1:` $P_I = T_I + R_I$ `i.e.,` $P_I = T_I \sqcup R_I$
`[`$\supseteq$`]`: Trivial.
`[`$\subseteq$`]`: we have

$$\begin{aligned} x \in P_I \quad & \Rightarrow \quad x = \sum \lambda_i x^i + \sum \mu_j r^j, \quad \text{since } P_I \subseteq P \\ & \Rightarrow \quad \underbrace{\sum \lambda_i x^i + \sum (\mu_j - \lfloor \mu_j \rfloor)\, r^j}_{\in T} + \underbrace{\sum \lfloor \mu_j \rfloor r^j}_{\in R_I} \\ & \Rightarrow \quad P_I \ni x = t + r \in T + R_I \\ & \Rightarrow \quad t = x - r \in \mathbb{Z}^n \\ & \Rightarrow \quad t \in T_I \end{aligned}$$

as desired.

`Claim 2:` $\text{conv}(\underbrace{T_I + R_I}_{P_I}) = \text{conv}(T_I) + \text{conv}(R_I)$.

`Claim 3:` $\text{conv}(R_I) = \text{rec}(P)$.

The proof for claim 2 and 3 are skipped, they are not trivial. $\square$

8.2 $\max\ c^T x$ s.t. $x \in P_I \equiv \max\ c^T x$ s.t. $x \in \text{conv}(P_I)$

Theorem 8.2

We have

$$\underbrace{\left\{\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in P_I \end{array}\right\}}_{\boxed{z_1}} = \underbrace{\left\{\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in \text{conv}(P_I) \end{array}\right\}}_{\boxed{z_2}}$$

Proof. Since $P_I \subseteq \text{conv}(P_I)$, we have $z_1 \leq z_2$.
Note that if $P_I = \varnothing$, then $\text{conv}(P_I) = \varnothing$, which implies that $z_1 = z_1 = -\infty$. Thus we may assume that $P_I \neq \varnothing$. Let $x^* \in \text{conv}(P_I)$, this tells us that

$$x^* = \sum_{i=1}^{p} \lambda_i x^i, \quad \sum_i \lambda_i = 1,\ \lambda \geq 0, x^1, \ldots, x^p \in P_I$$

This implies that there exists i such that
$$c^T x^i \geq c^T x^*$$

because otherwise we would have

$$c^T x^* = \sum \lambda_i c^T x^i < \sum \lambda_i c^T x^* = c^T x^*$$

Therefore, we must have $z_1 = z_2$ as desired. $\square$

Corollary 8.1

Suppose P is pointed, then $\text{conv}(P_I)$ is pointed, and any extreme point of $\text{conv}(P_I)$ is integral.

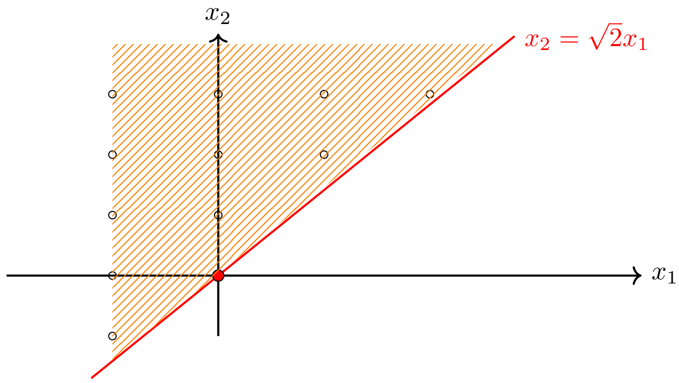
Proof. (*Pure integer case for simplicity*)
We have
$$\text{rec}(P) \supseteq \text{rec}(\text{conv}(P_I))$$

which implies that $\text{conv}(P_I)$ is pointed. Let x^* be an extreme point of $\text{conv}(P_I)$, so there exists c such that x^* is a unique solution to
$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in \text{conv}(P_I) \end{array}$$

However, the theorem tells us that there exists $\bar{x} \in P_I$ such that $c^T \bar{x} = c^T x^*$. But $\bar{x} \in \text{conv}(P_I)$ implies that $x^* = \bar{x} \in P_I$, which further implies that x^* is integral. $\square$

Example 8.1

Suppose we have $P = \{x \in \mathbb{R}^2 : x_2 \geq \sqrt{2}x_1\}$, thus $P_I = P \cap \mathbb{Z}^2$:



Note 8.3

Notice that $\text{conv}(P_I)$ is not a polyhedron since it is not a closed set, so our theorem wouldn't apply in this case.

Lecture 18 - Thursday, November 14

Consequences:

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \end{aligned} \tag{LP}$$

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \in \mathbb{Z},\ \forall j \in I \end{aligned} \tag{IP}$$

Then (IP) is either infeasible/ unbounded/ or has an optimal solution.

Result 8.1

- If (LP) is infeasible, then (IP) is infeasible.
- If (LP) is unbounded,
  - If $P_I \neq \varnothing$, then (IP) is unbounded;
  - If $P_I = \varnothing$, then (IP) is infeasible.
- If (LP) has an optimal solution,
  - If $P_I \neq \varnothing$, then (IP) has an optimal solution;
  - If $P_I = \varnothing$, then (IP) is infeasible.

Note 8.4

Separation, Optimization, and Feasibility for conv(P_I) are all "polynomially equivalent".

Note 8.5

Input size of (IP) is approximately the input size of (LP).

8.3 Cutting Plane Algorithm

It is hard to compute conv(P_I), so we try to produce sequence of polyhedra that are "closer" to conv(P_I).

Definition 8.6:

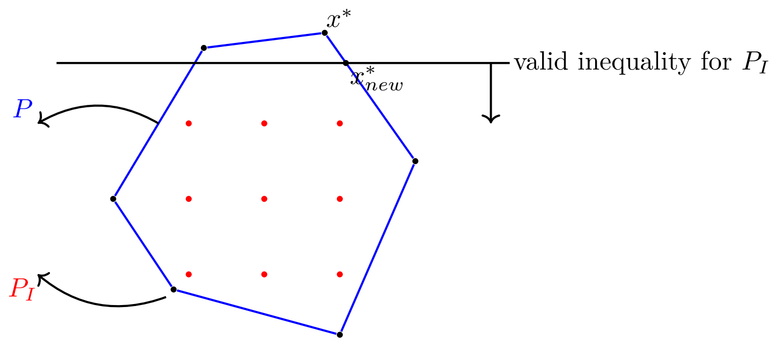
We call (LP) the LP-relaxation of (IP).

Definition 8.7: Valid

We say an inequality $\alpha^T x \leq \alpha_0$ is **valid** for $S \subseteq \mathbb{R}^n$ if $\alpha^T \bar{x} \leq \alpha_0$ for all $\bar{x} \in S$.

Example 8.2

Geometric interpretation:



It is reasonable to assume that (LP) has an optimal solution.

Algorithm 8.1

Let $R \leftarrow P = \{x : Ax \leq b\}$.
Do, while R neq $\varnothing$

```
| Let x* be opt. sol. to
|                    max c^T x
|                    s.t. x ∈ R
| If x* is integral, stop.
|    x* is opt. sol. for (IP)
| Else, find α^T x ≤ α_0 valid for P_I such that α^T x* > α_0,
```

| $R \leftarrow R \cap \{x : \alpha^T x \le \alpha_0\}$

Return $P_I = \varnothing$.

Note 8.6

At all times,

$$P_I \subseteq R \subseteq P$$

and we have

$$\underbrace{\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in P_I \end{array}}_{z_I} \le \underbrace{\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in R \end{array}}_{z_R}$$

If x^* is integral, then

$$x^* \in P_I \quad \Rightarrow \quad z_R = c^T x^* \le z_I \le z_R$$

which implies that x^* is optimal solutions for P_I.

Proposition 8.1

Let R be a pointed polyhedron such that $P_I \subseteq R \subseteq P$. Let x^* be a BFS of

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in R \end{array}$$

Then x^* is integral if and only if $x^* \in \text{conv}(P_I)$.

Proof. **exercise**. □

This suggests that cutting plane algorithm is well-defined if x^* is a BFS.

8.4 Chvátal-Gomory (CG) Cuts

- Assume $I = \{1, \ldots, n\}$ (this is crucial);
- Assume $P = \{x : Ax = b,\ x \ge 0\}$ (for convenience).

Now, x^* is a BFS, i.e.

$$x_N^* = 0 \qquad \text{and} \qquad x_B^* = \underbrace{A_B^{-1} b}_{\bar{b}}$$

and in canonical form for B:

$$x_B + \underbrace{A_B^{-1} A_N}_{\overline{A_N}} x_N = A_B^{-1} b := \bar{b}$$

$x_B, x_N \ge 0$. Since x^* is not integral, there exists i such that $\overline{b_i} \notin \mathbb{Z}$. Consider

$$\left\{ x \in \mathbb{R}^n : \begin{array}{rr} x_{B(i)} + \sum_{j \in N} \bar{a}_{ij} x_j & = \bar{b}_i \\ x_{B(i)} \ge 0, & x_N \ge 0 \end{array} \right\} =: K$$

Note 8.7

If $\alpha^T x \leq \alpha_0$ is valid for $K \cap \mathbb{Z}^n$, then it is valid for $P \cap \mathbb{Z}^n = P_I$, since

$$P \cap \mathbb{Z}^n \subseteq K \cap \mathbb{Z}^n$$

Note 8.8

If $\alpha^T x \leq \alpha_0$ is valid for $K \cap \mathbb{Z}^n$, then $\alpha_j = 0$ for all $j \in B \backslash \{B(i)\}$.

Now we have

$$x_{B(i)} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor x_j \leq \bar{b}_i$$

is valid for K, since $x_N \geq 0$. We know that $x_{B(i)}$ is an integer because we care about $K \cap \mathbb{Z}^n$, and $\lfloor \bar{a}_{ij} \rfloor$, x_j are all integers. Therefore we have

$$x_{B(i)} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor x_j \leq \lfloor \bar{b}_i \rfloor \tag{*}$$

is valid for $K \cap \mathbb{Z}^n$, hence is valid for P_I.

Note 8.9

For $x = x^*$,

$$x^*_{B(i)} = \bar{b}_i \qquad \text{and} \qquad x^*_N = 0$$

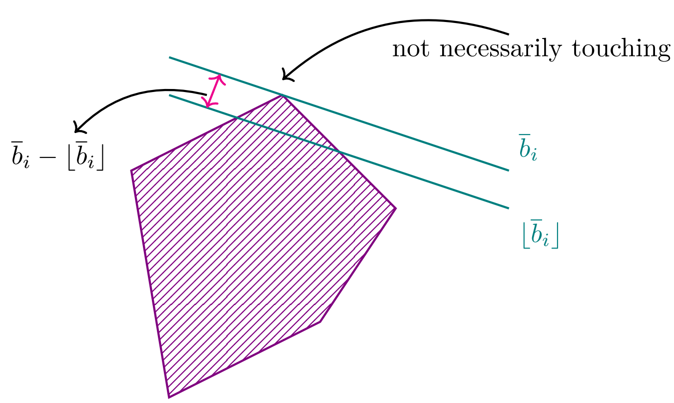
which implies

$$x^*_{B(i)} + \sum_{j \in N} \lfloor \bar{a}_{ij} \rfloor x^*_j = x^*_{B(i)} = \bar{b}_i > \lfloor \bar{b}_i \rfloor$$

i.e., we found $\alpha^T x \leq \alpha_0$ valid for P_I and with $\alpha^T x^* > \alpha_0$.

Discovery 8.1

Geometrically, we have



Result 8.2

(*) is called a CG-cut.

Note 8.10

A cut is just a valid inequality violated by x^*.

Theorem 8.3

CG-cutting plane algorithm terminates after finitely many iterations.

Lecture 19 - Tuesday, November 19

8.5 Branch and Bound

8.5.1 Branch

Consider the IP:

$$\begin{aligned} z_{IP} = \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \in \mathbb{Z}, \quad \forall\, j \in I \end{aligned} \tag{IP}$$

Note 8.11

We assume that $I = \{1, \dots, n\}$. Also assume IP relaxation has an optimal solution.

Discovery 8.2

Consider $a \in \mathbb{Z}$, $j \in I$ and two integer programs:

$$\begin{aligned} z_1 = \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \leq a \\ & x \in \mathbb{Z}^n \end{aligned} \tag{IP1}$$

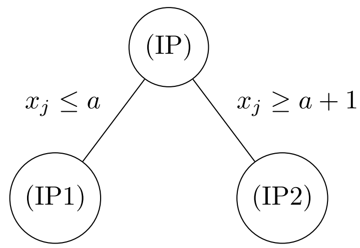
$$\begin{aligned} z_2 = \max \quad & c^T x \\ \text{s.t.} \quad & Ax \leq b \\ & x_j \geq a + 1 \\ & x \in \mathbb{Z}^n \end{aligned} \tag{IP2}$$

Notice that if x^* is the optimal solution for (IP), then it is either feasible for (IP1) or (IP2). Then

$$z_{IP} = \max\{z_1, z_2\}$$

We will represent this by a B&B tree:

Example 8.3



Althought the number of problems grow exponentially, but the problems are simplified as we branch.

Definition 8.8: Branching

This is what's called **branching**.

8.5.2 Bounding

The idea here is:

`Look at a node of the B&B tree.`
`Let`

$\max c^T x$

$s.t.\ x \in R$

`be the LP relaxation of the corresponding IP.`

Discovery 8.3: Fathom/ Prune by Infeasibility

If $R = \varnothing$, then the corresponding IP is also infeasible, thus we may stop branching from that node.
We call this *"Fathom/ Prune by Infeasibility"*.

Discovery 8.4: Fathom/ Prune by Optimality

If x^* is an optimal solution for LP relaxation

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & x \in R \end{array}$$

of a node in the B&B tree and is integral, then x^* is optimal for the IP corresponding to that node, we stop branching in this case as well.
We call this *"Fathom/ Prune by Optimality"*.

Note 8.12

At any point of the time, we keep track of the best integral solution and value found at that time. At the start, we set

$$\begin{aligned} z_{best} &\longleftarrow -\infty \\ x_{best} &\longleftarrow \text{undefined} \end{aligned}$$

When pruning by optimality, if $c^T x^* > z_{best}$, we set

$$\begin{aligned} z_{best} &\longleftarrow c^T x^* \\ x_{best} &\longleftarrow x^* \end{aligned}$$

Discovery 8.5: Fathom/ Prune by Bound

If x^* is optimal solution to LP relaxation of a note in the B&B tree and

$$c^T x^* \leq z_{best}$$

so one cannot find a better *integral* solution than x_{best}. We stop branching in this case.
We call this process *"Fathom/ Prune by Bound"*.

Result 8.3

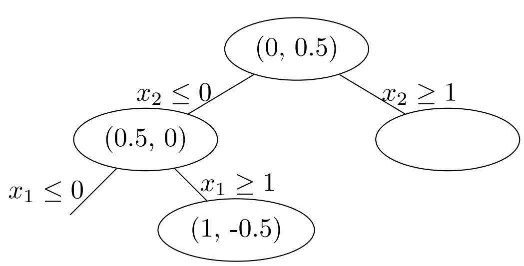
If all else fail, we look at x^*, pick j such that $x_j^* \notin \mathbb{Z}$. Set $a := \lfloor x_j^* \rfloor$ and branch by adding $x_j \leq a$, $x_j \geq a + 1$ in the "child" nodes.

Example 8.4

This is an example as the Branch and Bound could potentially be a bad algorithm. Consider the IP

$$\begin{array}{lll} \min & x_1 + x_2 & \\ \text{s.t.} & x_1 + x_2 & \geq 0.5 \\ & x_1 & \geq 0 \\ & x & \in \mathbb{Z}^2 \end{array}$$

Consider the LP-relaxation, we found the optimal solution $x^* = (0, 0.5)^T$ and the optimal value is 0.5. Apply the algorithm with x^*, we would have



Notice that if we choose the "bad" nodes, this might never terminate.

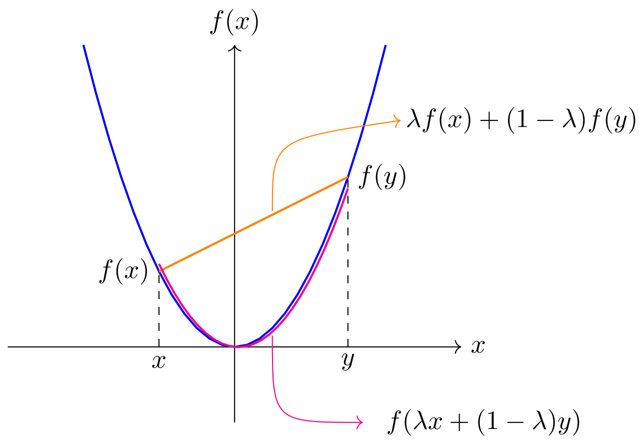
9 Nonlinear Programming (NLP)

Definition 9.1: Convex Functions

Let $S \subseteq \mathbb{R}^m$ be a convex set, the function $f : S \to \mathbb{R}$ is a **convex function** if for all $x, y \in S$ and any $\lambda \in [0,1]$,

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

Example 9.1



$f(x)$ is a convex function.

Definition 9.2: Strictly Convex

The function $f : S \to \mathbb{R}$ is **strictly convex** if

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

for all $x, y \in S$ and any $\lambda \in (0,1)$.

Result 9.1

Let $f_0, f_1, \ldots, f_m : \mathbb{R}^m \to \mathbb{R}$, NLP that we will focus on are

$$\begin{aligned} \inf \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0 \quad \forall\, i = 1, \ldots, m \end{aligned} \tag{NLP}$$

Note 9.1

This is a inf problem.

Definition 9.3: Convex Program

If $f_0, f_1, \ldots, f_m$ are convex functions, then (NLP) is called a **convex program**.

Proposition 9.1

If $f : \mathbb{R}^n \to \mathbb{R}$ is a convex function, then the set

$$\{x \in \mathbb{R}^n : f(x) \leq \alpha\} =: S$$

is a convex set, where $\alpha \in \mathbb{R}$ is a real number.

Proof. Pick $x, y \in S$, for any $\lambda \in [0, 1]$, we have

$$\begin{aligned} f(\lambda x + (1-\lambda)y) &\leq \lambda f(x) + (1-\lambda) f(y) \\ &\leq \lambda\alpha + (1-\lambda)\alpha \\ &= \alpha \end{aligned}$$

which implies that $\lambda x + (1-\lambda)y \in S$ as desired. □

9.1 Gradient and Hessian

Definition 9.4: Gradient & Hessian

Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable, then

$$\nabla f(\bar{x}) = \begin{pmatrix} \frac{\partial f}{\partial x_1}(\bar{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\bar{x}) \end{pmatrix} \qquad \text{and} \qquad \nabla^2 f(\bar{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}(\bar{x})\right]$$

where ∇ is known as gradient and ∇^2 is known as Hessian.

Theorem 9.1

Let $S \subseteq \mathbb{R}$ be a convex set, let $f : S \to \mathbb{R}$ be twice differentiable, TFAE:

1. f is convex on S;
2. $f(x) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x})$ for all $x, \bar{x} \in S$;
3. $(f'(x) - f'(\bar{x}))(x - \bar{x}) \geq 0$ for all $x, \bar{x} \in S$;

4. $f''(x) \geq 0$ for all $x \in S$.

Proof. `[(1) ⇒ (2)]`: Suppose $x > \bar{x}$, then

$$\lim_{\lambda \to 0} \frac{f(\bar{x} + \lambda(x - \bar{x})) - f(\bar{x})}{\lambda(x - \bar{x})} = f'(\bar{x})$$

Since we know that

$$f(\lambda x + (1-\lambda)\bar{x}) \leq \lambda f(x) + (1-\lambda) f(\bar{x})$$

so for $\lambda \in (0,1)$,

$$\begin{aligned}
\frac{f(\lambda x + (1-\lambda)\bar{x})}{\lambda(x - \bar{x})} &\leq \frac{\lambda f(x) + (1-\lambda) f(\bar{x})}{\lambda(x - \bar{x})} \\
\frac{f(\lambda x + (1-\lambda)\bar{x}) - f(\bar{x})}{\lambda(x - \bar{x})} &\leq \frac{\lambda f(x) + (1-\lambda) f(\bar{x}) - f(\bar{x})}{\lambda(x - \bar{x})} = \frac{f(x) - f(\bar{x})}{x - \bar{x}}
\end{aligned}$$

We can now observe that

$$f'(\bar{x}) \leq \frac{f(x) - f(\bar{x})}{x - \bar{x}}$$

`[(2) ⇒ (1)]`: Let $x, \bar{x} \in S$ and $\lambda \in [0,1]$, let $y = \lambda x + (1-\lambda)\bar{x}$, so we have

$$\begin{aligned}
(1-\lambda)\times \qquad & f(x) \geq f(y) + f'(y)(x - y) \\
\lambda\times \qquad & f(\bar{x}) \geq f(y) + f'(y)(\bar{x} - y)
\end{aligned}$$

Sum them up we obtain

$$\begin{aligned}
(1-\lambda) f(x) + \lambda f(\bar{x}) &\geq (1-\lambda) f(y) + \lambda f(y) + \\
&\quad (1-\lambda) f'(y)(x - y) + \lambda f'(y)(\bar{x} - y) \\
&= f(y) + f'(y) \underbrace{[(1-\lambda)(x - y) + \lambda(\bar{x} - y)]}_{=0}
\end{aligned}$$

The rest of the proofs are left as **exercise**.

`The following proofs are my attempted proofs:`

`[(2) ⇒ (3)]`: By [2], we know that

$$f(x) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x}) \qquad \text{and} \qquad f(\bar{x}) \geq f(x) + f'(x)(\bar{x} - x)$$

Summing them up we have

$$f(x) + f(\bar{x}) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x}) + f(x) + f'(x)(\bar{x} - x)$$

which further yields us

$$0 \geq (f'(\bar{x}) - f'(x))(x - \bar{x})$$

`[(3) ⇒ (2)]`: WLOG assume $\bar{x} < x$. By MVT, we know that there exists $c \in (x, \bar{x})$ such that

$$f'(c) = \frac{f(x) - f(\bar{x})}{x - \bar{x}}$$

Moreover, by [3] we know that

$$[f'(x) - f'(c)]\,(x - c) \geq 0 \qquad \text{and} \qquad [f'(c) - f'(\bar{x})]\,(c - \bar{x}) \geq 0$$

which implies that $f'(\bar{x}) \leq f'(c) \leq f'(c)$. Thus we have

$$f(x) = f(\bar{x}) + f'(c)(x - \bar{x}) \geq f(\bar{x}) + f'(\bar{x})(x - \bar{x})$$

□

Discovery 9.1

In general, we know that if f is convex on S, then

- For all $x, \bar{x} \in S$,
$$f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$
- For all $x, \bar{x} \in S$,
$$(\nabla f(x) - \nabla f(\bar{x}))^T (x - \bar{x}) \geq 0$$
- For all $x \in S$,
$$\nabla^2 f(x) \text{ is P.S.D. (positive semi-definite)}$$

Note 9.2

A symmetric $n \times n$ matrix Q is **positive semi-definite** (PSD) if for all $y \in \mathbb{R}^n$, $y^T Q y \geq 0$. We denote

$$Q \succeq 0$$

It is positive definite (PD) if for all $y \in \mathbb{R}^n$ with $y \neq 0$, $y^T Q y > 0$, we denote

$$Q \succ 0$$

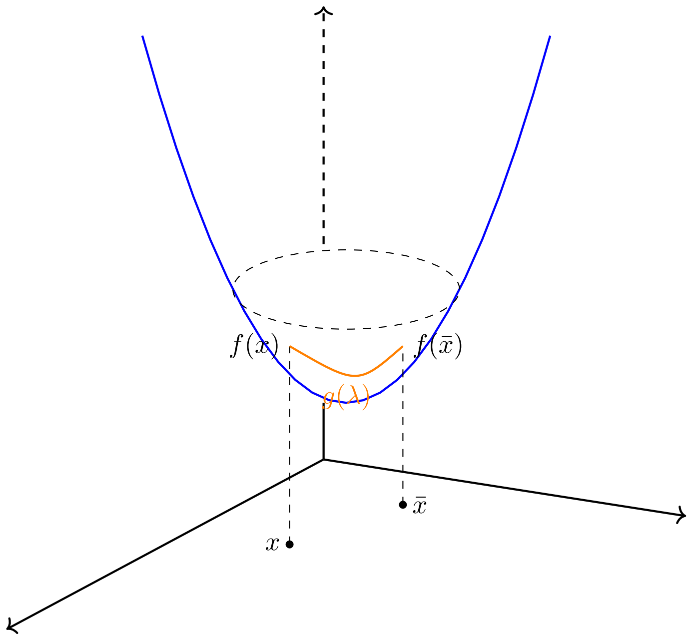
Theorem 9.2

Let $S \subseteq \mathbb{R}^n$ be a convex set, let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable, TFAE:

1. f is convex on S;
2. $f(x) \geq f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$ for all $x, \bar{x} \in S$;
3. $(\nabla f(x) - \nabla f(\bar{x}))^T (x - \bar{x}) \geq 0$ for all $x, \bar{x} \in S$;

4. $\nabla^2 f(x) \succeq 0$ for all $x \in S$.

Proof. The idea is that for $x, \bar{x} \in S$, define $g(\lambda) = f(\lambda x + (1-\lambda)\bar{x})$.



so we have $g'(0) = \nabla f(\bar{x})^T (x - \bar{x})$. ☐

Example 9.2

Consider $f : \mathbb{R}^n \to \mathbb{R}$ defined as

$$f(x) = \|x\|^2 = \sum_j x_j^2$$

then we have

$$\nabla f(x) = \begin{pmatrix} 2x_1 \\ \vdots \\ 2x_n \end{pmatrix}$$

and

$$\nabla^2 f(x) = 2I$$

Note 9.3

We have

$$y^T [\nabla^2 f(x)] y = 2y^T y = 2\|y\|^2 \geq 0$$

hence f is indeed a convex function.

9.2 Local vs. Global Optimality

Consider the (NLP) and let S denote its feasible region.

Definition 9.5: Global Optimum

A point $x^* \in S$ is a **global optimum** if

$$f_0(x^*) \leq f_0(x), \qquad \forall\, x \in S$$

Definition 9.6: Local Optimum

A point $x' \in S$ is a **local optimum** if there exists $\varepsilon > 0$ such that

$$f_0(x') \leq f(x), \qquad \forall\, x \in \mathcal{B}_\varepsilon(x') \cap S$$

Theorem 9.3

If (NLP) is a convex program, then x^* is a local optimum if and only if x^* is a global optimum.

Proof. `[Backward direction]`: this is the easy direction.
`[Forward direction]`: Suppose for a contradiction that there exists $\bar{x} \in S$ such that $f_0(\bar{x}) < f_0(x^*)$. Define

$$y := \lambda \bar{x} + (1-\lambda)x^*, \qquad \text{for } \lambda \in (0,1)$$

we know that f_0 is a convex function, so

$$f_0(y) \leq \lambda f_0(\bar{x}) + (1-\lambda) f_0(x^*) < f_0(x^*)$$

Then for any $\varepsilon > 0$, we may pick λ small enough so that $y \in \mathcal{B}_\varepsilon(x^*) \cap S$. $\square$

Discovery 9.2

Easy to see that if f_i's are non-convex, then we have a hard problem. For instance, suppose we have a binary program

$$\begin{array}{lll} \min & c^T x & \\ \text{s.t.} & Ax & \leq b \\ & x_1 & \in \{0,1\}^n \end{array}$$

which no one knows a good algorithm to solve (yet). Note that the above program is equivalent to

$$\begin{array}{llll} \min & c^T x & & \\ \text{s.t.} & a_i^T x - b_i & = 0, & \forall\, i = 1, \ldots, m \\ & x_j(1-x_j) & = 0, & \forall\, j = 1, \ldots, n \end{array}$$

Lecture 21 - Tuesday, November 26

Proposition 9.2

Consider a convex optimization problem and f_0 differentiable. Let S be the feasible region, then x^* is optimal if and only if

$$\nabla f_0(x^*)^T(x - x^*) \geq 0, \qquad \forall\, x \in S$$

Proof. `[Backward:]`
By convexity, we have

$$f_0(x) \geq f_0(x^*) + \nabla f_0(x^*)^T(x - x_0) \geq f_0(x^*) \qquad \forall\, x \in S$$

`[Forward:]`
The rough idea is that if there exists $x \in S$ such that

$$\nabla f_0(x^*)^T(x - x^*) < 0$$

Then define $g(\lambda) = f(\lambda x + (1-\lambda)x^*)$. This way, we have $g'(0) = \nabla f_0(x^*)^T(x - x^*) < 0$, which gives us that x^* is not optimal. □

Corollary 9.1

If f_0 is convex and differentiable, then x^* is optimal solution to

$$\begin{array}{ll} \inf & f_0(x) \\ \text{s.t.} & x \in \mathbb{R}^n \end{array}$$

if and only if $\nabla f_0(x^*) = 0$.

9.3 Lagrangian Duality

Consider the non-linear program

$$\begin{array}{ll} \inf & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \quad \forall\, i = 1, \dots, m \end{array} \tag{NLP}$$

Definition 9.7: Lagrangian

We define

$$L(x, \lambda) := f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$

L is a function from $\mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and is called the **Lagrangian**. $\lambda_1, \dots, \lambda_m$ are called **Lagrangian multipliers** accosiated with (NLP).

Proposition 9.3

Let

$$S = \{x \in \mathbb{R}^n : f_i(x) \le 0, \quad \forall\, i = 1, \dots, m\}$$

If $\bar{x} \in S$, $\lambda \ge 0$, then

$$L(\bar{x}, \lambda) \le f_0(\bar{x})$$

Proof. We have $\sum_i \lambda_i f_i(\bar{x}) \le 0$. □

Discovery 9.3

Let

$$g(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda)$$

it follows that for any $\lambda \ge 0$ and $\bar{x} \in S$, we have

$$g(\lambda) \le f_0(\bar{x})$$

Thus we get a lower bound for the optimal value for any $\lambda \ge 0$. Recall that in duality, we wanted to find the highest lower bound, thus we want

$$\begin{array}{rl} \max & g(\lambda) \\ \text{s.t.} & \lambda \ge 0 \end{array}$$

This is called **Lagrangian Dual**.

Definition 9.8: Lagrangian Dual

This is called **Lagrangian Dual**.

9.3.1 Weak Duality

Proposition 9.4: Weak Duality

If $\bar{x} \in S$, $\lambda \ge 0$, then

$$g(\lambda) \le f_0(\bar{x})$$

Example 9.3

Consider

$$\begin{array}{rll} \inf & (x_1 - 1)^2 + (x_2 - 1)^2 & \\ \text{s.t.} & x_1 + 2x_2 - 1 & \le 0 \\ & 2x_2 + x_2 - 1 & \le 0 \end{array}$$

Then we have

$$\begin{aligned} L(x,\lambda) = \quad & (x_1-1)^2 + (x_2-1)^2 + \\ & \lambda_1(x_1+2x_2-1) + \\ & \lambda_2(2x_2+x_2-1) \end{aligned}$$

(Check if $L(x,\lambda)$ is convex on x for fixed λ). We wish to solve for $g(\lambda) = \min_{x\in\mathbb{R}^2} L(x,\lambda)$. Compute to solve $\nabla_x L(x,\lambda) = 0$ we obtain

$$\begin{pmatrix} 2(x_1-1)+\lambda_1+2\lambda_2 \\ 2(x_2-1)+2\lambda_1+\lambda_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

which yields us the solution

$$x_1^* = \frac{-\lambda_1 - 2\lambda_2}{2} + 1 \qquad \text{and} \qquad x_2^* = \frac{-2\lambda_1 - \lambda_2}{2} + 1$$

Thus we have

$$L(x^*,\lambda) = -1.25\lambda_1^2 - 1.25\lambda_2^2 - 2\lambda_1\lambda_2 + 2\lambda_1 + 2\lambda_2$$

so

$$\begin{array}{lll} \sup & g(\lambda) & \\ \text{s.t.} & \lambda & \geq 0 \end{array} = \begin{array}{lll} \sup & L(x^*,\lambda) & \\ \text{s.t.} & \lambda & \geq 0 \end{array}$$

For $\lambda = \left(\frac{4}{9}, \frac{4}{9}\right)$, we have $L(x^*,\lambda) = 8/9$. We also have $f_0\left(\frac{1}{3}, \frac{1}{3}\right) = 8/9$. So $(1/3, 1/3)$ is our optimal solution.

Note 9.4

Here comes a natural question to ask: "When does strong duality hold?"

For now, let us assume that $f_0, f_1, \ldots, f_m$ are all convex. Also assume that there exists $\bar{x}$ such that

$$f_i(\bar{x}) < 0, \qquad \forall\, i = 1, \ldots, m$$

9.4 Slater's Condition

Definition 9.9: Slater's Condition

The existence of such $\bar{x}$ is called the **Slater's Condition**.

Theorem 9.4

If $f_0, f_1, \ldots, f_m$ are all convex, p^* is finite, and Slater's condition holds, then there exists λ^* such that

$$\underbrace{\inf_{x\in\mathbb{R}^n} L(x,\lambda^*)}_{g(\lambda^*)} = \begin{array}{ll} \inf & f_0(x) \\ \text{s.t.} & f_i(x) \leq 0, \qquad \forall\, i = 1, \ldots, m \end{array}$$

Proof. Later (see 9.6.1). □

Example 9.4

Consider

$$\begin{aligned} \inf \quad & x \\ \text{s.t.} \quad & x^2 \leq 0 \end{aligned}$$

Hence $x = 0$ is our only feasible solution, which is hence optimal. Notice that Slater's condition does not hold here. Here, we have

$$L(x, \lambda) = x + \lambda x^2$$

and

$$g(\lambda) = \min_x x + \lambda x^2 = \begin{cases} -\infty & \text{if } \lambda = 0 \\ -\frac{1}{2\lambda} & \text{if } \lambda > 0 \end{cases}$$

which is an counter example of the above theorem when Slater's condition does not hold, proving it is necessary.

Suppose we have x^*, λ^* optimal for primal/ dual (not always exist), then

$$\begin{aligned} g(\lambda^*) = \min_{x \in \mathbb{R}^n} & \left[f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) \right] \\ & \leq f_0(x^*) + \sum_{i=1}^m \lambda_i^* f_i(x^*) \\ & \leq f_0(x^*) \end{aligned}$$

If we want strong duality to hold, i.e.,

$$g(\lambda^*) = f_0(x^*)$$

Hence we want both inequalities above to hold as equality.

1. *If the first inequality holds as equality:* we know that x^* is optimal solution to

   $$\min_{x \in \mathbb{R}^n} f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) \equiv \min_{x \in \mathbb{R}^n} L(x, \lambda^*)$$

   This means that $\min_{x \in \mathbb{R}^n} L(x, \lambda^*) = L(x^*, \lambda^*)$. If $f_0, f_1, \ldots, f_m$ are all convex and differentiable, then

   $$\nabla_x L(x^*, \lambda^*) = 0$$

   Strongerly, if $f_0, f_1, \ldots, f_m$ are all differentiable, then

   $$\nabla_x L(x^*, \lambda^*) = 0$$

   Hence strong duality tells us that if f's are differentiable, then

   $$\nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) = 0$$

2. *If the second inequality holds as equality:* we know that

$$\lambda_i^* f_i(x^*) = 0, \qquad \forall\, i = 1, \ldots, m$$

which is the complementary slackness-type. Because it implies that we have (inclusive or)

$$\boxed{\lambda_i^* = 0 \qquad \text{or} \qquad f_i(x^*) = 0}$$

Lecture 22 - Thursday, November 28

9.5 Karush-Kuhn-Tucker Conditions

Definition 9.10: Karush-Kuhn-Tucker Conditions

1. $f_i(x) \leq 0$, for all $i = 1, \ldots, m$
2. $\lambda \geq 0$
3. $\lambda_i f_i(x) = 0$ for all $i = 1, \ldots, m$
4. $\nabla f_0(x) + \sum_{i=1}^m \lambda_i \nabla f_i(x) = 0$

Consider

$$\begin{array}{lll} \inf & f_0(x) & \\ \text{s.t.} & f_i(x) & \leq 0 \quad \forall\, i = 1, \ldots, m \end{array} \tag{NLP}$$

where f_i are differentiable.

9.5.1 Necessary Optimality Conditions

Theorem 9.5: Necessary Optimality Conditions

Suppose $\bar{x}$, $\bar{\lambda}$ are feasible for (NLP) and its dual such that $f_0(\bar{x}) = g(\bar{\lambda})$, then KKT conditions hold.

Note 9.5

For p^* and d^* defined by

$$p^* = \begin{array}{lll} \inf & f_0(x) & \\ \text{s.t.} & f_i(x) & \leq 0 \quad \forall\, i = 1, \ldots, m \end{array}$$

and

$$d^* = \begin{array}{lll} \sup & g(\lambda) & \\ \text{s.t.} & \lambda & \geq 0 \end{array}$$

We may have

1. $p^* > d^*$, or
2. $p^* = d^*$ but there does not exist $(\bar{x}, \bar{\lambda})$ such that they satisfy KKT.

9.5.2 Sufficient Optimality Conditions

Theorem 9.6: Sufficient Optimality Conditions

If $\bar{x}, \bar{\lambda}$ satisfy KKT and f_i is differentiable and convex for all $i = 0, \dots, m$, then

$$f_0(\bar{x}) = g(\bar{\lambda})$$

9.6 Summary on NLP's

	Generic (NLP)	Generic and Differentiable	Convex	Convex and differentiable
Weak Duality, $d^* \leq p^*$	Yes	Yes	Yes	Yes
Slater $\Rightarrow \exists\, \bar{\lambda} : g(\bar{\lambda}) = p^*$	No	No	Yes	Yes
Necassary KKT	No	Yes	No	Yes
Sufficient KKT	No	No	No	Yes

Example 9.5

Consider the example

$$\begin{array}{lll} \inf & (x_1 - 1)^2 + (x_2 - 1)^2 & \\ \text{s.t.} & x_1 + 2x_2 - 1 & \leq 0 \\ & 2x_2 + x_2 - 1 & \leq 0 \end{array}$$

Then for the fourth condition in KKT, we have

$$\begin{pmatrix} 2(x_1 - 1) \\ 2(x_2 - 1) \end{pmatrix} + \lambda_1 \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \lambda_2 \begin{pmatrix} 2 \\ 1 \end{pmatrix} = 0$$

which could further gives

$$x_1^* = \frac{-\lambda_1 - 2\lambda_2}{2} + 1 \qquad \text{and} \qquad x_2^* = \frac{-2\lambda_1 - \lambda_2}{2} + 1$$

Then, one can try to see which λ_i is allowed to be zero by going through all the four possibilities to get a KKT point.

9.6.1 Slater's Condition Proof

Proof. This is the proof for the Theorem (9.4) in regards to Slater's Condition.
Let

$$\mathcal{G} = \left\{ (f_0(x), \dots, f_m(x)) \in \mathbb{R} \times \mathbb{R}^m : x \in \mathbb{R}^n \right\}$$

Also let

$$\mathcal{A} = \left\{ (t, u) \in \mathbb{R} \times \mathbb{R}^m : \exists\, x \in \mathbb{R} \text{ s.t. } f_0(x) \leq t, f_i(x) \leq u_i \right\}$$

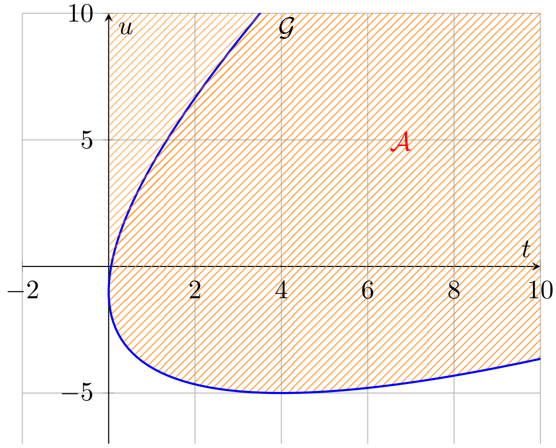
Here are two examples for the sake of illustration

Example 9.6

For instance, consider

$$f_0(x) = (x-1)^2$$
$$f_1(x) = (x-3)^2 - 5$$

so we have $\mathcal{G} = \{(x-1)^2, (x-3)^2 - 5 : x \in \mathbb{R}^n\}$

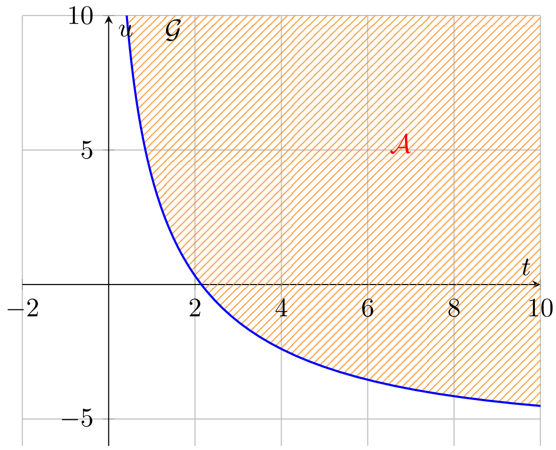


Example 9.7

For instance, consider

$$f_0(x) = e^x$$
$$f_1(x) = (x-3)^2 - 5$$

so we have $\mathcal{G} = \{e^x, (x-3)^2 - 5 : x \in \mathbb{R}^n\}$



`Claim 1:` $\mathcal{A}$ `is a convex set`

Proof of claim 1: **exercise**.

`Claim 2:` `we have` $p^* = \inf\{t : (t, 0_m) \in \mathcal{A}\}$
Proof of claim 2: For any $(t, 0) \in \mathcal{A}$, there exists $\bar{x}$ such that

$$f_0(\bar{x}) \leq t \qquad \text{and} \qquad f_i(\bar{x}) \leq 0$$

This implies that $p^* \leq t$. If for some $\varepsilon > 0$ we have $p^* + \varepsilon$, then this implies that there exists $\bar{x}$ such that

$$f_0(\bar{x}) \leq p^* + \varepsilon \qquad \text{and} \qquad f_i(\bar{x}) \leq 0 \qquad \forall\, i \in [m]$$

because $p^* = \inf f_0(x)$ such that $f_i(x) \leq 0$. Furthermore, this means that $p^* + \varepsilon$ is not a lower bound on $t : (t, 0) \in \mathcal{A}$.
Also fix $\bar{\lambda} \geq 0$. We consider

$$g(\bar{\lambda}) := \inf L(x, \bar{\lambda}) \qquad \text{s.t. } x \in \mathbb{R}^n$$

`Claim 3:` `we have`

$$g(\bar{\lambda}) = \inf \left\{ t + \sum_{i=1}^m \bar{\lambda}_i u_i : (t, u) \in \mathcal{A} \right\}$$

Proof of Claim 3: We have

$$g(\bar{\lambda}) \leq f_0(x) + \sum_{i=1}^m \bar{\lambda}_i f_i(x) \qquad \forall\, x \in \mathbb{R}^n$$

which implies that

$$g(\bar{\lambda}) \leq t + \sum_{i=1}^m \bar{\lambda}_i u_i \qquad \forall\, (t, u) \in \mathcal{A}$$

and if $\varepsilon > 0$, consider $g(\bar{\lambda}) + \varepsilon$, there exists $\bar{x}$ such that $g(\bar{\lambda}) + \varepsilon > L(\bar{x}, \bar{\lambda})$, which implies that there exists $(\bar{t}, \bar{u}) \in \mathcal{A}$ such that

$$\bar{t} + \sum_{i=1}^m \bar{\lambda}_i \bar{u}_i < g(\bar{\lambda}) + \varepsilon$$

Now let $B = \{(s, 0) \in \mathbb{R} \times \mathbb{R}^m : s < p^*\}$. We know that B is a convex set and $B \cap \mathcal{A} = \varnothing$. This implies that there exists $\tilde{\gamma}, \tilde{\lambda} \in \mathbb{R} \times \mathbb{R}^m$ such that $(\tilde{\gamma}, \tilde{\lambda}) \neq 0$ and $\tilde{\alpha} \in \mathbb{R}$ such that

$$\begin{aligned} \tilde{\lambda}^T u + \tilde{\gamma} t &\geq \tilde{\alpha}, \qquad \forall\, (t, u) \in \mathcal{A} \\ \tilde{\lambda}^T u + \tilde{\gamma} t &\leq \tilde{\alpha}, \qquad \forall\, (t, u) \in B \end{aligned}$$

But since we can increase u, t arbitrarily large by the definition of $\mathcal{A}$, we must have $\tilde{\lambda}, \tilde{\gamma} \geq 0$.

Lecture 23 - Tuesday, December 03

For $(s, 0) \in B$,

$$\tilde{\gamma} \cdot s \leq \tilde{\alpha}, \qquad \forall\, s < p^*$$

thich implies $\tilde{\gamma} p^* \leq \tilde{\alpha}$. Therefore, for all $x \in \mathbb{R}^n$,

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(x) + \tilde{\gamma} f_0(x) \geq \tilde{\alpha} \geq \tilde{\gamma} p^*$$

1. `Case 1:` $\tilde{\gamma} > 0$

We define $\lambda^* := \dfrac{\tilde{\lambda}}{\tilde{\gamma}}$, so for all $x \in \mathbb{R}^n$, we have

$$\sum_{i=1}^{m} \lambda^* f_i(x) + f_0(x) \geq p^* \quad \Rightarrow \quad g(\lambda^*) \geq p^*$$

and since weak duality holds, or $g(\lambda^*) \leq p^*$, we must have $g(\lambda^*) = p^*$.

2. `Case 1:` $\tilde{\gamma} = 0$
   For all $x \in \mathbb{R}^n$, we get
   $$\sum_{i=1}^{m} \tilde{\lambda}_i f_i(x) \geq 0$$
   Let $\bar{x}$ be such that $f_i(x) < 0$ for all $i = 1, \dots, m$. Combining with the fact that $\tilde{\lambda} \geq 0$, we must have $\tilde{\lambda} = 0$. However, this contradicts the fact that $(\tilde{\lambda}, \tilde{\gamma}) \neq 0$.

□

9.7 Sufficient Conditions for Having Optimal Solution to (NLP)

Theorem 9.7

If f_0 is continuous and feasible region is compact, then the (NLP) has an optimal solution.

Theorem 9.8

If there exists α such that the set

$$\left\{ x \in \mathbb{R}^n : \begin{array}{ll} f_0(x) & \leq \alpha \\ f_i(x) & \leq 0, \qquad \forall\, i = 1, \dots, m \end{array} \right\}$$

is non-empty, closed, and bounded, and f_0 is continuous. Then there exists an optimal solution.

10 Algorithms for Convex NLP

I am going to be really Handwavy

Ricardo Fukasawa

10.1 Unconstrained Problems

Consider the program

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & x \in \mathbb{R}^n \end{aligned}$$

Our goal is to find x^* such that $\nabla f_0(x^*) = 0$.

10.1.1 Descent Methods

Algorithm 10.1

1. Start from $x^i \in \mathbb{R}^n$;
2. Find $d^k \in \mathbb{R}^n$ such that
$$\left(d^k\right)^T \nabla f_0(x^i) < 0$$
3. Find step size t^k;
4. $x^{k+1} \longleftarrow x^k + t^k d^k$.

10.2 Constrained Problems

Consider the problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leq 0, \qquad \forall\, i = 1, \ldots, m \end{aligned}$$

We also assume the following things:

- There exists an optimal solution;
- f is convex and differentiable;
- Slater's Condition holds here.

Our problem is equivalent to

$$\min \quad f_0(x) + \sum_{i=1}^{m} I_-(f_i(x))$$

where we define $I_- : \mathbb{R} \to \mathbb{R} \cup \{\infty\}$ by

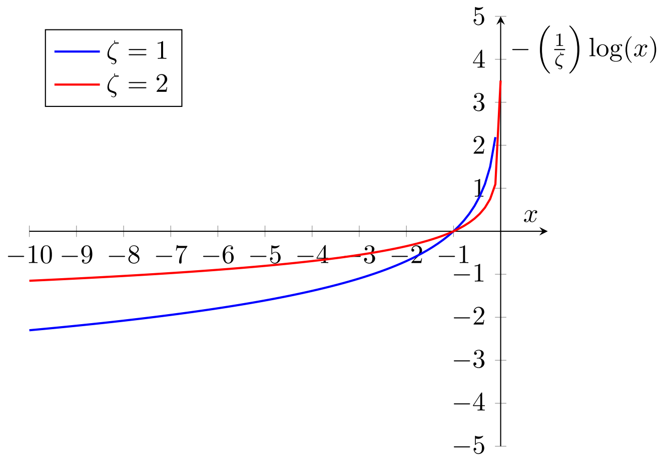
$$I_-(u) = \begin{cases} 0, & u \leq 0 \\ \infty, & u > 0 \end{cases}$$

Instead we consider the function

$$-\left(\frac{1}{\zeta}\right)\log(-u), \qquad \text{for } \zeta > 0$$

which has the image as

Example 10.1: Plot of the function $-\left(\frac{1}{\zeta}\right)\log(x)$ for $\zeta = 1$ and 2



Observe that our function is continuous and convex. Hence we wish to solve the unconstrained problem:

$$\begin{array}{ll} \min & f_0(x) + \sum_{i=1}^{m}\left(-\frac{1}{\zeta}\right)\log(-f_i(x)) \\ \text{s.t.} & x \in \mathbb{R}^m \end{array}$$

11 Comments on CO Courses

- `Discrete Optimization:`

$$CO452 \qquad CO450 \qquad CO353$$

- `NLP:`

$$CO367 \qquad CO471(SDP) \qquad CO463(Convex)$$

- `Network Flows:`

$$CO351$$

- `Game Theory`

$$CO456$$

Index