

MATH561/IOE510/TO518  
Linear Programming

Pingbang Hu

December 9, 2022

## Abstract

This is the first course in the series of graduate-level, large-scale and rigorous mathematical programming courses taught by [Jon Lee](#), and in particular, we will use the book write by professor Lee [[Lee22](#)]. This is a dynamic book which may changes and update constantly. In this course, we focus on developing a rigorous understanding on large-scale linear programming problems and also introduce the basic concept of integer programming.

# Contents

<b>1</b>	<b>Introduction to Linear Programming</b>	<b>2</b>
1.1	General Linear Programming Problem . . . . .	2
1.2	First Glance of Duality . . . . .	4
<b>2</b>	<b>Production Problem</b>	<b>6</b>
2.1	Production Problem . . . . .	6
2.2	Norm . . . . .	6
<b>3</b>	<b>Algebra v.s. Geometry</b>	<b>8</b>
3.1	Elementary Row Operations . . . . .	8
3.2	Basic Partition . . . . .	8
3.3	Convex Set . . . . .	9
3.4	Extreme Point . . . . .	10
3.5	Feasible Directions . . . . .	10
3.6	Feasible Rays . . . . .	13
<b>4</b>	<b>Simplex Algorithm</b>	<b>15</b>
4.1	Simplex Algorithm . . . . .	15
4.2	Remaining Problem . . . . .	19
4.3	The Word <i>Simplex</i> . . . . .	22
4.4	Complete Simplex Algorithm . . . . .	24
<b>5</b>	<b>Duality</b>	<b>25</b>
5.1	Strong Duality Theorem . . . . .	25
5.2	Complementary . . . . .	26
5.3	Duality for General Linear Optimization Problems . . . . .	27
5.4	Geometrically Understanding of Duality . . . . .	31
5.5	The Big Picture of Cones . . . . .	34
<b>6</b>	<b>Sensitivity Analysis</b>	<b>40</b>
6.1	Local Analysis . . . . .	40
6.2	Global Analysis . . . . .	41
6.3	More on Local Analysis . . . . .	44
6.4	More on Global Analysis . . . . .	46
<b>7</b>	<b>Large-Scale Linear Optimization</b>	<b>48</b>
7.1	Decomposition Algorithm . . . . .	48
7.2	Solution of the Master Problem via the Simplex Algorithm . . . . .	52
7.3	Lagrangian Relaxation . . . . .	55
7.4	Cutting-Stock Problem . . . . .	62
<b>8</b>	<b>Integer-Linear Optimization</b>	<b>66</b>
8.1	Modeling Techniques . . . . .	67
8.2	Algorithmically Solving Integer-Programming Problem . . . . .	70

# Chapter 1

## Introduction to Linear Programming

### Lecture 1: Introduction

#### 1.1 General Linear Programming Problem

30 Aug. 08:00

Let's start with the definition of a [linear programming problem](#).

**Definition 1.1.1** (General linear programming problem). A *general linear programming problem* is to either minimize or maximize an [objective function](#) with the [constraints](#) defined as follows.

**Definition 1.1.2** (Objective function). An *objective function* is in the form of

$$c_1x_1 + c_2x_2 + \dots + c_nx_n,$$

where  $x_i$  are our variables,  $i = 1, \dots, n$ .

**Definition 1.1.3** (Constraints). The *constraints* are the combination of [structured constraints](#) and also the [signed constraints](#).

**Definition 1.1.4** (Structured constraints). The *structured constraints* are in the form of

$$\begin{array}{ccccccc} a_{11}x_1 + \dots + a_{1n}x_n & \begin{array}{c} \geq \\ \leq \end{array} & b_1 \\ a_{21}x_1 + \dots + a_{2n}x_n & \begin{array}{c} \geq \\ \leq \end{array} & b_2 \\ \vdots & & \vdots \\ a_{n1}x_1 + \dots + a_{nn}x_n & \begin{array}{c} \geq \\ \leq \end{array} & b_n. \end{array}$$

**Definition 1.1.5** (Signed constraints). The *signed constraints* are referring to  $x_i \begin{array}{c} \geq \\ \leq \end{array} 0$ .

Given a [general linear programming problem](#), we have the following definitions.

**Definition 1.1.6** (Solution). We called an assignment of values to variable  $x$  as a *solution*.

**Definition 1.1.7** (Feasible solution). If this [solution](#) satisfies the [linear constraints](#), we say that this [solution](#) is a *feasible solution*.

**Definition 1.1.8** (Feasible region). The set of [feasible solutions](#) is called *feasible region*.

**Definition 1.1.9 (Optimal solution).** And a **solution** is an *optimal solution* if there is no **feasible solution** with better objective value.

**Remark.** A **feasible region** is a polyhedron.

**Notation.** We often referred  $\geq$  to either  $\geq, \leq$  or  $=$ .

We will denote

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}, \quad b = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

It's convenient to only consider linear programming in the **standard form** defined as follows.

**Definition 1.1.10 (Standard form).** The *standard form* linear programming problem has the form of

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

with the condition that rows of  $A$  are linear independent.

**Remark (Compactness of the solution space).** Notice that we only consider finitely many of **constraints**, since the property that the **objective function** can attain its extremum only on a compact set, which requires finite dimensional vector space.

**Remark (Convert to standard form).** Surprisingly, every **general linear programming problem** can be converted to **standard form**, we now see how is this done.

**Proof.** Given a **general linear programming problem** with our notation, we see that

- Sign:
  - If  $x_j \leq 0 \Rightarrow x_j \rightarrow -x_j^-$ , where  $x_j^- \geq 0$ .
  - If  $x_j$  is unrestricted  $\Rightarrow x_j \rightarrow x_j^+ - x_j^-$ , where  $x_j^\pm \geq 0$ .
- Constraints:
  - $\sum_{j=1}^n a_{ij}x_j \leq b \Rightarrow \sum_{j=1}^n a_{ij}x_j + s_i = b_i$ , where  $s_i \geq 0$ .

**Definition 1.1.11 (Slack variable).** This  $s_i$  sometimes called *slack variable*.

$$- \sum_{j=1}^n a_{ij}x_j \geq b \Rightarrow \sum_{j=1}^n a_{ij}x_j - s_i = b_i, \text{ where } s_i \geq 0.$$

**Definition 1.1.12 (Surplus variable).** This  $s_i$  sometimes called *surplus variable*.

- Maximize:  $\max \sum c_j x_j \Rightarrow -\min -\sum c_j x_j$ .

⊛

## Lecture 2: Duality

### 1.2 First Glance of Duality

1 Sep. 08:00

We can associate the [standard form](#) problem with another linear programming problem, called the [dual](#) of the original problem.

**Definition 1.2.1 (Primal and Dual).** Given a [standard form](#) linear programming problem denoted as  $(P)$ , we have the induced problem denoted as  $(D)$  as follows.

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \quad \begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top. \\ (D) & \end{array}$$

**Definition 1.2.2 (Primal).** We sometimes called  $(P)$  the *primal*.

**Definition 1.2.3 (Dual).** The *dual* of the [primal](#) is the problem  $(D)$ .

**Note.** We see that the [dual](#) is equivalent to

$$\begin{array}{ll} \max & b^\top y \\ & A^\top y \leq c. \end{array}$$

Then we have a direct, but important theorem.

**Theorem 1.2.1 (Weak duality theorem).** If  $\hat{x}$  is [feasible](#) for  $(P)$ , and  $\hat{y}$  is [feasible](#) for  $(D)$ , then we have

$$c^\top \hat{x} \geq \hat{y}^\top b.$$

**Proof.** Since we have

$$\hat{y}^\top A \leq c^\top \xRightarrow{\hat{x} \geq 0} \hat{y}^\top A \hat{x} \leq \hat{c}^\top \hat{x} \xRightarrow{A\hat{x}=b} \hat{y}^\top b \leq c^\top \hat{x},$$

the result follows. ■

**Example.** Consider

$$\begin{array}{ll} \min & c^\top x \\ & Ax \geq b, \end{array}$$

turn this into the [standard form](#) problem and find the [dual](#).

**Proof.** We see that  $x$  is unrestricted. We first minus a [surplus variable](#)  $S$ , we have

$$\begin{array}{ll} \min & c^\top x \\ & Ax - S = b \\ & S \geq 0. \end{array}$$

Now, we turn  $x$  into  $x^+ - x^-$ , namely

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad x^+ := \begin{pmatrix} x_1^+ \\ \vdots \\ x_n^+ \end{pmatrix}, \quad x^- := \begin{pmatrix} x_1^- \\ \vdots \\ x_n^- \end{pmatrix}, \quad x^\pm \geq \vec{0}.$$

Then we see the original problem becomes

$$\begin{aligned} \min \quad & c^\top (x^+ - x^-) \\ & A(x^+ - x^-) - S = b \\ & x^+, x^-, S \geq 0 \end{aligned}$$

We can further have

$$\begin{aligned} \min \quad & (c^\top \quad -c^\top \quad 0) \begin{pmatrix} x^+ \\ x^- \\ S \end{pmatrix} \\ & (A \quad -A \quad -I) \begin{pmatrix} x^+ \\ x^- \\ S \end{pmatrix} = b \\ & \begin{pmatrix} x^+ \\ x^- \\ S \end{pmatrix} \geq 0. \end{aligned}$$

Set the **dual** variable being  $y$ , we further have

$$\begin{aligned} \max \quad & y^\top b \\ & y^\top (A \quad -A \quad -I) \leq (c^\top \quad -c^\top \quad 0^\top.) \end{aligned}$$

⊗

**Note.** The **dual** of the **dual** is the **primal**.

**Exercise.** Show the above assertion.

# Chapter 2

## Production Problem

### Lecture 3: Production Problem

#### 2.1 Production Problem

8 Sep. 08:00

The *production problem* can be formulated as follows.

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

- $n$  products activities
- $c_j$  = per-unit revenue for activity  $j = 1 \dots n$
- $b_i$  = resource endowment for resource  $i = 1 \dots m$
- $a_{ij}$  = amount of resource  $i$  consumed by activity  $j$

#### 2.2 Norm

We start with some definitions.

**Definition 2.2.1.** We define the following different norms.

**Definition 2.2.2 (Maximum norm).** The *maximum norm* is defined as

$$\|x\|_\infty := \max_{1 \leq i \leq n} \{|x_i|\}.$$

**Definition 2.2.3 (1-norm).** The *1-norm* is defined as

$$\|x\|_1 := \sum_{i=1}^n |x_i|.$$

**Definition 2.2.4 (2-norm).** The *2-norm* is defined as

$$\|x\|_2 := \sqrt{\sum_{i=1}^n x_i^2}.$$

We can easily find the respective norm for  $x$  by following linear optimization problems.



### 2.2.1 Maximum (Infinity) Norm

Consider

$$\begin{aligned} \min \quad & \|x\|_\infty \\ & Ax = b, \end{aligned}$$

we set up

$$\begin{aligned} \min \quad & t \\ & t \geq x_i, \text{ for } i = 1, \dots, n \\ & t \leq -x_i, \text{ for } i = 1, \dots, n \\ & Ax = b. \end{aligned}$$

We see that the linear optimization **pressure** will force the maximum of  $|x_i|$  being small, hence we'll get the minimum among  $|x_i|$ .

### 2.2.2 1-Norm

Consider

$$\begin{aligned} \min \quad & \|x\|_1 \\ & Ax = b, \end{aligned}$$

we set up

$$\begin{aligned} \min \quad & \sum_{i=1}^n t_i \\ & t_i \geq x_i, \text{ for } i = 1, \dots, n \\ & t_i \leq -x_i, \text{ for } i = 1, \dots, n \\ & Ax = b. \end{aligned}$$

Again, we see that the linear optimization pressure will force  $t_i$  goes to  $|x_i|$ , resulting  $\sum_{i=1}^n t_i$  being  $\|x\|_1$ .

**Remark.** Minimize  $\|x\|_1$  tends to make  $x$  **spars** (lots of zeros).

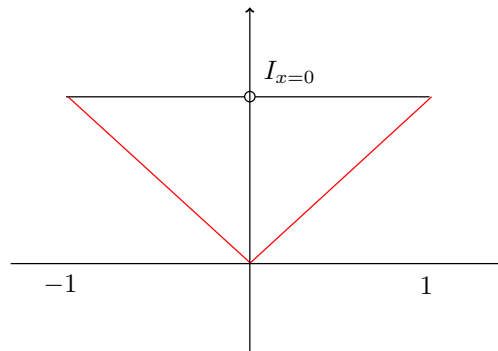


Figure 2.1: The best approximated convex function of  $I_{x=0}$

# Chapter 3

## Algebra v.s. Geometry

### Lecture 4: Basis Partition

#### 3.1 Elementary Row Operations

13 Sep. 08:00

There are some simple operations we can apply to a matrix called [elementary row operations](#).

**Definition 3.1.1** (Elementary row operations). The following operations are called *elementary row operations*.

- (a) Permute rows.
- (b) Multiply a row by a non-zero factor.
- (c) Add a multiple of a row to another row.
- (d) Permutation in columns.

**Note** (Permutation in columns). Consider

$$A = [A_1, \dots, A_n]_{m \times n}.$$

A permutation is a function like

$$(1, \dots, n) \rightarrow (\sigma(1), \dots, \sigma(n)).$$

Then the permuted matrix  $A_\sigma$  is

$$A_\sigma = [A_{\sigma(1)}, \dots, A_{\sigma(n)}].$$

With the same permutation for  $x$ , we have

$$x_\sigma = \begin{pmatrix} x_{\sigma(1)} \\ \vdots \\ x_{\sigma(n)} \end{pmatrix}.$$

We then easily see that

$$Ax = \sum_{j=1}^n A_j x_j = \sum_{j=1}^n A_{\sigma(j)} x_{\sigma(j)}.$$

Hence,

$$Ax = b \Leftrightarrow A_\sigma x_\sigma = b.$$

#### 3.2 Basic Partition

**Definition 3.2.1 (Partition).** We denote a *partition* by

$$\beta := (\beta_1, \dots, \beta_m), \quad \eta := (\eta_1, \dots, \eta_{n-m}),$$

which is a partition of  $\{1, \dots, n\}$ .

**Definition 3.2.2 (Basic).**  $\beta$  is called *basic*

**Definition 3.2.3 (Non-basic).**  $\eta$  is called *non-basic*.

Given the above definitions, we have the following naturally induced notion.

**Definition 3.2.4 (Basic partition).** A *partition* is a *basic partition* if

$$A_\beta = [A_{\beta_1}, \dots, A_{\beta_m}]_{m \times m}$$

is invertible.

**Definition 3.2.5 (Basic solution).** Associate a *basic partition* with a *basic solution*  $\bar{x}$ , which is defined as

$$\bar{x}_\eta = \begin{pmatrix} \bar{x}_{\eta_1} \\ \vdots \\ \bar{x}_{\eta_{n-m}} \end{pmatrix} := \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \bar{x}_\beta = \begin{pmatrix} \bar{x}_{\beta_1} \\ \vdots \\ \bar{x}_{\beta_m} \end{pmatrix} := A_\beta^{-1} b.$$

**Intuition.** This of course makes sense, since we know that if this is a *feasible solution* for a *standard form* problem, then  $\bar{A}\bar{x} = b$ , which means

$$[A_\beta, A_\eta] \begin{pmatrix} \bar{x}_\beta \\ \bar{x}_\eta \end{pmatrix} = b \Rightarrow A_\beta \bar{x}_\beta + A_\eta \underbrace{\bar{x}_\eta}_{=0} = b \Rightarrow \bar{x}_\beta = \underbrace{A_\beta^{-1}}_{\text{invertible}} b$$

**Remark.** After choosing  $\eta$ , we see that  $\bar{x}_\beta$  is determined.

## Lecture 5: Convex Set

### 3.3 Convex Set

15 Sep. 08:00

**Definition 3.3.1 (Convex set).** A set  $S \subseteq \mathbb{R}^n$  is a *convex set* if

$$x^1, x^2 \in S, \text{ and } 0 < \lambda < 1 \Rightarrow \lambda x^1 + (1 - \lambda)x^2 \in S.$$

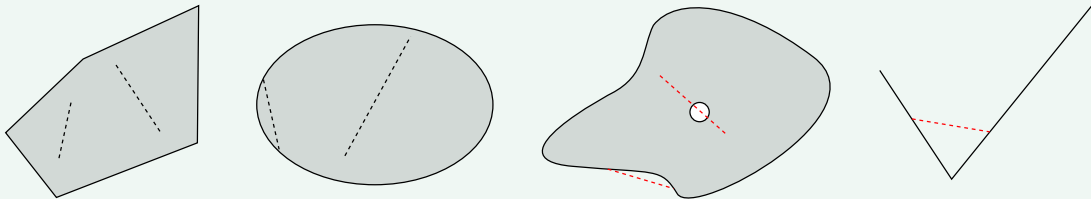


Figure 3.1: The first two are *convex sets*, while the latter two are not since (some parts of) those red lines are outside the set.

**Intuition.** A **convex set** is a set that contains every line segment between two points in which.

**Remark.** The **feasible region** of any linear program is a **convex set**.

**Proof.** Suppose there are two points  $x^1$  and  $x^2 \in S$  which means they are **feasible**. Consider a **standard form** problem, then we know

$$\begin{cases} Ax^1 = b, & x^1 \geq 0 \\ Ax^2 = b, & x^2 \geq 0 \end{cases}.$$

Then we simply have

$$A(\underbrace{\lambda x^1 + (1 - \lambda)x^2}_{\geq 0}) = \lambda Ax^1 + (1 - \lambda)Ax^2 = (\lambda + (1 - \lambda))b = b$$

for every  $\lambda \in (0, 1)$ . With the fact that  $\lambda x^1 + (1 - \lambda)x^2$  is non-negative, hence it's **feasible**.  $\otimes$

### 3.4 Extreme Point

**Definition 3.4.1** (Extreme point). Suppose  $S$  is a **convex set**. Consider  $\hat{x} \in S$ , then  $\hat{x}$  is an *extreme point* of  $S$  if we **cannot** write

$$\hat{x} = \lambda x^1 + (1 - \lambda)x^2$$

with  $x^1 \neq x^2$ ,  $x^1, x^2 \in S$ ,  $0 < \lambda < 1$ .

Then we have an important theorem.

**Theorem 3.4.1.** Every **basic feasible** solution of **standard form** problem  $(P)$  is an **extreme point** of the **feasible region** of  $(P)$ .

**Proof.** Consider a **basic feasible** solution  $\bar{x}$ :  $\bar{x}_\eta = \vec{0}$ ,  $\bar{x}_\beta = A_\beta^{-1}b \geq \vec{0}$ . If it is not an **extreme point**, then we have

$$\exists x^1 \neq x^2 \text{ which is } \text{feasible}, \text{ for } 0 < \lambda < 1 \text{ with } \bar{x} = \lambda x^1 + (1 - \lambda)x^2,$$

we will have

$$\bar{x}_\eta = \underbrace{\lambda}_{>0} \underbrace{x_\eta^1}_{>0} + \underbrace{(1 - \lambda)}_{>0} \underbrace{x_\eta^2}_{\geq 0} \Rightarrow x_\eta^1 = x_\eta^2 = 0 \Rightarrow x_\beta^1 = x_\beta^2 = A_\beta^{-1}b.$$

Hence, we see that  $\bar{x} = x^1 = x^2 \not\Leftarrow$  ■

The converse is also true, but it's harder to show.

**Theorem 3.4.2.** If  $\hat{x}$  is an **extreme point** of the **feasible region** of  $(P)$ , then  $\hat{x}$  is **basic**.

**Proof.** Skip. We leave it here. ■

## Lecture 6: Feasible Direction and Ray

### 3.5 Feasible Directions

20 Sep. 08:00

We'll talk about an important concept, but before this, we first play around with the **standard form** problem a little. Consider

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

It's obvious that it's equivalent to

$$\begin{aligned} \min \quad & c_\beta^\top x_\beta + c_\eta^\top x_\eta \\ & A_\beta x_\beta + A_\eta x_\eta = b \\ & x_\beta \geq 0, x_\eta \geq 0 \end{aligned}$$

Further, we have

$$\begin{aligned} \min \quad & c_\beta^\top (A_\beta^{-1}b - A_\beta^{-1}A_\eta x_\eta) + c_\eta^\top x_\eta \\ & x_\beta + A_\beta^{-1}A_\eta x_\eta = A_\beta^{-1}b \\ & x_\beta \geq 0, x_\eta \geq 0 \end{aligned}$$

since from the **constraints**, we have  $x_\beta = A_\beta^{-1}b - A_\beta^{-1}A_\eta x_\eta$ . Finally, we see that the **objective function** now only depends on  $x_\eta$ , hence,

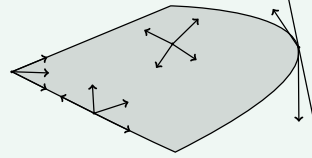
$$\begin{aligned} c_\beta^\top A_\beta^{-1}b + \min \quad & (c_\eta^\top - c_\beta^\top A_\beta^{-1}A_\eta)x_\eta \\ & A_\beta^{-1}A_\eta x_\eta \leq A_\beta^{-1}b \\ & x_\beta \geq 0, x_\eta \geq 0. \end{aligned}$$

**Note (Reduced cost).**  $c_\eta^\top - c_\beta^\top A_\beta^{-1}A_\eta$  is what we called *reduced costs*. We'll see that we want **reduced costs** to be zero.

Now, with this intuition, we have the following definition.

**Definition 3.5.1 (Feasible direction).** Suppose  $\hat{x} \in \mathcal{S}$ , where  $\mathcal{S}$  is a **convex set**.  $\hat{z}$  is a *feasible direction* relative to  $\hat{x}$  if there exists some  $\epsilon > 0$  such that

$$\hat{x} + \epsilon \hat{z} \in \mathcal{S}.$$



**Remark.** For a **primal** (P), we must have  $A\hat{z} = 0$  if  $\hat{z}$  is a **feasible direction**.

**Proof.** We see that in order to let  $\hat{z}$  to be a **feasible direction**, we need to have

$$A(\hat{x} + \epsilon \hat{z}) = \underbrace{A\hat{x}}_{=b} + \epsilon A\hat{z} = b \Leftrightarrow A\hat{z} = 0$$

⊗

Let the **basic partition**  $\beta, \eta$  being

$$\beta = (\beta_1, \dots, \beta_m), \quad \eta = (\eta_1, \dots, \underset{\uparrow \eta_j}{\eta_{n-m}}),$$

where we choose  $j$  from  $1 \leq j \leq n - m$ , which means we choose an  $\eta_j$  from  $\eta$ . Then, we see that there is a **basic direction**  $\bar{z}$  associated with this particular **basic** and this  $j$  defined as follows.

**Definition 3.5.2 (Basic direction).** Given a **basic partition**  $\beta, \eta$ , we say that  $\bar{z}$  is a *basic direction*

associated with this **basic**  $\beta$  and a  $j$  such that  $1 \leq j \leq n - m$  if

$$\bar{z}_{\eta_j} = 1 \Rightarrow \begin{cases} \bar{z}_\eta := e_j = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} \leftarrow j \\ \bar{z}_\beta := -A_\beta^{-1} A_{\eta_j}. \end{cases}$$

**Lemma 3.5.1.** Given a **basic direction**  $\bar{z}$ ,  $\bar{z}$  is **feasible** from  $\bar{x}$  if

$$0 < \min \left\{ \frac{\bar{b}_i}{\bar{a}_{i\eta_j}} \geq 0 \text{ for } i \text{ such that } \bar{a}_{i\eta_j} > 0 \right\}.$$

**Proof.** For a **basic direction**  $\bar{z}$  being **feasible**, we need to check

1.  $A(\bar{x} + \epsilon \bar{z}) = b$ : Since

$$A\bar{z} = 0 \Leftrightarrow A_\beta \bar{z}_\beta + A_\eta \bar{z}_\eta = 0 \Leftrightarrow A_\beta \bar{z}_\beta + A_\eta e_j = A_\beta \bar{z}_\beta + A_{\eta_j} = 0,$$

hence  $A\bar{z} = 0$  from the fact that  $\bar{z}_\beta = -A_\beta^{-1} A_{\eta_j}$ , which implies

$$A_\beta \bar{z}_\beta + A_{\eta_j} = A_\beta (-A_\beta^{-1} A_{\eta_j}) + A_{\eta_j} = -A_{\eta_j} + A_{\eta_j} = 0,$$

hence  $A\bar{z} = 0$ , which means  $A(\bar{x} + \epsilon \bar{z}) = A\bar{x} = b$ .  $\checkmark$

2.  $\bar{x} + \epsilon \bar{z} \geq 0$ : Since

$$\begin{aligned} \bar{x}_\eta + \epsilon \bar{z}_\eta &= 0 + \epsilon e_j \geq 0 \\ \bar{x}_\beta + \epsilon \bar{z}_\beta &= \underbrace{A_\beta^{-1} b}_{\geq 0} - \underbrace{\epsilon A_\beta^{-1} A_{\eta_j}}_{>0} \stackrel{?}{\geq} 0, \end{aligned}$$

hence we just need to make sure  $\bar{x}_\beta + \epsilon \bar{z}_\beta \geq 0$ . Denote  $\bar{b} := A_\beta^{-1} b$ ,  $\bar{A}_{\eta_j} := A_\beta^{-1} A_{\eta_j}$ , then the requirement becomes

$$\begin{aligned} \bar{b} - \epsilon \bar{A}_{\eta_j} &\geq 0 \Leftrightarrow \bar{b}_i - \epsilon \bar{a}_{i\eta_j} \geq 0, \text{ for } i = 1, \dots, m \\ &\Leftrightarrow \underbrace{\bar{b}_i}_{\geq 0} \geq \epsilon \bar{a}_{i\eta_j}, \text{ for } i = 1, \dots, m. \end{aligned}$$

We finally have

$$\epsilon \leq \frac{\bar{b}_i}{\bar{a}_{i\eta_j}}, \quad \forall_{1 \leq i \leq m} \bar{a}_{i\eta_j} > 0.$$

Notice that if  $\bar{a}_{i\eta_j} \leq 0$ , there is no restriction on  $\epsilon$  being  $\geq 0$ , so the result follows. ■

**Note.** Notice that we can denote  $A$  by

$$A = \begin{bmatrix} A_\eta & A_\beta \end{bmatrix}.$$

Then since  $A_\beta$  is invertible, so

$$A_\beta^{-1} \begin{bmatrix} A_\eta & A_\beta \end{bmatrix} = \begin{bmatrix} A_\beta^{-1} A_\eta & I \end{bmatrix}_{m \times n}.$$

Considering

$$\begin{bmatrix} I \\ -A_\beta^{-1} A_\eta \end{bmatrix},$$

we have

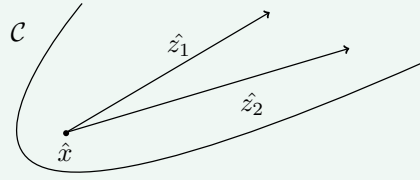
$$\underbrace{\begin{bmatrix} I \\ -A_\beta^{-1} A_\eta \end{bmatrix}}_{\dim(CS)=n-m} \underbrace{\begin{bmatrix} A_\beta^{-1} A_\eta & I \end{bmatrix}}_{\dim(RS)=m} = 0.$$

And since the dimension for the first matrix is  $n \times (m - n)$ , we see that the columns of the first matrix form a **basis** for the null space of  $[A_\eta \ A_\beta]$ , namely  $A$ . Furthermore, one can see that  $\bar{z}$  is the  $j^{th}$  columns of  $\begin{bmatrix} I \\ -A_\beta^{-1} A_\eta \end{bmatrix}$  for a choice of  $j$ .

### 3.6 Feasible Rays

**Definition 3.6.1 (Ray).**  $\hat{z}$  is called a *ray* of a **convex set**  $\mathcal{C}$  of  $\hat{x} \in \mathcal{C}$  if

$$\forall \lambda > 0 \quad \hat{x} + \lambda \hat{z} \in \mathcal{C}.$$



Suppose  $\hat{x} \in \mathcal{C}$ , where  $\mathcal{C}$  is the **feasible region** of

$$\begin{aligned} Ax &\geq b \\ x &\geq 0, \end{aligned}$$

then we see that in order to let  $\lambda$  arbitrarily large, we need

$$A(\hat{x} + \lambda \hat{z}) = \underbrace{A\hat{x}}_{=b} + \lambda \underbrace{A\hat{z}}_{=0} = b \Rightarrow \hat{z} \in n.s.(A).$$

**Problem 3.6.1.**

$$\underbrace{\hat{x}}_{\geq 0} + \underbrace{\lambda}_{> 0} \hat{z} \stackrel{?}{\geq} 0 \Rightarrow \hat{z} \geq 0.$$

This means that starts from the idea of **basic direction**,  $\hat{z}$  is a **ray** if

$$\hat{z} \geq 0 \Leftrightarrow A_\beta^{-1} A_{\eta_j} \leq 0.$$

We have another concept about **ray**.

**Definition 3.6.2 (Extreme ray).**  $\hat{z}$  is an *extreme ray* of a **convex set**  $\mathcal{S}$  if we **cannot** write

$$\hat{z} = z^1 + z^2 \text{ with } z^1 \neq \mu z^2,$$

where  $z^1, z^2$  being **rays** of  $\mathcal{S}$  and  $\mu \neq 0$ .

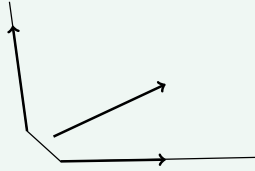


Figure 3.2: All three arrows are rays, but only the red ones are extreme.

**Remark.** We can compare the *non-negative basic direction* with *extreme ray*.

$$\text{Basic solution } \bar{x} = \begin{cases} \bar{x}_\beta := A_\beta^{-1}b \geq 0 \\ \bar{x}_\eta := 0 \end{cases} \Leftrightarrow \text{Extreme points of the feasible region}$$

<b>basic feasible direction</b>	v.s.	<b>Geometry</b>
(Basic direction that are non-negative)		(Extreme Ray)



# Chapter 4

## Simplex Algorithm

### Lecture 7: Worry-Free Simplex Algorithm

#### 4.1 Simplex Algorithm

22 Sep. 08:00

We start by considering the [standard form](#) problem

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \quad \begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top. \\ (D) & \end{array}$$

**Definition 4.1.1** (Dual basic solution). The *dual basic solution*  $\bar{y} \in \mathbb{R}^m$  is defined as

$$\bar{y}^\top = c_\beta^\top A_\beta^{-1}.$$

**Lemma 4.1.1.** If  $\beta, \eta$  is a [basic partition](#), and  $\bar{x}$  is the associated [primal basic solution](#) and  $\bar{y}$  is the associated [dual basic solution](#), then

$$c^\top \bar{x} = \bar{y}^\top b.$$

**Proof.**

$$c^\top \bar{x} = \begin{pmatrix} c_\beta^\top & c_\eta^\top \end{pmatrix} \begin{pmatrix} \bar{x}_\beta \\ \bar{x}_\eta \end{pmatrix} = c_\beta^\top \bar{x}_\beta + c_\eta^\top \bar{x}_\eta = c_\beta^\top A_\beta^{-1} b = \bar{y}^\top b.$$

Recall that

**As previously seen.**

$$\begin{array}{ll} \min & c_\beta^\top x_\beta + c_\eta^\top x_\eta \\ & A_\beta x_\beta + A_\eta x_\eta = b \\ & x_\beta \geq 0, x_\eta \geq 0, \end{array}$$

and hence

$$\begin{array}{ll} c_\beta^\top A_\beta^{-1} b + \min & (c_\eta^\top - c_\beta^\top A_\beta^{-1} A_\eta) x_\eta \\ & A_\beta^{-1} A_\eta x_\eta \leq A_\beta^{-1} b \\ & x_\beta \geq 0, x_\eta \geq 0. \end{array}$$

We now formalize the concept of [reduced cost](#).

**Definition 4.1.2** (Reduced cost).  $\bar{c}_\eta$  is called *reduced cost* for [non-basic](#) variables, where  $\bar{c}_\eta$  is defined as

$$\bar{c}_\eta^\top := c_\eta^\top - c_\beta^\top A_\beta^{-1} A_\eta = c_\eta^\top - \bar{y}^\top A_\eta.$$

### 4.1.1 Dual Feasibility

**Lemma 4.1.2.**  $\bar{y}$  is **feasible** for  $(D)$  if and only if  $\bar{c}_\eta \geq 0$ .

**Proof.**

$$y^\top A \leq c^\top \Leftrightarrow y^\top [A_\beta \quad A_\eta] \leq (c_\beta^\top \quad c_\eta^\top)$$

since

$$\begin{cases} y^\top A_\beta \leq c_\beta^\top \\ y^\top A_\eta \leq c_\eta^\top \Rightarrow c_\eta^\top - y^\top A_\eta \geq 0. \end{cases}$$

■

**Corollary 4.1.1.** If  $\hat{x}$  is **feasible** for  $(P)$  and  $\hat{y}$  is **feasible** for  $(D)$ , and if  $c^\top \hat{x} = \hat{y}^\top b$ , then  $\hat{x}$  and  $\hat{y}$  are **optimal**.

**Theorem 4.1.1** (Weak optimal basis theorem). Let  $\bar{x}$  and  $\bar{y}$  are **basic primal** and **dual solutions** for  $(P)$  and  $(D)$ . Then if  $\beta$  is a feasible **basis** and  $\bar{c}_\eta \geq 0$ ,  $\bar{x}$  and  $\bar{y}$  are **optimal**.

**Proof.** Obvious from the **standard problem** in the form of

$$\begin{aligned} c_\beta^\top A_\beta^{-1} b + \min \quad & (c_\eta^\top - c_\beta^\top A_\beta^{-1} A_\eta) x_\eta \\ & A_\beta^{-1} A_\eta x_\eta \leq A_\beta^{-1} b \\ & x_\beta \geq 0, x_\eta \geq 0. \end{aligned}$$

■

**Note.** The order of the arguments in text book for **Theorem 4.1.1** is slightly different.

### 4.1.2 Naive Algorithmic Approach

We see that from the above discussion, we can come up with the following algorithmic approach to find the optimal solution  $\bar{x}$  and  $\bar{y}$  given a **standard form** LP.

1. Start with a **basis partition**  $\beta, \eta$  with  $\bar{x}_\beta \geq 0$ .
2. If  $\bar{c}_\eta \geq 0$ , then  $\bar{x}$  and  $\bar{y}$  are **optimal** and *STOP*.
3. Otherwise, choose  $\eta_j$  with  $\bar{c}_{\eta_j} < 0$ . Consider the associated **basis direction**  $\bar{z}$ . (Idea:  $\bar{x} \rightarrow \bar{x} + \lambda \bar{z}$  with  $\lambda > 0$ ) Then

$$c^\top (\bar{x} + \lambda \bar{z}) = c^\top \bar{x} + \lambda c^\top \bar{z} = c^\top \bar{x} + \lambda \bar{c}_{\eta_j},$$

where

- $c^\top \bar{x}$  is the current objective value
- $c^\top \bar{z}$  is

$$\begin{aligned} c^\top \bar{z} &= c_\eta^\top \bar{z}_\eta + c_\beta^\top \bar{z}_\beta \\ &= c_\eta^\top e_j - c_\beta^\top (A_\beta^{-1} A_{\eta_j}) \\ &= c_{\eta_j} - c_\beta^\top A_\beta^{-1} A_{\eta_j} \\ &= \bar{c}_{\eta_j} \end{aligned}$$

- $\lambda \bar{c}_{\eta_j}$  is the *rate* of change of objective value as we move in direction  $\bar{z}$ .

Then we move from  $\bar{x}$  to  $\bar{x} + \lambda \bar{z}$ , where we let  $\lambda$  as large as possible. Operationally, since we need

$$\bar{x}_\beta + \lambda \bar{z}_\beta \geq 0,$$

where  $\bar{z}_\eta = e_j$ ,  $\bar{z}_\beta = -A_\beta^{-1}A_{\eta_j}$ . We then have

$$\begin{aligned}\bar{x}_{\beta_i} - \lambda \bar{a}_{i,\eta_j} &\geq 0, \text{ for } i = 1, \dots, m \\ \lambda &\leq \frac{\bar{x}_{\beta_i}}{\bar{a}_{i,\eta_j}}, \quad \text{for } i \text{ such that } \bar{a}_{i,\eta_j} > 0.\end{aligned}$$

Hence,

$$\bar{\lambda} := \min_{i: \bar{a}_{i,\eta_j} > 0} \left\{ \frac{\bar{x}_{\beta_i}}{\bar{a}_{i,\eta_j}} \right\} \geq 0.$$

**Remark.** If  $\bar{a}_{i,\eta_j} \leq 0$  for all  $i = 1, \dots, m$ , namely

$$\bar{A}_{i,\eta_j} \leq 0 \Leftrightarrow -A_\beta^{-1}A_{\eta_j} \geq 0 \Leftrightarrow \bar{z} \geq 0,$$

then  $\bar{z}$  is a ray. This means  $(P)$  is unbounded below, hence we *STOP*.

### 4.1.3 Worry-Free Simplex Algorithm

Now, we give the very first version of the simplex algorithm called **worry-free simplex algorithm**. Consider the **standard form** problem

$$\begin{aligned}\min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0.\end{aligned}$$

---

#### Algorithm 4.1: Worry-Free Simplex Algorithm

---

**Data:** **standard form**  $(P)$ , **basic partition**  $\beta, \eta$  with  $x_\beta \geq 0$

**Result:** **optimal solutions**  $\bar{x}, \bar{y}$  or report  $(P)$  is unbounded

```

1 while True do
2    $\bar{x}_\beta \leftarrow A_\beta^{-1}b (\geq 0)$ 
3    $\bar{c}_\eta^\top \leftarrow c_\eta^\top - c_\beta^\top A_\beta^{-1}A_\eta$ 
4   if  $\bar{c}_\eta \geq 0$  then                                     //  $\bar{x}$  is optimal for  $(P)$ 
5      $\bar{y} \leftarrow c_\beta^\top A_\beta^{-1}$                              // From Theorem 4.1.1
6     return  $\bar{x}, \bar{y}$ 
7   else                                                  // Basic direction  $\bar{z}$ , then  $c^\top \bar{z} = \bar{c}_{\eta_j} < 0$ 
8     choose  $j$  where  $1 \leq j \leq n - m$  such that  $\bar{c}_{\eta_j} < 0$ 
9     if  $\bar{A}_{\eta_j} \leq 0$  then                                //  $\bar{A}_{\eta_j} \leq 0 \Rightarrow (P)$  is unbounded
10      return  $(P)$  is unbounded
11     $\lambda \leftarrow \min_{i: \bar{a}_{i,\eta_j} > 0} \left\{ \frac{\bar{x}_{\beta_i}}{\bar{a}_{i,\eta_j}} \right\}$  // Largest choice so that  $\bar{x} + \lambda \bar{z} \geq 0$ 
12     $\bar{x} \leftarrow \bar{x} + \lambda \bar{z}$ 
13    Redetermine  $\beta, \eta$ 
```

---

**Remark.** Note that  $x$  is assumed to be a **basic feasible solution**.

**Problem 4.1.1.** The problem is that is  $\bar{x} + \lambda \bar{z}$  still a **basic solution**? And if it is, what is the **basic partition** that goes with it?

**Answer.** We see that after one iteration, one of the **basic** index  $i^*$  will become **non-basic**, namely

$$(\bar{x} + \lambda \bar{z})_{\beta_{i^*}} = 0;$$

while one of the **non-basic** index will need to become **basic**, since

$$(\bar{x} + \lambda \bar{z})_{\beta_{i^*}} = \lambda \bar{e}_j.$$

Namely,

	$\bar{x}$	$\bar{z}$	$\bar{x} + \lambda \bar{z}$	
$\beta_{i^*} \rightarrow \beta$	$\bar{x}_{\beta}$	$\bar{z}_{\beta}$	$\rightarrow 0$	$\beta_{i^*}$ becomes <i>non-basic</i>
$\eta$	$\bar{x}_{\eta} = 0$	$\bar{x}_{\eta} = e_j$	$\lambda \bar{e}_j$	$\eta_j$ becomes <i>basic</i>

⊛

Now, suppose  $i^*$  is that chosen index, which means  $\bar{a}_{i^*\eta_j} > 0$  and  $\frac{\bar{x}_{\beta_{i^*}}}{\bar{a}_{i^*\eta_j}} = \bar{\lambda}$ . Then we have  $\beta_{i^*}$  such that

$$\bar{x} + \lambda \bar{z} \Rightarrow \bar{x}_{\beta_{i^*}} + \bar{\lambda} \bar{z}_{\beta_{i^*}} = \bar{x}_{\beta_{i^*}} + \frac{\bar{x}_{\beta_{i^*}}}{\bar{a}_{i^*\eta_j}} (-\bar{a}_{i^*\eta_j}) = 0.$$

So we reasonably suspect that there is a new *basic partition* such that

$$\begin{aligned} \tilde{\beta} &:= (\beta_1, \beta_2, \dots, \beta_{i^*-1}, \eta_j, \beta_{i^*+1}, \dots, \beta_m) \\ &\quad \updownarrow \\ \tilde{\eta} &:= (\eta_1, \eta_2, \dots, \eta_{j-1}, \beta_{i^*}, \eta_{j+1}, \dots, \eta_{n-m}). \end{aligned}$$

The remaining question is that, is  $A_{\tilde{\beta}}$  still invertible? Namely, is  $\det(A_{\tilde{\beta}}) \neq 0$ ?

**Lemma 4.1.3.** After one iteration of *worry-free simplex algorithm*,  $A_{\tilde{\beta}}$  is still invertible.

**Proof.** We see that  $A_{\tilde{\beta}}$  is invertible if and only if  $A_{\tilde{\beta}}^{-1} A_{\tilde{\beta}}$  is invertible. And since

$$A_{\tilde{\beta}}^{-1} A_{\tilde{\beta}} = [e_1 \ e_2 \ \dots \ e_{i^*-1} \ \bar{A}_{\eta_j} \ e_{i^*+1} \ \dots \ e_m],$$

and since  $\det(A_{\tilde{\beta}}^{-1} A_{\tilde{\beta}}) = \bar{a}_{i^*\eta_j}$ , if  $\bar{a}_{i^*\eta_j} \neq 0$ , then we see that this is indeed invertible. But this is a obvious fact by our choice of  $i^*$ . ■

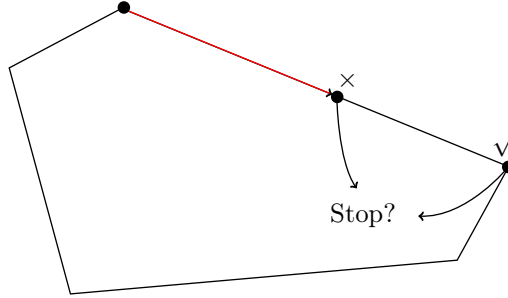


Figure 4.1: Pivot Swap in terms of *feasible region*.

Finally, we check that the unique *basic solution* for this *basic partition*  $\tilde{\beta}, \tilde{\eta}$  are exactly  $\bar{x} + \bar{\lambda} \bar{z}$ .

**Lemma 4.1.4.** The unique solution of  $Ax = b$  having  $x_{\tilde{\eta}} = 0$  is  $\bar{x} + \bar{\lambda} \bar{z}$ .

**Proof.** Firstly,  $(\bar{x} + \bar{\lambda} \bar{z})_j = 0$  for  $j \in \tilde{\eta}$ . Moreover,  $\bar{x} + \bar{\lambda} \bar{z}$  is the unique solution to  $Ax = b$  having  $x_{\tilde{\eta}} = 0$  because  $A_{\tilde{\beta}}$  is invertible, namely

$$Ax = b \Rightarrow \underbrace{A_{\tilde{\eta}} x_{\tilde{\eta}}}_{=0} + A_{\tilde{\beta}} x_{\tilde{\beta}} = b \Rightarrow x_{\tilde{\beta}} = A_{\tilde{\beta}}^{-1} b.$$

■

## Lecture 8: Simplex Algorithm

### 4.2 Remaining Problem

27 Sep. 08:00

**Problem 4.2.1.** How do we start with a [basic](#) feasible partition?

**Answer.** We consider the so-called [phase one problem](#). ⊗

#### 4.2.1 Phase one problem

**Definition 4.2.1** (Phase one problem). Given the [primal](#)  $(P)$ , the so-called *phase one problem*, denoted as  $(\Phi)$  is defined as follows.

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \qquad \begin{array}{ll} \min & x_{n+1} \\ & Ax + A_{n+1}x_{n+1} = b \\ (\Phi) & x \geq 0, x_{n+1} \geq 0. \end{array}$$

**Remark.** We see that

1. If min value of  $x_{n+1}$  in  $(\Phi)$  is 0, then we get a [feasible solution](#) of  $(P)$ .
2. If min value of  $x_{n+1}$  in  $(\Phi)$  is  $> 0$ , then there is no [feasible solution](#) of  $(P)$ .

We see that by solving  $(\Phi)$ , we will get a [feasible solution](#) for  $(P)$  or determine whether  $(P)$  is solvable in the first place. But to solve the linear program  $(\Phi)$ , we're facing the same problem as  $(P)$ ...

**Problem 4.2.2.** How do we get an initial [basic feasible solution](#) for  $(\Phi)$ ?

**Answer.** Thankfully, in this case, we know how to get a [basic feasible solution](#) for  $(\Phi)$ .

1. Start with a [basic solution](#) of  $(P)$ ,  $\tilde{\beta}, \tilde{\eta}$  is the [basic partition](#).
2. If  $\bar{x}_{\tilde{\beta}}$  is [feasible](#) then we just use  $\tilde{\beta}$  and  $\tilde{\eta}$  for  $\beta$  and  $\eta$ .
3. Otherwise, set  $A_{n+1} = -A_{\tilde{\beta}}^{-1}\tilde{1}$ . If  $\eta_j = n+1$

$$\bar{z} : \bar{z}_{\tilde{\eta}} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}, \quad \bar{z}_{\tilde{\beta}} := -A_{\tilde{\beta}}^{-1}(A_{n+1}) = \tilde{1}$$

and

$$\vec{x} \rightarrow \vec{x} + \lambda \vec{z} \geq \vec{0}.$$

**Example.**

$$\vec{x}_{\tilde{\beta}} + \lambda \vec{z}_{\tilde{\beta}} = \begin{pmatrix} 7 \\ 0 \\ 3 \\ -5 \\ 6 \\ -8 \end{pmatrix} + \lambda \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

then

$$i^* = \arg \min_{i: \bar{x}_{\tilde{\beta}} < 0} \{-\bar{x}_{\tilde{\beta}}\}.$$

⊗

**Problem 4.2.3.** What if  $x_{n+1} = 0$ ?

**Answer.** Just stop right before  $x_{n+1} = 0$ , let other variable do that.

⊗

### 4.2.2 Perturbed Problem

Though we now know how to get [basic feasible solution](#) for  $(P)$  to start our [worry-free simplex algorithm](#), we have one more problem:

**Problem 4.2.4 (Degenerate problem).** What if  $\lambda = 0$ , i.e.  $x_{\beta_i} = 0$  for some  $i$ ?

We see that we'll need to introduce so-called [non-degeneracy hypothesis](#).

**Definition 4.2.2 (Non-degeneracy hypothesis).**  $x_{\beta_i} > 0$  for all  $i$  at every iteration of [worry-free simplex algorithm](#).

**Remark (Termination analysis).** With [non-degeneracy hypothesis](#), we see that [worry-free simplex algorithm](#) will certainly terminate.

**Proof.** We have

$$\begin{aligned} \vec{x}_{\beta_i} > 0 \text{ for all } i \text{ at every iteration} &\Rightarrow \bar{\lambda} \neq 0 \\ &\Rightarrow \text{objective value decrease at each iteration.} \\ &\Rightarrow \text{algorithm must terminate} \end{aligned}$$

because there are only finitely many bases.

⊗

Though the [non-degeneracy hypothesis](#) helps avoid the mess, but since we want to be able to solve any [general LP](#), hence we now try to avoid using this hypothesis. We first consider the following problem called [perturbed problem](#).

**Definition 4.2.3 (Perturbed problem).** Given a [standard form](#) problem, the following induced problem is called *perturbed problem*.

$$\begin{aligned} \min \quad & c^\top x \\ Ax = b + B & \begin{pmatrix} \epsilon \\ \epsilon^2 \\ \epsilon^3 \\ \vdots \\ \epsilon^m \end{pmatrix} \\ x \geq 0 \end{aligned}$$

where  $\epsilon$  is an arbitrarily small *indeterminate*.

**Remark.** Note that  $\epsilon \neq 0$ .

**Note.** We see that

$$\vec{x}_\beta = A_\beta^{-1} \left( b + B \begin{pmatrix} \epsilon \\ \epsilon^2 \\ \vdots \\ \epsilon^m \end{pmatrix} \right) = A_\beta^{-1} b + A_\beta^{-1} B \begin{pmatrix} \epsilon \\ \epsilon^2 \\ \vdots \\ \epsilon^m \end{pmatrix},$$

which is just a **polynomial in  $\epsilon$** .

**Definition 4.2.4** (Polynomial in  $\epsilon$ ). We denote polynomials in  $\epsilon$  as

$$p(\epsilon) = p_0 + p_1\epsilon + p_2\epsilon^2 + \cdots + p_m\epsilon^m,$$

where  $p_i \in \mathbb{R}$ .

Which suggest the following definitions.

**Definition 4.2.5** (Sign of polynomial in  $\epsilon$ ). Let  $K$  be the minimal index with  $p_K \neq 0$ .

- If  $p_K < 0$ , then  $p(\epsilon) < 0$
- If  $p_K > 0$ , then  $p(\epsilon) > 0$
- If  $p_K = 0$ , namely  $p_0 = p_1 = \cdots = p_m = 0$ , then  $p(\epsilon) = 0$

**Note.** Given

$$\begin{aligned} p(\epsilon) &= p_0 + p_1\epsilon + p_2\epsilon^2 + \cdots + p_m\epsilon^m \\ q(\epsilon) &= q_0 + q_1\epsilon + q_2\epsilon^2 + \cdots + q_m\epsilon^m \end{aligned}$$

with  $K_p$  and  $K_q$ . Then  $K_{p+q}$  depends on  $K_p$  and  $K_q$ . We then see that if  $p(\epsilon) - q(\epsilon) \geq 0$ , then  $p(\epsilon) \geq q(\epsilon)$ .

**Problem 4.2.5.** Where does this  $\epsilon$  thing links with the **worry-free simplex algorithm**, and how can it solve the **degenerate problem**?

**Answer.** Suppose

$$\overbrace{p(\epsilon)}^{\text{value of some basic variable}} = p_0 + p_1\epsilon + p_2\epsilon^2 + \cdots + p_m\epsilon^m.$$

Feasibility for the **perturbed problem** means  $p(\epsilon) \geq \vec{0} \Rightarrow p(0) = p_0 \geq 0$ .

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b + B\vec{\epsilon} \\ & x \geq 0. \end{aligned}$$

Find an initial feasible basis  $\beta, \eta$  for unperturbed problem,  $B := A_\beta$ ,

$$\vec{x}_\beta = A_\beta^{-1}(b + A_\beta\vec{\epsilon}) = \underbrace{A_\beta^{-1}b}_{\geq \vec{0}} + \vec{\epsilon} = \vec{x}_\beta + \begin{pmatrix} \epsilon \\ \epsilon^2 \\ \epsilon^3 \\ \vdots \\ \epsilon^m \end{pmatrix} = \begin{pmatrix} \vec{x}_{\beta_1} + \epsilon \\ \vec{x}_{\beta_2} + \epsilon^2 \\ \vdots \\ \vec{x}_{\beta_m} + \epsilon^m \end{pmatrix} \geq \vec{0}.$$

**Claim.** **Perturbed problem** is non-degenerate.

**Proof.** This is equivalent to show that there are no  $i$  in the later basis  $\tilde{\beta}$  such that  $\vec{x}_{\tilde{\beta}_i} = 0$ . Suppose there is an  $i$  such that  $\vec{x}_{\tilde{\beta}_i} = 0$ . But since

$$\vec{x}_{\tilde{\beta}} := A_{\tilde{\beta}}^{-1}(b + A_{\beta}\vec{\epsilon}) = A_{\tilde{\beta}}^{-1}b + A_{\tilde{\beta}}^{-1}A_{\beta}\vec{\epsilon},$$

if  $\vec{x}_{\tilde{\beta}_i} = 0$ , we must have

$$i^{\text{th}} \text{ element of } A_{\tilde{\beta}}^{-1}A_{\beta} \begin{pmatrix} \epsilon \\ \epsilon^2 \\ \epsilon^3 \\ \vdots \\ \epsilon^m \end{pmatrix} = 0 \Rightarrow \langle i^{\text{th}} \text{ row of } A_{\tilde{\beta}}^{-1}A_{\beta}, \vec{\epsilon} \rangle = 0$$

$$\Rightarrow i^{\text{th}} \text{ row of } A_{\tilde{\beta}}^{-1}A_{\beta} = \vec{0} \nmid$$

because  $A_{\tilde{\beta}}^{-1}A_{\beta}$  is invertible where  $A_{\tilde{\beta}}^{-1}A_{\tilde{\beta}}$  is its inverse. ⊗

⊗

## Lecture 9: Practical Simplex Algorithm

**Note** ( $A_{\tilde{\beta}}^{-1}$  in reality). In reality, we don't really calculate  $A_{\tilde{\beta}}^{-1}$ , since when calculating

$$A_{\beta}x_{\beta} = b,$$

we do not use

$$\bar{x}_{\beta} = A_{\beta}^{-1}b,$$

instead, we use *LU-Factorization*.<sup>a</sup> And since after applying pivot change, there is only a column change in  $A_{\beta}^{-1}$ , we can use the previous result to calculate the new  $\bar{x}_{\beta}$  much faster.

<sup>a</sup>[https://en.wikipedia.org/wiki/LU\\_decomposition](https://en.wikipedia.org/wiki/LU_decomposition)

29 Sep. 08:00

### 4.3 The Word *Simplex*

For a **standard form** problem

$$\begin{aligned} \min \quad & c^{\top}x \\ & Ax = b \\ (P) \quad & x \geq 0, \end{aligned}$$

we can instead consider an equivalent problem formulated as

$$\begin{aligned} \min \quad & z \\ & z - c^{\top}x = 0 \Leftrightarrow (c^{\top}x = z) \\ & Ax = b \\ & x \geq 0. \end{aligned}$$

**As previously seen.** Our picture is in  $\mathbb{R}^{n-m}$ , but we consider *Dantzig picture*, which is in  $\mathbb{R}^{m+1}$

#### 4.3.1 Column geometry

Plot columns:

$$\underbrace{\begin{pmatrix} c_1 \\ A_1 \end{pmatrix} \begin{pmatrix} c_2 \\ A_2 \end{pmatrix} \cdots \begin{pmatrix} c_n \\ A_n \end{pmatrix}}_{n \text{ points in } \mathbb{R}^{m+1}}$$



The requirement line is

$$\begin{pmatrix} z \\ b \end{pmatrix}.$$

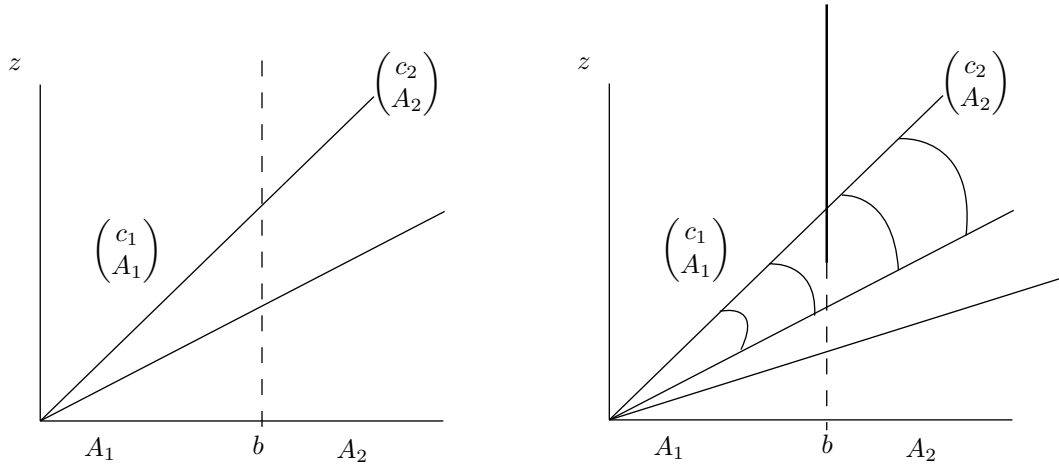


Figure 4.2: Column Geometry.

### 4.3.2 Simplices (Plural of Simplex)

**Example (Simplex).** An  $n-1$  dimensional simplex in  $\mathbb{R}^n$  with  $n$  standard unit vectors are the corner can be described as

$$\left\{ x \in \mathbb{R}^n : \sum_{i=1}^n x_i = 1, x_i \geq 0 \right\}.$$



Figure 4.3: Simplex.

**Note.**  $m+1$  points of a simplex of dimension  $m$ .

A simplicial cone is rather simple, the graph below is informative enough.

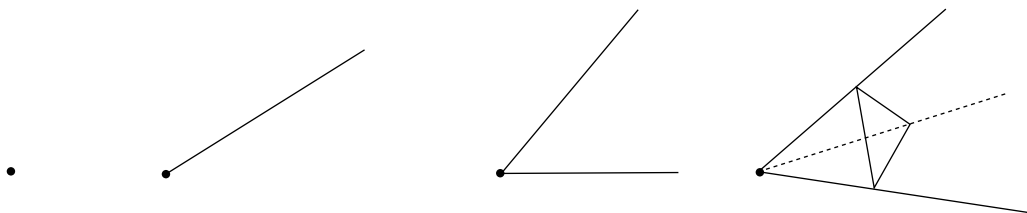


Figure 4.4: Simplicial Cones

## 4.4 Complete Simplex Algorithm

Now, given a [standard form](#) problem  $(P)$ :

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0, \end{aligned}$$

we can then solve  $(P)$  in a mathematical rigorous and complete way.

---

**Algorithm 4.2:** Simplex Algorithm
 

---

**Data:** [standard form](#) LP  $(P)$   
**Result:** [optimal solutions](#)  $\bar{x}, \bar{y}$ , or report  $(P)$  is unbounded/infeasible

```

1  $(\Phi_\epsilon) \leftarrow$  algebraic perturbation to the phase one problem  $(\Phi)$ 
2  $\beta \leftarrow$  basic feasible solution for  $(\Phi_\epsilon)$ 
3 result  $\leftarrow$  WorryFreeSimplexAlgorithm $((\Phi_\epsilon), \beta)$ a
4 if result = unbounded then
5   | return  $(P)$  has no solution
6 else
7   |  $\beta \leftarrow$  result // Retrieve the feasible basis for  $(P)$ 
8
9  $(P_\epsilon) \leftarrow$  algebraic perturbation to  $(P)$ 
10 return WorryFreeSimplexAlgorithm $((P_\epsilon), \beta)$ b
```

---

<sup>a</sup>Adapted to [algebraically perturbed problems](#), and always giving preference to  $x_{n+1}$  for leaving the [basis](#) whenever it's eligible to leave for the unperturbed problem.

<sup>b</sup>Again, adapted to [algebraically perturbed problems](#).

# Chapter 5

## Duality

Consider the [standard problem](#) and its [dual](#)

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \quad \begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top. \\ (D) & \end{array}$$

### 5.1 Strong Duality Theorem

**As previously seen.** The [weak duality theorem](#) states that if  $\hat{x}$  is [feasible](#) for  $(P)$ , and  $\hat{y}$  is [feasible](#) for  $(D)$ , then

$$c^\top \hat{x} \geq \hat{y}^\top b.$$

Moreover, the equality holds if and only if  $\hat{x}$  and  $\hat{y}$  are [optimal](#).

Also, the [weak optimal basis theorem](#) states that if we have a [basic partition](#)  $\beta, \eta$ , and we also have  $\bar{x}_\beta \geq \bar{0}$  ( $\bar{x}$  is [feasible](#) for  $(P)$ ) and  $\bar{c}_\eta \geq \bar{0}$  ( $\bar{y}$  is [feasible](#) for  $(D)$ ), then  $\bar{x}$  and  $\bar{y}$  are both [optimal](#).

Now, we have so-called [strong optimal basis theorem](#).

**Theorem 5.1.1** (Strong optimal basis theorem). If  $(P)$  has a [feasible solution](#), and if  $(P)$  is not unbounded, then there exist a [basic partition](#)  $\beta, \eta$  such that  $\bar{x}$  and  $\bar{y}$  are [optimal](#), and

$$c^\top \bar{x} = \bar{y}^\top b.$$

**Proof.** Since if  $(P)$  has a [feasible solution](#) and is not unbounded, we can just run the [simplex algorithm](#), which will terminate with a basis  $\beta$  such that the associated [basic solution](#)  $\bar{x}$  and the associated [dual solution](#)  $\bar{y}$  are [optimal](#). ■

We see that this leads to another similar result.

**Theorem 5.1.2** (Strong duality theorem). If  $(P)$  has a [feasible solution](#) and  $(P)$  is not unbounded, then there exist [optimal solutions](#)  $\hat{x}$  and  $\hat{y}$  with

$$c^\top \hat{x} = \hat{y}^\top b.$$

**Note.** The proof of these two theorems are by directly using the *mathematical complete* version of [simplex algorithm](#), hence the completeness of [simplex algorithm](#) (namely the [phase one problem](#) and the [perturbation](#)) is important.

<i>Simplex Algorithm</i>	$(P) \setminus (D)$	optimal solution	infeasible	unbounded
$\bar{c}_\eta \geq \vec{0} \Rightarrow \text{Stop}$	optimal solution	✓	×	×
optimal $x_{n+1}$ in $\Phi$ is positive	infeasible	×	✓	✓
$\bar{A}_{\eta_j} \leq \vec{0} \Rightarrow \text{Stop}$	unbounded	×	✓	×

Table 5.1: Comparison between  $(P)$  and  $(D)$ 

## Lecture 10: Complementary

### 5.2 Complementary

4 Oct. 08:00

**Definition 5.2.1** (Complementary). Solutions  $\hat{x}$  to  $(P)$  and  $\hat{y}$  to  $(D)$  are *complementary* if

$$m+n \text{ equations } \begin{cases} \underbrace{(c_j - \hat{y}^\top A_{\cdot j})}_{=0 \text{ for } j \in \beta} \underbrace{\hat{x}_j}_{=0 \text{ for } j \in \eta} = 0, & j = 1 \cdots n; \\ \hat{y}_i \underbrace{(A_{i \cdot} \hat{x} - b_i)}_{=0 \text{ for } \bar{x}} = 0, & i = 1 \cdots m. \end{cases}$$

Now, suppose we have a **basic partition**  $\beta, \eta$  such that

$$\begin{aligned} \bar{x}: \bar{x}_\beta &= A_\beta^{-1} b, \quad \bar{x}_\eta = \vec{0} \\ \bar{y}: \bar{y}^\top &= c_\beta^\top A_\beta^{-1}. \end{aligned}$$

**Note.** Specifically, we see that  $c_j - \hat{y}^\top A_{\cdot j} = 0$  for  $j \in \beta$  is because  $\bar{y}^\top = c_\beta^\top A_\beta^{-1}$ , and then

$$c_j - \hat{y}^\top A_{\cdot j} = c_j - c_\beta^\top \underbrace{A_\beta^{-1} A_{\cdot j}}_{e_j} = c_j - c_j = 0.$$

Then just from above, we see that the following theorems hold.

**Theorem 5.2.1.** If  $\bar{x}$  and  $\bar{y}$  are **basic solutions** for  $\beta, \eta$ , then  $\bar{x}$  and  $\bar{y}$  are **complementary**.

**Theorem 5.2.2** (Complementary with equal objective value). If  $\hat{x}$  and  $\hat{y}$  are **complementary**, then

$$c^\top \hat{x} = \hat{y}^\top b.$$

**Note.**

$$c_\beta^\top A_\beta^{-1} b = \bar{y}^\top b, \quad c^\top (A_\beta^{-1} b) = c_\beta^\top \bar{x}_\beta = c^\top \bar{x}.$$

**Proof.** We show that

$$c^\top \hat{x} - \hat{y}^\top b = 0.$$

We have

$$\begin{aligned}
 c^\top \hat{x} - \hat{y}^\top b &= (c^\top - \underbrace{\hat{y}^\top A}_{\text{added terms}}) \hat{x} + \hat{y}^\top (A\hat{x} - b) \\
 &= \sum_{j=1}^n \underbrace{(c_j - \hat{y}^\top A_{\cdot j})}_{=0 \text{ for } i=1\dots n} x_j + \sum_{i=1}^m \underbrace{\hat{y}_i (A_{i\cdot} \hat{x} - b_i)}_{=0 \text{ for } i=1\dots m} \\
 &= 0.
 \end{aligned}$$

⊛

**Theorem 5.2.3** (Weak complementary slackness theorem). If  $\hat{x}$  and  $\hat{y}$  are **feasible** and **complementary**, then they are **optimal**.

**Proof.** Follows from [Theorem 1.2.1](#) and **complementary solutions** having equal objective value from [Theorem 5.2.2](#). ■

**Theorem 5.2.4** (Strong complementary slackness theorem). If  $\hat{x}$  and  $\hat{y}$  are **optimal**, then  $\hat{x}$  and  $\hat{y}$  are **complementary**.

**Proof.** Recall that

$$\underbrace{\sum_{j=1}^n \underbrace{(c_j - \hat{y}^\top A_{\cdot j})}_{\geq 0 \text{ for each } j} \underbrace{\hat{x}_j}_{\geq 0 \text{ for each } j}}_{\geq 0} + \underbrace{\sum_{i=1}^m \underbrace{\hat{y}_i (A_{i\cdot} \hat{x} - b_i)}_{=0 \text{ for each } i}}_{=0} = 0 = c^\top \hat{x} - \hat{y}^\top b$$

if  $\hat{x}$  and  $\hat{y}$  are **optimal**:  
same object value

Hence, the equality can only hold if

$$(c_j - \hat{y}^\top A_{\cdot j}) \hat{x}_j = 0, \text{ for } j = 1, 2, \dots, n;$$

with the obvious fact that

$$\hat{y}_i (A_{i\cdot} \hat{x} - b_i) = 0, \text{ for } i = 1, 2, \dots, m,$$

so they are **complementary**. ■

### 5.3 Duality for General Linear Optimization Problems

So far, we only discuss the **dual** of the **standard form** problem. But we will see that *every* linear optimization problem has a natural **dual**.

Now consider a **general linear programming problem**

$$\begin{aligned}
 \min \quad & c_P^\top x_P + c_N^\top x_N + c_U^\top x_U \\
 & A_{GP} x_P + A_{GN} x_N + A_{GU} x_U \geq b_G \\
 & A_{LP} x_P + A_{LN} x_N + A_{LU} x_U \leq b_L \\
 & A_{EP} x_P + A_{EN} x_N + A_{EU} x_U = b_E \\
 (\mathcal{G}) \quad & x_P \geq 0, x_N \leq 0, x_U \text{ unrestricted.}
 \end{aligned}$$

We first turn this into a **standard form** problem:

1.  $\tilde{x}_N := -x_N$ :

$$\begin{aligned}
 \min \quad & c_P^\top x_P + c_N^\top x_N + c_u^\top x_U \\
 & A_{GP} x_P - A_{GN} x_N + A_{GU} x_U \geq b_G \\
 & A_{LP} x_P - A_{LN} x_N + A_{LU} x_U \leq b_L \\
 & A_{EP} x_P - A_{EN} x_N + A_{EU} x_U = b_E \\
 & x_P \geq 0, x_N \leq 0, x_U \text{ unrestricted}
 \end{aligned}$$

2.  $x_U = \tilde{x}_U - \tilde{\tilde{x}}_U$ , where  $\tilde{x}_U, \tilde{\tilde{x}}_U \geq 0$ :

$$\begin{aligned} \min \quad & c_P^\top x_P + c_N^\top x_N + c_U^\top \tilde{x}_U - c_U \tilde{\tilde{x}}_U \\ & A_{GP}x_P - A_{GN}x_N + A_{GU}\tilde{x}_U - A_{GU}\tilde{\tilde{x}}_U \geq b_G \\ & A_{LP}x_P - A_{LN}x_N + A_{LU}\tilde{x}_U - A_{LU}\tilde{\tilde{x}}_U \leq b_L \\ & A_{EP}x_P - A_{EN}x_N + A_{EU}\tilde{x}_U - A_{EU}\tilde{\tilde{x}}_U = b_E \\ & x_P \geq 0, x_N \leq 0, \tilde{x}_U \geq 0, \tilde{\tilde{x}}_U \geq 0 \end{aligned}$$

3. Adding [slack variables](#):

$$\begin{aligned} \min \quad & c_P^\top x_P + c_N^\top x_N + c_U^\top \tilde{x}_U - c_U \tilde{\tilde{x}}_U \\ & A_{GP}x_P - A_{GN}x_N + A_{Gu}\tilde{x}_U - A_{GU}\tilde{\tilde{x}}_U - s_G = b_G \\ & A_{LP}x_P - A_{LN}x_N + A_{Lu}\tilde{x}_U - A_{LU}\tilde{\tilde{x}}_U + t_L = b_L \\ & A_{EP}x_P - A_{EN}x_N + A_{Eu}\tilde{x}_U - A_{EU}\tilde{\tilde{x}}_U = b_E \\ & x_P \geq 0, x_N \leq 0, \tilde{x}_U \geq 0, \tilde{\tilde{x}}_U \geq 0, s_G \geq 0, t_L \geq 0 \end{aligned}$$

With [dual variables](#)  $y_G, y_L, y_E$ , we have

$$\begin{aligned} \max \quad & y_G^\top b_G + y_L^\top b_L + y_E^\top b_E \\ & y_G^\top A_{GP} + y_L^\top A_{LP} + y_E^\top A_{EP} \leq c_P^\top \\ & -y_G^\top A_{GN} - y_L^\top A_{LN} - y_E^\top A_{EN} \leq -c_N^\top \\ & y_G^\top A_{GU} + y_L^\top A_{LU} + y_E^\top A_{EU} \leq c_U^\top \\ & -y_G^\top A_{GU} - y_L^\top A_{LU} - y_E^\top A_{EU} \leq -c_U^\top \\ & y_G^\top \geq 0, y_L^\top \leq 0. \end{aligned}$$

We time  $-1$  to the both side of the second constraint, and we see that last two [structure constraints](#) can be reduced to a single equality, results in

$$\begin{aligned} \max \quad & y_G^\top b_G + y_L^\top b_L + y_E^\top b_E \\ & y_G^\top A_{GP} + y_L^\top A_{LP} + y_E^\top A_{EP} \leq c_P^\top \\ & y_G^\top A_{GN} + y_L^\top A_{LN} + y_E^\top A_{EN} \geq c_N^\top \\ & y_G^\top A_{GU} + y_L^\top A_{LU} + y_E^\top A_{EU} = c_U^\top \\ (\mathcal{H}) \quad & y_G^\top \geq 0, y_L^\top \leq 0. \end{aligned}$$

Finally, we remark that this gives us a simple result as we have already seen before.

**Theorem** (Duality for general LP). We rephrase the weak and strong duality theorem in a more general term.

**Theorem 5.3.1** (Weak duality theorem). If  $(\hat{x}_P, \hat{x}_N, \hat{x}_U)$  is [feasible](#) in  $\mathcal{G}$  and the [dual variables](#)  $(\hat{y}_G, \hat{y}_L, \hat{y}_E)$  is [feasible](#) in  $\mathcal{H}$ , then

$$c_P^\top \hat{x}_P + c_N^\top \hat{x}_N + c_U^\top \hat{x}_U \geq \hat{y}_G^\top b_G + \hat{y}_L^\top b_L + \hat{y}_E^\top b_E.$$

**Theorem 5.3.2** (Strong duality theorem). If  $\mathcal{G}$  has a [feasible solution](#), and  $\mathcal{G}$  is not unbounded, then there exist [feasible solutions](#)  $(\hat{x}_P, \hat{x}_N, \hat{x}_U)$  for  $\mathcal{G}$  and  $(\hat{y}_G, \hat{y}_L, \hat{y}_E)$  for  $\mathcal{H}$  that are [optimal](#). Moreover,

$$c_P^\top \hat{x}_P + c_N^\top \hat{x}_N + c_U^\top \hat{x}_U = \hat{y}_G^\top b_G + \hat{y}_L^\top b_L + \hat{y}_E^\top b_E.$$

**Remark.** We can also rephrase the [Theorem 5.2.3](#) and [Theorem 5.2.4](#) in this setup. The proof follows the same idea, but with some more works.

## Lecture 11: Duality

As previously seen (The production problem). The [primal](#):

6 Oct. 08:00

$$\begin{aligned} \max \quad & c^\top x \\ & Ax \leq b \\ & x \geq \vec{0} \end{aligned}$$

- $n$  products activities
- $c_j$  = per-unit revenue for activity  $j = 1 \dots n$
- $b_i$  = resource endowment for resource  $i = 1 \dots m$
- $a_{ij}$  = amount of resource  $i$  consumed by activity  $j$

$$\begin{aligned} \min \quad & y^\top b \\ & y^\top A \geq \vec{c} \\ & y \geq \vec{0} \end{aligned}$$

where

$$y^\top A_{\cdot j} \geq c_j \left( \sum_{i=1}^m y_i a_{ij} \right) \geq c_j.$$

**Note.** We have

	min	max	
constraints	$\geq$	$\geq 0$	variables
	$\leq$	$\leq 0$	
	$\leq$	unres.	
	$=$		
variables	$\geq 0$	$\leq$	constraints
	$\leq 0$	$\geq$	
	unres.	$=$	

for a general rule to find a [primal's dual](#).

Come back to [complementary](#).

$$\begin{aligned} \hat{y}^\top A_{\cdot j} - c_j \hat{x}_j &= 0 \text{ for } j = 1 \dots n \\ \hat{y}_i (b_i - A_{i \cdot} \hat{x}) &= 0 \text{ for } i = 1 \dots m \end{aligned}$$

**Note.** For [feasible solutions](#) of  $(P)$  and  $(D)$ , at most one of  $\hat{y}^\top A_{\cdot j} - c_j$  and  $\hat{x}_j$  is positive for  $j = 1 \dots n$ ; while at most one of  $b_i - A_{i \cdot} \hat{x}$  and  $\hat{y}_i$  is positive for  $i = 1 \dots m$ ;

**Problem 5.3.1.** We are looking for a way to find out the upper bound of  $c^\top x$  from the [dual](#).

**Answer.** Since

$$c^\top x \underset{?}{\leq} \underbrace{y^\top A}_{\geq c^\top} \underbrace{x}_{\geq \vec{0}} \leq \underbrace{y^\top}_{\geq \vec{0}} b \Leftrightarrow \sum_{i=1}^m y_i \left( \sum_{j=1}^n a_{ij} x_j \right) \leq \sum_{i=1}^m y_i b_i.$$

We want

$$c^\top \leq y^\top A \Rightarrow c^\top x \leq y^\top Ax$$

Now, return to the **standard form** problem, we have

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \quad \begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top \\ (D) & \end{array}$$

with  $y$  unrestricted. Then we have

$$c^\top x \underset{?}{\geq} \underbrace{y^\top A}_{\leq c^\top} \underbrace{x}_{\geq 0} = y^\top b$$

since

$$y^\top Ax \leq c^\top x.$$

⊛

**Intuition.** For a minimization problem, we are just trying to find the lower bound of the **objective function**'s value.

**Example.** Consider the following linear programming problem:

$$\begin{array}{ll} \max & c^\top x + d^\top z \\ & Ax \geq b \\ & Bx - Fz = g \\ & x \leq 0, z \text{ unrestricted} \end{array}$$

Then the **dual** is (with **dual** variables  $y, w$ )

$$\begin{array}{ll} \min & y^\top b + w^\top g \\ & y^\top A + w^\top B \leq c^\top \\ & -w^\top F = d^\top \\ & y \leq 0, w \text{ unrestricted}, \end{array}$$

where we just look up the table for finding the **dual**. Or, we can also find the **dual** from

$$\begin{array}{l} y^\top A + w^\top B \leq c^\top \\ (y^\top A + w^\top B)x \geq c^\top x, \end{array}$$

hence

$$\begin{array}{l} \underbrace{y^\top}_{\leq 0} (Ax \geq b) \\ + w^\top (Bx - Fz = g) \end{array} \quad \begin{array}{l} \hline c^\top x + d^\top z \stackrel{\text{want}}{\leq} \underbrace{y^\top Ax + w^\top Bx - w^\top Fz}_{\underbrace{(y^\top A + w^\top B)x}_{\leq c^\top} \underbrace{x}_{\leq 0} - \underbrace{(w^\top F)z}_{=d^\top}} \stackrel{\text{want}}{\leq} y^\top b + w^\top g \end{array}$$

**Remark.** Think about what if all are equal sign? (both in constraints and variables, namely unrestricted)



## 5.4 Geometrically Understanding of Duality

### 5.4.1 Farkas' Lemma

**Lemma 5.4.1** (Farkas' Lemma). Let (I) and (II) being

$$\begin{aligned} (I) \quad & Ax = b \\ & x \geq 0 \\ (II) \quad & y^\top b > 0 \\ & y^\top A \leq 0 \end{aligned}$$

for any data  $A$  and  $b$ , then exactly one of (I) or (II) has a solution.

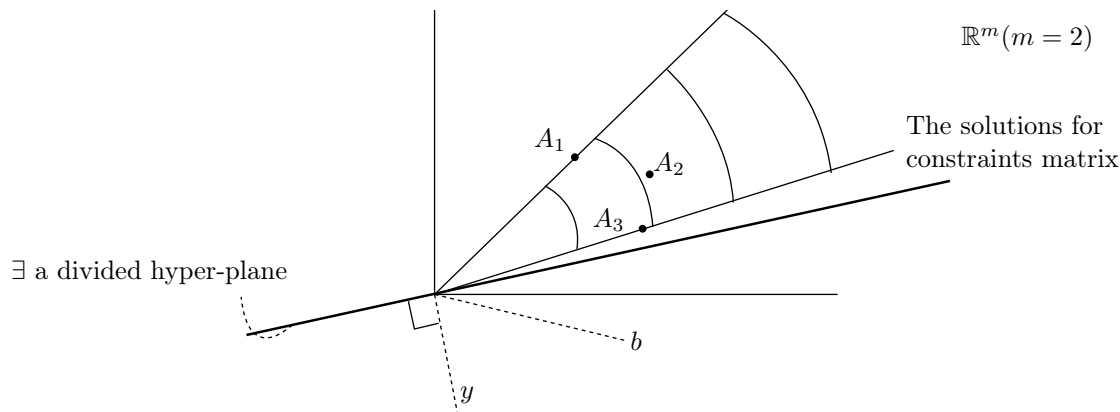


Figure 5.1: Geometrically point of view with  $\mathbb{R}^m, m = 2$

**Intuition.** We outline the idea about the proof.

- Step 1: (I) and (II) can't both have solutions for the same  $A, b$ . Suppose  $\hat{x}$  solves (I) and  $\hat{y}$  solves (II). Then we have

$$0 \geq \underbrace{\hat{y}^\top A}_{\leq \vec{0}} \underbrace{\hat{x}}_{\geq \vec{0}} = \hat{y}^\top b \not\leq 0$$

- Step 2: Show that if (I) has no solution, then (II) has a solution.

## Lecture 12: Farkas' Lemma

Before we prove [Farkas' Lemma](#), we first see something similar. There is a lemma called [Gauss' Lemma](#), 11 Oct. 08:00 which is highly related to [Farkas' Lemma](#).

**Lemma 5.4.2** (Gauss' Lemma). : Exactly one of the following has a solution :

$$\begin{aligned} (I) \quad & Ax = b \\ (II) \quad & y^\top A \geq 0 \\ & y^\top b \neq 0 \end{aligned}$$

**Proof.** This just follows from the Gauss elimination. By doing the elimination, there are two cases :

1. The system has no solution.

2. There is a(some) solution(s).

For second case, it's just  $Ax = b$  is solvable. For the first case, we see that after the elimination, we will have something like

$$\begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} a \end{pmatrix}$$

where  $a \neq 0$ , which just indicates this system is unsolvable. ■

Now we start to proof [Farkas' Lemma](#).

**Proof of Lemma 5.4.1.** As what we have outlined, we divide the proof into two cases.

**Claim.** (I) and (II) can't both have solutions.

**Proof.** Suppose  $\hat{x}$  solves I and  $\hat{y}$  solves (II). Then we have

$$\hat{y}^\top (\hat{A}x = b) \Rightarrow \underbrace{(\hat{y}^\top A)}_{\geq \vec{0}} \underbrace{\hat{x}}_{\geq \vec{0}} = \hat{y}^\top b > 0 \nmid$$

⊗

**Claim.** At least one of (I) or (II) has a solution  $\cong$  If (I) has no solution, then (II) has a solution.

**Proof.** Assume that (I) has no solution, which means that (P) is infeasible with (P) being

$$\begin{aligned} \min \quad & \vec{0}^\top x \\ & Ax = b \\ (P) \quad & x \geq 0. \end{aligned}$$

The dual of this (P) is

$$\begin{aligned} \max \quad & y^\top b \\ (D) \quad & y^\top A \leq \vec{0}^\top. \end{aligned}$$

But this means that (D) is infeasible or unbounded. But we see that (D) can't be infeasible, because  $y = \vec{0}$  is a [feasible solution](#), then we know

$$\begin{aligned} \Rightarrow D \text{ is unbounded} \\ \Rightarrow \text{there exist a feasible solution } \tilde{y} \text{ to } (D) \text{ with positive objective} \end{aligned}$$

⊗

**Remark.** Now, consider  $\lambda \tilde{y}$  (feasible for (D)). Drive to  $+\infty$  by increasing  $\lambda$ . We now see what [Farkas' Lemma](#) really tells us.

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0 \quad \text{feasibility} \\ & \Updownarrow \\ \max \quad & y^\top b \quad \text{unbounded direction} \\ (D) \quad & y^\top A \leq c^\top \end{aligned}$$

Suppose  $\tilde{y}$  is feasible to (D) and suppose  $\hat{y}$  satisfies (II), then

$$(\tilde{y} + \lambda \hat{y})^\top A = \underbrace{\tilde{y}^\top A}_{\leq c^\top} + \underbrace{\lambda}_{>0} \underbrace{\hat{y}^\top A}_{\leq \vec{0}} \leq c^\top.$$

Furthermore, we have

$$(\tilde{y} + \lambda \hat{y})^\top b = \tilde{y}^\top b + \lambda \hat{y}^\top b \Rightarrow \infty \text{ as } \lambda \uparrow.$$

**Example.**

$$\begin{aligned} (I) \quad & Ax \leq b \\ (II) \quad & ? \end{aligned}$$

Find out what (II) is.

**Proof.** We simply set up the (P) and then find its dual.

$$\begin{aligned} \min \quad & \vec{0}^\top x \\ & Ax \leq b \\ (P) \quad & \\ \max \quad & y^\top b \\ & y^\top A = \vec{0}. \\ (D) \quad & y \leq \vec{0} \end{aligned}$$

Then we have

$$\begin{aligned} (I) \quad & Ax \leq b \\ (II) \quad & y^\top A = \vec{0} \\ & y \leq \vec{0} \\ & y^\top b > 0 \end{aligned}$$

Check:

$$0 = \underbrace{\hat{y}^\top A}_{=\vec{0}} \hat{x} \geq \underbrace{\hat{y}^\top b}_{\hat{y} \leq \vec{0}} > 0 \nmid$$

or,

$$\begin{aligned} Ax &\stackrel{y \leq \vec{0}}{\leq} b \quad (y^\top b > 0) \\ 0 &\stackrel{?}{\geq} \underbrace{y^\top A}_{=\vec{0}} x \geq y^\top b > 0 \nmid \end{aligned}$$

⊛

Let's look at another example.

**Example.**

$$\begin{aligned} & (\min \quad \vec{0}^\top x + \vec{0}^\top w) \\ & \quad A x + B w = b \\ & \quad -F w \geq f \\ (I) \quad & x \geq 0, w \text{ unrestricted} \end{aligned}$$

with the dual variables  $y, w$ , we have

$$\begin{aligned} & (\text{Suppose (I) has no solution.}) \\ & \max \quad y^\top b + v^\top b (> 0) \\ & \quad y^\top A \leq \vec{0} \\ (II) \quad & y^\top B - v^\top F = \vec{0} \end{aligned}$$

with  $y$  unrestricted,  $v \geq \vec{0}$ .

Now, we should have a general picture about what [Farkas' Lemma](#) really means. For conditions (I)

and (II), we have

$$\begin{aligned}
 (I) \quad & Ax = b \\
 & x \geq 0 \quad \Leftrightarrow b \text{ is in the cone } K \\
 (II) \quad & y^\top b > 0 \quad \Leftrightarrow y \text{ makes an acute angle with } b. \\
 & y^\top A \leq 0^\top \quad y \text{ makes a non-acute angle with all columns of } A
 \end{aligned}$$

Suppose  $\hat{z}$  in  $K$ , then

$$\hat{z} = A\hat{x} \text{ for some } \hat{x} \geq \vec{0}.$$

Then we have

$$y^\top \hat{z} = \underbrace{y^\top A}_{\leq 0^\top} \underbrace{\vec{x}}_{\geq \vec{0}} \leq 0.$$

We see that  $y$  makes a non-acute angle with everything in  $K$ . Now, suppose  $\hat{y}$  solves (II). Consider

$$\underbrace{\hat{y}^\top}_{\text{numbers}} \underbrace{z}_{\text{variables}} = 0.$$

Now, we have the hyperplane:  $\{z: \hat{y}^\top z = 0\}$  separates  $b$  and  $K$ .

$\mathbb{R}^m (m = 2)$

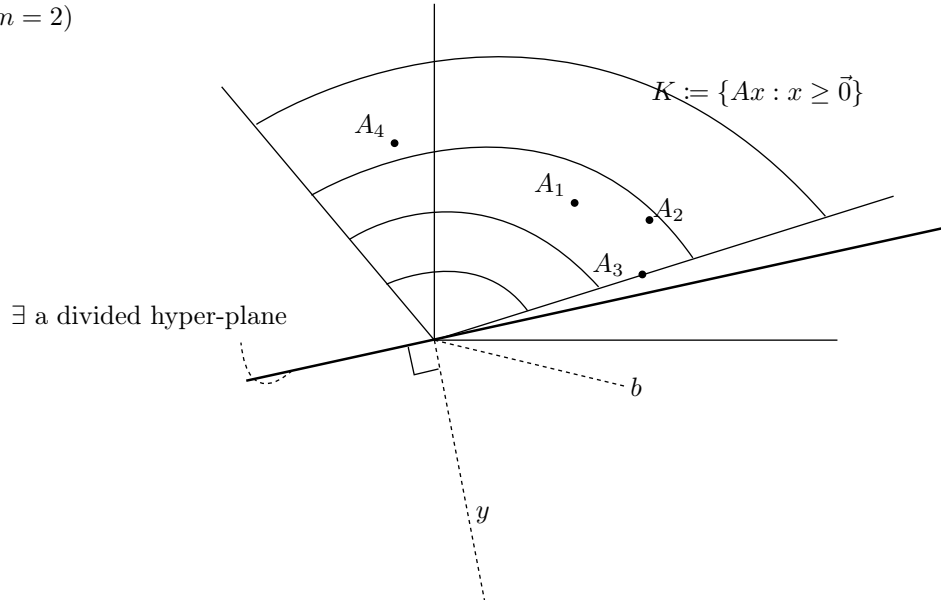


Figure 5.2: Case (II) of the [Farkas' Lemma](#) with  $m = 2$

## 5.5 The Big Picture of Cones

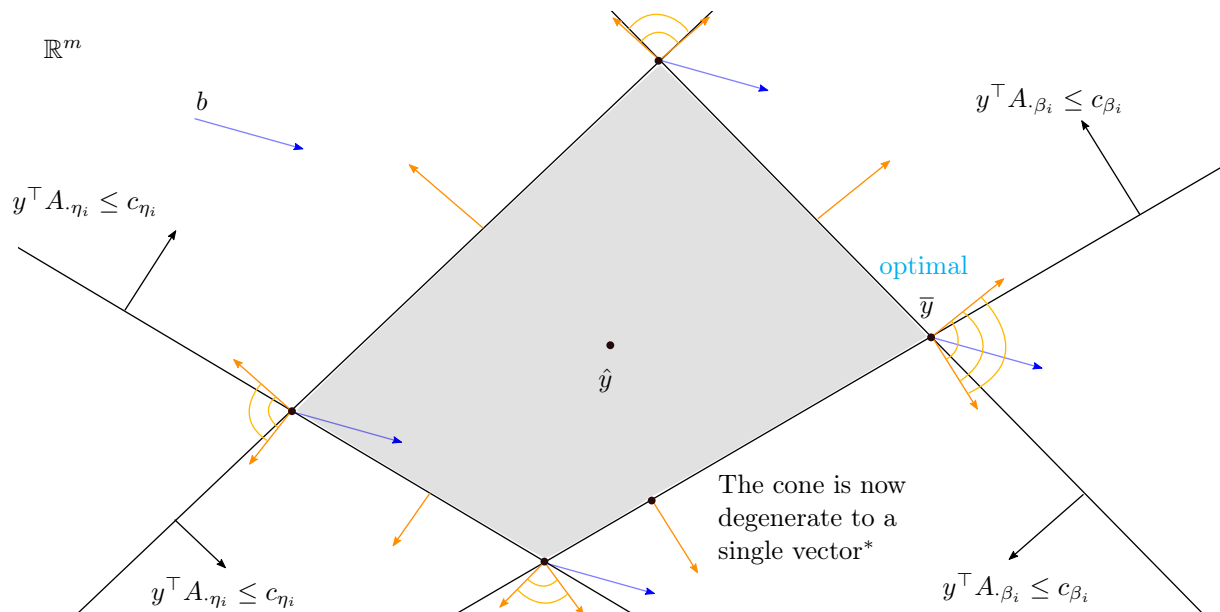
Consider the linear programming problem

$$\begin{aligned}
 \max \quad & y^\top b \\
 \text{s.t.} \quad & y^\top A \leq c^\top
 \end{aligned}$$

with the [partition](#)  $\beta, \eta$ , we see that

$$y^\top A \leq c^\top \Rightarrow \begin{cases} y^\top A_\beta & \leq c_\beta^\top \\ y^\top A_\eta & \leq c_\eta^\top \end{cases}.$$

By solving only for  $\beta$ , then we have  $\bar{y}^\top = c_\beta^\top A_\beta^{-1}$ . And then, by considering the cones, we have

Figure 5.3: Optimality of Cones<sup>1</sup>

with

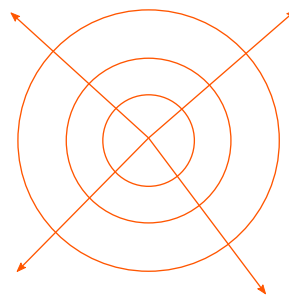


Figure 5.4: Cones join together

**Note.** Consider  $b = \vec{0}(\hat{y})$ . It's in every cone  $\Rightarrow$  every point is **optimal**.

**Remark.** We see that each corner (extreme point) corresponds to a **solution** for  $\beta$ , while the blue vector  $\vec{b}$  corresponds to the **dual** constraints  $y^\top A_\eta < c_\eta^\top$ . Only when the blue vector are in the region of orange sectors span by two *normal vectors* of  $y^\top A_{\beta_i} \leq c_{\beta_i}$ , the constraints are satisfied.

**Example** (~~Over~~ Strictly Complementarity (Exercise 5.5. in [Lee22])). Consider

$$\begin{array}{ll}
 \min & c^\top x \\
 & Ax = b \\
 (P) & x \geq 0
 \end{array}
 \qquad
 \begin{array}{ll}
 \max & y^\top b \\
 & y^\top A \leq \vec{0}. \\
 (D) &
 \end{array}$$

<sup>1</sup>This corresponds to the case that we run into the overlapping issue in Figure 5.4.

As previously seen. **Complementarity** of  $\hat{x}$  and  $\hat{y}$ :

$$\begin{aligned}(c_j - \hat{y}^\top A_{\cdot j})\hat{x}_j &= 0, \text{ for } j = 1 \dots n \\ y_i^\top (A_i \hat{x} - b_i) &= 0, \text{ for } i = 1 \dots m\end{aligned}$$

**Definition 5.5.1** (Strictly complementary). For **feasible solutions**  $\hat{x}$  and  $\hat{y}$  are *strictly complementary* if they are **complementary** and exactly one of

$$c_j - \hat{y}^\top A_{\cdot j} \text{ and } \hat{x}_j \text{ is } 0.$$

**Theorem 5.5.1** (Strictly complementarity). If  $(P)$  and  $(D)$  are both **feasible**, then for  $(P)$  and  $(D)$  there exist strictly complementary (**feasible**) **optimal** solutions.

**Intuition.** Let  $v$  be the **optimal** value of  $(P)$ :

$$\begin{aligned}v &= \min c^\top x \\ Ax &= b \\ (P) \quad x &\geq 0\end{aligned}$$

Now, we try to find an **optimal solution** with

$$x_j > 0, \quad \text{fix } j$$

by formulating the following linear programming

$$\begin{aligned}\max \quad & x_j \\ c^\top x &\leq v \\ Ax &= b \\ (P_j) \quad x &\geq 0\end{aligned}$$

where  $P_j$  seeks an **optimal solution** of  $(P)$  that has  $x_j$  being positive. If failed, then construct an **optimal solution**  $\hat{y}$  to  $(D)$  with

$$c_j - \hat{y}^\top A_{\cdot j} > 0.$$

We then see for any **fixed**  $j$ , the desired property holds. The only thing we need to do is combine these  $n$  pairs of  $\hat{x}$  and  $\hat{y}$  appropriately to construct **optimal**  $\hat{x}$  and  $\hat{y}$  that are **overly complementary**.

## Lecture 13: Duality

We now formally prove **Theorem 5.5.1**.

18 Oct. 08:00

**Proof.** First prove for one fixed  $j$ . Consider

$$\begin{aligned}\max \quad & x_j \\ c^\top x &\leq v \\ Ax &= b \\ (P_j) \quad x &\geq 0,\end{aligned}$$

where

$$\begin{aligned}c^\top x \\ Ax &= b \\ x &\geq 0\end{aligned}$$

is trying to model the set of **optimal solutions** to  $(P)$ , and  $P_j$  is trying to find an **optimal solution** of  $(P)$  with  $x_j > 0$ .

We see that there are three cases.

1.  $P_j$  has an **optimal solution**.  $\hat{x}$  with  $\hat{x}_j > 0$ . Take  $\hat{x}$  **optimal** for  $P_j \Rightarrow \hat{x}$  **optimal** for  $(P)$ . Take a  $\hat{y}$  **optimal** for  $(D)$ .
2.  $P_j$  is unbounded. Take any **feasible solutions**  $\hat{x}$  of  $P_j$  with  $\hat{x}_j > 0$ .
3. The **optimal** value of  $P_j$  is zero. Then consider the **dual** of  $P_j$ , denoted by  $D_j$  with the **dual** variables  $w \in \mathbb{R}$ ,  $y \in \mathbb{R}^m$ . We then have

$$\begin{aligned} \min \quad & wv + y^\top b \\ & wc^\top + y^\top A \geq e_j^\top \\ (D_j) \quad & w \geq 0, \ y \text{ unres.} \end{aligned}$$

Suppose  $\hat{w}$  and  $\hat{y}$  is **optimal** for  $D_j$ .

Case 1.  $\hat{w} > 0$ : Then

$$\begin{aligned} & -c^\top + \left( \frac{\hat{y}^\top}{-\hat{w}} \right) A \not\leq \frac{1}{-\hat{w}} e_j^\top \\ \Rightarrow \underbrace{\left( \frac{\hat{y}^\top}{-\hat{w}} \right) A}_{\hat{y}} & \leq c^\top - \frac{1}{\hat{w}} e_j^\top \\ \Rightarrow \hat{y}^\top A & \leq c^\top - \frac{1}{\hat{w}_j} e_j^\top \\ \Rightarrow \hat{y}^\top A & \leq c^\top \text{ with a little slack in the } j^{th} \text{ constraint.} \\ \Rightarrow \hat{y}^\top A_{\cdot j} & \leq c_j - \frac{1}{\hat{w}} < c_j, \forall j. \end{aligned}$$

Note that the **optimal** value of  $D_j$  is zero since the **optimal** value of  $P_j$  is zero. Then

$$\begin{aligned} \hat{w}v + \hat{y}^\top b &= 0 \\ \Rightarrow -v + \left( \frac{\hat{y}^\top}{-\hat{w}} \right) b &= 0 \\ \Rightarrow \hat{y}^\top b &= v \\ \Rightarrow \hat{y} & \text{ is optimal for } D. \end{aligned}$$

Case 2.  $\hat{w} = 0$ : Then

$$\hat{y}^\top A \geq e_j^\top.$$

Let  $\tilde{y}$  be an **optimal solution** of  $(D)$ . Now consider  $\tilde{y} - \hat{y}$ , we have

$$(\tilde{y} - \hat{y})^\top A = \underbrace{\tilde{y}^\top A}_{\leq c^\top} - \underbrace{\hat{y}^\top A}_{\geq e_j^\top} \leq c^\top - e_j^\top,$$

we see that  $(\tilde{y} - \hat{y})$  is **feasible** for  $(D)$  with **slackness** in the right-hand side in the  $j^{th}$  constraint.

Then the objective value of  $\tilde{y} - \hat{y}$  of  $(D)$  is

$$(\tilde{y} - \hat{y})^\top b = \tilde{y}^\top b - \hat{y}^\top b = v - \hat{y}^\top b = v$$

since  $\hat{y}^\top b$  is the **optimal** value of  $D_j$ , which is zero.

Notice that this is just for a fixed  $j$ !

$j$	$\hat{x}^\top$		$c^\top - \hat{y}^\top A$	
1	$\ddots$	0	$\ddots$	
$\vdots$		$\vdots$		
$j$	$\rightarrow \hat{x}^{(j)}$	0/+	+ / 0	$\leftarrow c^\top - \hat{y}^{(j)\top} A$
$\vdots$		$\ddots$		$\ddots$
$n$		0 $\ddots$		+ $\ddots$
	$\hat{x}$	$\uparrow$	$c^\top - \hat{y}^\top A$	
		0	+	

**Intuition.** We average out for all  $j$ , then we have

$$\hat{x} := \sum_{j=1}^n \frac{1}{n} \hat{x}^{(j)}, \quad \hat{y} := \sum_{j=1}^n \frac{1}{n} \hat{y}^{(j)}$$

We check that  $\hat{x}$  and  $\hat{y}$  are **feasible**. Since

$$A\hat{x} = A \left( \frac{1}{n} \sum_{j=1}^n \hat{x}^{(j)} \right) = \frac{1}{n} \sum_{j=1}^n \underbrace{A\hat{x}^{(j)}}_b = b.$$

⊛

**Problem 5.5.1.** For multicommodity flow problem, we see that

$$\begin{aligned} \min \quad & \sum_{(i,j) \in \mathcal{A}} c_{ij} x_{ij} \\ \text{s.t.} \quad & \underbrace{\sum_{j: (i,j) \in \mathcal{A}} x_{ij}}_{\text{flow out of } i} - \underbrace{\sum_{j: (j,i) \in \mathcal{A}} x_{ji}}_{\text{flow into } i} = b_i, i \in \mathcal{N} \\ & x_{ij} \geq 0 \leq u_{ij} \text{ for } (i,j) \in \mathcal{A} \end{aligned}$$

**Answer.** Write it in the matrix form, we have

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ & 0 \leq x \leq u, \end{aligned}$$

write it in another way, we have

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ & Ix \leq u \\ & x \geq 0 \end{aligned}$$

with the dual variables  $y$  and  $\Pi$ , we have the **dual**

$$\begin{aligned} \max \quad & y^\top b + \Pi^\top u \\ & y^\top A + \Pi^\top I \leq c^\top \\ & y \text{ unres.}, \Pi \leq 0. \end{aligned}$$



The  $A$  looks like

$$A_{(m \times n)} = \begin{matrix} & \text{arc}(i,j) \\ \begin{matrix} \vdots \\ \vdots \\ \vdots \\ \vdots \end{matrix} & \begin{pmatrix} \ddots & 0 & & & \\ & \ddots & \vdots & & \\ \dots & \dots & +1 & \dots & \dots \\ & & \vdots & & \\ \dots & \dots & -1 & \dots & \dots \\ & & \vdots & \ddots & \\ & 0 & & \ddots & \end{pmatrix} \end{matrix} \begin{matrix} i \\ j \end{matrix}$$

Then we see the [dual](#) is just

$$\begin{aligned} \max \quad & \sum_{i \in \mathcal{N}} y_i b_i + \sum_{(i,j) \in \mathcal{A}} \Pi_{ij} u_{ij} \\ & y_i - y_j + \Pi_{ij} \leq c_{ij} \quad \text{for all } (i,j) \in \mathcal{A} \\ & \Pi_{ij} \leq 0 \quad \text{for all } (i,j) \in \mathcal{A}. \end{aligned}$$

⊛

# Chapter 6

## Sensitivity Analysis

### Lecture 14: Sensitivity Analysis

As usual, we start with the [primal](#) and the [dual](#)

25 Oct. 08:00

$$\begin{array}{ll} \min & c^\top x \\ & Ax = b \\ (P) & x \geq 0 \end{array} \quad \begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top \\ (D) & \end{array}$$

with an [optimal basic partition](#)  $\beta, \eta$  such that

$$\bar{x} := \begin{cases} \bar{x}_\beta := A_\beta^{-1}b \geq \vec{0} \\ \bar{x}_\eta := \vec{0} \end{cases}, \quad \bar{y}^\top := c_\beta^\top A_\beta^{-1}.$$

As previously seen. The [dual](#) feasibility is

$$\bar{c}_\eta := c_\eta - c_\beta^\top A_\beta^{-1} A_\eta = c_\eta - \bar{y}^\top A_\eta \geq \vec{0}$$

from [Lemma 4.1.2](#).

## 6.1 Local Analysis

### 6.1.1 Change $b$ on the right-hand side

We let

$$b \rightarrow b + \Delta_i e_i = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_i + \Delta_i \\ \vdots \\ b_m \end{pmatrix},$$

then

$$A_\beta^{-1}(b + \Delta_i e_i) = A_\beta^{-1}b + \Delta_i \underbrace{A_\beta^{-1}e_i}_{h^i},$$

where  $h_i$  is the  $i^{th}$  column of  $A_\beta^{-1}$ . So now we have

$$\bar{x}_\beta + \Delta_i h^i \geq \vec{0},$$

where we need  $\beta, \eta$  to still be an [optimal partition](#).

### 6.1.2 Objective Value

Now, the objective value is

$$c_\beta^\top (\bar{x}_\beta + \Delta_i A_\beta^{-1} e_i) + c_\eta^\top \vec{0} = \underbrace{c_\beta^\top \bar{x}_\beta}_{\text{old obj. value}} + \Delta_i \underbrace{c_\beta^\top A_\beta^{-1} e_i}_{\bar{y}^\top} = c_\beta^\top \bar{x}_\beta + \Delta_i \bar{y}_i^\top.$$

### 6.1.3 Analysis

Let  $f$  be

$$\begin{aligned} f(b) &:= \min c^\top x \\ &\quad Ax = b \\ (P_b) \quad &x \geq 0 \end{aligned}$$

where

$$f : \mathbb{R}^m \rightarrow \mathbb{R}.$$

We see that since the **optimal** objective value is equal for the **dual** of  $P_b$ , then  $f(b) = y^\top b$ . Then

$$\frac{\partial f}{\partial b_i} = \bar{y}_i$$

if  $\bar{x}_\beta > \vec{0}$ .

**Problem 6.1.1.** For what values of  $\Delta_i$  is

$$\bar{x}_\beta + \Delta_i h^i \geq \vec{0}?$$

**Answer.** Firstly, we see that we need

$$\bar{x}_{\beta_K} + \Delta_i h_K^i \geq 0 \text{ for } K = 1, \dots, m.$$

Equivalently,

$$\Delta_i h_K^i \geq -\bar{x}_{\beta_K},$$

hence

$$\begin{cases} \Delta_i \geq \frac{-\bar{x}_{\beta_K}}{h_K^i}, & \text{if } h_K^i > 0, \\ \Delta_i \leq \frac{-\bar{x}_{\beta_K}}{h_K^i}, & \text{if } h_K^i < 0. \end{cases}$$

We define  $L_i, U_i$  such that

$$L_i \leq \Delta_i \leq U_i$$

where

$$L_i := \max_{K: h_K^i > 0} \{-\bar{x}_{\beta_K}/h_K^i\}, \quad U_i := \min_{K: h_K^i < 0} \{-\bar{x}_{\beta_K}/h_K^i\}.$$

**Reality Check.** We see that

$$L_i \leq 0 \leq U_i.$$

**Remark.** Noting that if  $h_K^i \leq 0$  for all  $K$ , then we define  $L_i := -\infty$ . Similarly, if  $h_K^i \geq 0$  for all  $K$ , we define  $U_i := \infty$ .

⊛

## 6.2 Global Analysis

We start with a theorem.

**Theorem 6.2.1.** The domain of  $f$  is a **convex set**.

**Proof.** Assume that the **dual** of  $P_b$  is **feasible**, where we denote the **dual** as  $D_b$ :

$$\begin{aligned} \max \quad & y^\top b \\ (D_b) \quad & y^\top A \leq c^\top. \end{aligned}$$

Now, the domain is the set of  $b$  such that  $P_b$  is **feasible**. Mathematically,

$$S := \{b: Ax = b, x \geq 0 \text{ are feasible.}\} \subseteq \mathbb{R}^m.$$

Suppose  $b^1, b^2 \in S$ . We want to check

$$\lambda b^1 + (1 - \lambda)b^2 \in S \text{ for } 0 < \lambda < 1.$$

Notice that there is an  $x^1$  such that

$$Ax^1 = b^1, x^1 \geq \vec{0}$$

and there is an  $x^2$  such that

$$Ax^2 = b^2, x^2 \geq \vec{0}.$$

Firstly, we check that  $\lambda x^1 + (1 - \lambda)x^2$  is non-negative. This is clear since all components are non-negative. Then we check

$$A(\lambda x^1 + (1 - \lambda)x^2) = \lambda b^1 + (1 - \lambda)b^2.$$

This is clear since

$$A(\lambda x^1 + (1 - \lambda)x^2) = \lambda Ax^1 + (1 - \lambda)Ax^2 = \lambda b^1 + (1 - \lambda)b^2.$$

■

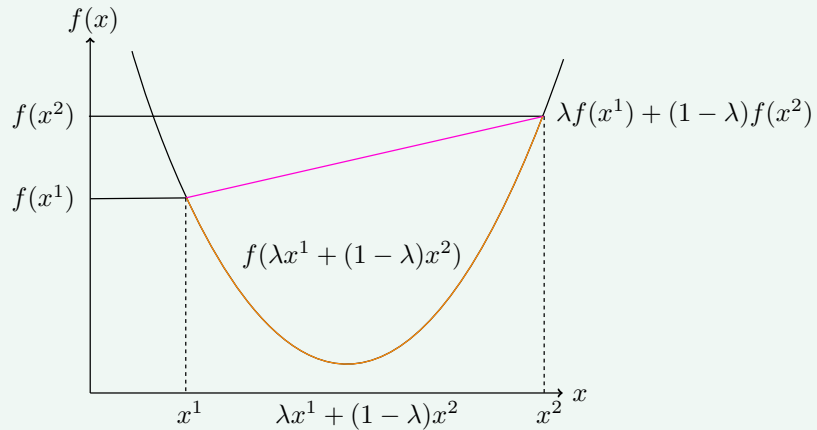
We now introduce the convexity of a function.

**Definition 6.2.1 (Convex function).** We say that  $f$  is a *convex function* on a **convex domain**  $S$  if

$$x^1, x^2 \in S \text{ and } 0 < \lambda < 1,$$

then

$$f(\lambda x^1 + (1 - \lambda)x^2) \leq \lambda f(x^1) + (1 - \lambda)f(x^2).$$



### 6.2.1 Affine Function

Before we go further, we need to have several definitions.

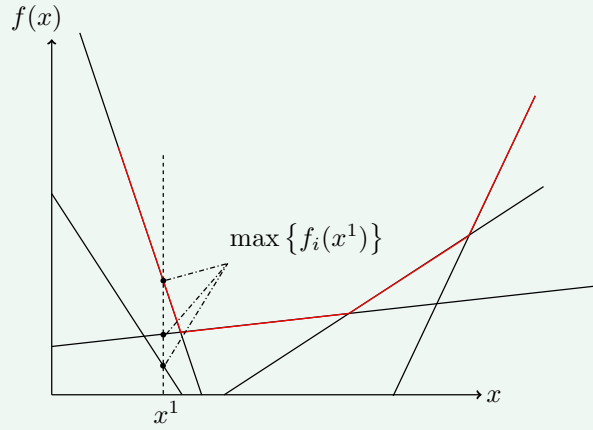
**Definition 6.2.2** (Affine function). A function  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  is *affine* if

$$f(u_1, u_2, \dots, u_m) = a_0 + \sum_{i=1}^m a_i u_i$$

for  $a_i \in \mathbb{R}$ ,  $i = 0, \dots, m$ .

**Remark.** If  $a_0 = 0$ , then  $f$  is a linear function.

**Definition 6.2.3** (Convex piece-wise linear function). A function  $f: \mathbb{R}^m \rightarrow \mathbb{R}$  is a *convex piece-wise linear function* if  $f$  is the point-wise maximum of *affine functions*.



Now, suppose  $f_i: \mathbb{R}^m \rightarrow \mathbb{R}$  for  $i = 1, \dots, K$  and assume that each is *affine*. Then define

$$f(x) := \max_{1 \leq i \leq K} \{f_i(x)\}.$$

**Theorem 6.2.2.** The point-wise maximum of *affine function* is a *convex function*.

**Proof.** We see that

$$\begin{aligned} f(\lambda x^1 + (1 - \lambda)x^2) &= \max_{1 \leq i \leq K} \{f_i(\lambda x^1 + (1 - \lambda)x^2)\} \\ &= \max_{1 \leq i \leq K} \{\lambda f_i(x^1) + (1 - \lambda)f_i(x^2)\} \\ &\geq \max_{1 \leq i \leq K} \{\lambda f_i(x^1)\} + \max_{1 \leq i \leq K} \{(1 - \lambda)f_i(x^2)\} \\ &= \lambda \max_{1 \leq i \leq K} \{f_i(x^1)\} + (1 - \lambda) \max_{1 \leq i \leq K} \{f_i(x^2)\} = \lambda f(x^1) + (1 - \lambda)f(x^2). \end{aligned}$$

**Remark.** The second equality holds since

$$\begin{aligned} &\max_{1 \leq i \leq K} \left\{ a_{i0} + \sum_{l=1}^m a_{il}(\lambda u_l^1 + (1 - \lambda)u_l^2) \right\} \\ &= \max_{1 \leq i \leq K} \left\{ \lambda a_{i0} + (1 - \lambda)a_{i0} + \sum_{l=1}^m a_{il}(\lambda u_l^1 + (1 - \lambda)u_l^2) \right\} \end{aligned}$$

■

## Lecture 15: Sensitivity Analysis

### 6.3 More on Local Analysis

27 Oct. 08:00

As previously seen. Based on an optimal basic solution:

$$\bar{x}_\beta := A_\beta^{-1} \mathbf{b} \geq \vec{0}$$

and the reduced cost

$$\bar{c}_\eta := \mathbf{c}_\eta^\top - \mathbf{c}_\beta^\top A_\beta^{-1} \mathbf{A}_\eta \geq \vec{0},$$

we see that  $\mathbf{c}, \mathbf{b}, \mathbf{A}_\eta$  are linear respect to the objective value. Therefore, there is no limitation for us to only do local analysis respect to  $b$ , we can do this for any one of the data mentioned above.

#### 6.3.1 Change $A_\eta$ on the right-hand side

We now change  $A_\eta$  to do the local analysis for example. If

$$a_{i,\eta_j} \rightarrow a_{i,\eta_j} + \Delta,$$

then

$$A_{\eta_j} = \begin{pmatrix} a_{1,\eta_j} \\ a_{2,\eta_j} \\ \vdots \\ a_{m,\eta_j} \end{pmatrix}.$$

**Problem 6.3.1.** For what  $\Delta$  is  $\beta, \eta$  still an optimal partition?

**Answer.** We see that the reduced cost is now

$$\begin{aligned} \bar{c}'_{\eta_j} &= c_{\eta_j} - \mathbf{c}_\beta^\top A_\beta^{-1} (A_{\eta_j} + \Delta \mathbf{e}_i) \\ &= c_{\eta_j} - \bar{\mathbf{y}}^\top (A_{\eta_j} + \Delta \mathbf{e}_i) \\ &= \bar{c}_{\eta_j} - \Delta \bar{y}_i \underset{\text{want}}{\geq} 0. \end{aligned}$$

Hence, the condition becomes

$$\bar{c}_{\eta_j} \geq \Delta \bar{y}_i.$$

⊛

#### 6.3.2 Change $c$ on the right-hand side

We can also try to change  $c$  for local analysis. Firstly, consider changing  $c_{\eta_j}$ , we have

$$c_{\eta_j} \rightarrow c_{\eta_j} + \Delta,$$

then the reduced cost for  $x_{\eta_j}$  becomes

$$(c_{\eta_j} + \Delta) - \bar{\mathbf{y}}^\top A_{\eta_j} = \bar{c}_{\eta_j} + \Delta \underset{\text{want}}{\geq} 0.$$

Hence, the condition becomes

$$\Delta \geq -\bar{c}_{\eta_j}.$$

Now, for  $c_{\beta_i}$ ,

$$c_{\beta_i} \rightarrow c_{\beta_i} + \Delta.$$

Then

$$\underline{c}_\eta^\top - (\underline{c}_\beta^\top + \Delta \mathbf{e}_i^\top) \underline{A}_\beta^{-1} \mathbf{A}_\eta \underset{\text{want}}{\geq} \vec{0}.$$

We see that the underlined part is just  $\bar{c}_\eta$ , hence the **reduced cost** is just

$$\bar{c}_\eta^\top - \Delta e_i^\top \bar{A}_\eta = (\bar{c}_{\eta_1}, \dots, \bar{c}_{\eta_{n-m}}) - \Delta(\bar{a}_{i,\eta_1}, \bar{a}_{i,\eta_2}, \dots, \bar{a}_{i,\eta_{n-m}}) \underset{\text{want}}{\geq} 0$$

Separate them, we see

$$\bar{c}_{\eta_j} - \Delta \bar{a}_{i,\eta_j} \geq 0 \text{ for } j = 1, \dots, n-m.$$

Equivalently,

$$\Delta \leq \frac{\bar{c}_{\eta_j}}{\bar{a}_{i,\eta_j}} \text{ for } j \text{ such that } \bar{a}_{i,\eta_j} > 0$$

and

$$\Delta \geq \frac{\bar{c}_{\eta_j}}{\bar{a}_{i,\eta_j}} \text{ for } j \text{ such that } \bar{a}_{i,\eta_j} < 0.$$

Recall the definition of  $L$  and  $U$ , we can have the similar inequality for  $\Delta$  such that  $L \leq \Delta \leq U$ , where

$$L := \max_{j: \bar{a}_{i,\eta_j} < 0} \left\{ \frac{\bar{c}_{\eta_j}}{\bar{a}_{i,\eta_j}} \right\} \leq \Delta \leq \min_{j: \bar{a}_{i,\eta_j} > 0} \left\{ \frac{\bar{c}_{\eta_j}}{\bar{a}_{i,\eta_j}} \right\} =: U.$$

### 6.3.3 Change $b$ on the right-hand side, but in two entries

There is no limitation for us to change two entries for  $b$ . Consider

$$b \rightarrow b + \Delta(e_i - e_K).$$

Then

$$\begin{aligned} \bar{x}'_\beta &= A_\beta^{-1}(b + \Delta(e_i - e_K)) \\ &= \bar{x}_\beta + \Delta A_\beta^{-1}(e_i - e_K) \\ &= \bar{x}_\beta + \Delta(h_i - h_K) \underset{\text{want}}{\geq} \vec{0}. \end{aligned}$$

Writing things separately, we have

$$\bar{x}_{\beta_l} + \Delta(h_{il} - h_{Kl}) \geq 0 \text{ for } l = 1, \dots, m$$

where  $H := A_\beta^{-1}$ . Then,

$$\Delta \geq \frac{-\bar{x}_{\beta_l}}{h_{il} - h_{Kl}} \text{ if } h_{il} - h_{Kl} > 0$$

and

$$\Delta \leq \frac{-\bar{x}_{\beta_l}}{h_{il} - h_{Kl}} \text{ if } h_{il} - h_{Kl} < 0.$$

But when we want to change more than one variable in the same time, it becomes more complicated. Consider

$$b \rightarrow b + \Delta_i e_i, \quad c_{\beta_l} \rightarrow c_{\beta_l} + \Delta_l e_l.$$

The condition for  $\beta, \eta$  still being a **basic partition** is

$$A_\beta^{-1}(b + \Delta_i e_i) \geq \vec{0}, \quad c_\eta - (c_\beta + \Delta_l e_l)^\top A_\eta \geq \vec{0}.$$

Originally, the objective value is

$$\underbrace{c_\beta^\top \bar{x}_\beta}_{c_\beta^\top (A_\beta^{-1} b)} = \underbrace{\bar{y}^\top b}_{(c_\beta^\top A_\beta^{-1}) b},$$

after considering the changes, we have

$$(c_\beta + \Delta_l e_l)^\top A_\beta^{-1}(b + \Delta_i e_i).$$

We see that this is a *quadratic* relation. Expanding the expression, we have

$$\begin{aligned} &c_\beta^\top A_\beta^{-1} b + \Delta_i c_\beta^\top A_\beta^{-1} e_i + \Delta_l e_l^\top A_\beta^{-1} b + \Delta_i \Delta_l e_l^\top A_\beta^{-1} e_i \\ &= c_\beta^\top A_\beta^{-1} b + \Delta_i \bar{y}_i + \Delta_l \bar{x}_{\beta_l} + \Delta_i \Delta_l h_{li} \end{aligned}$$

where again,  $H := A_\beta^{-1}$ .

**Remark.** We see that if we hold one of  $\Delta_i$  or  $\Delta_l$  being 0, it's still a linear relation.

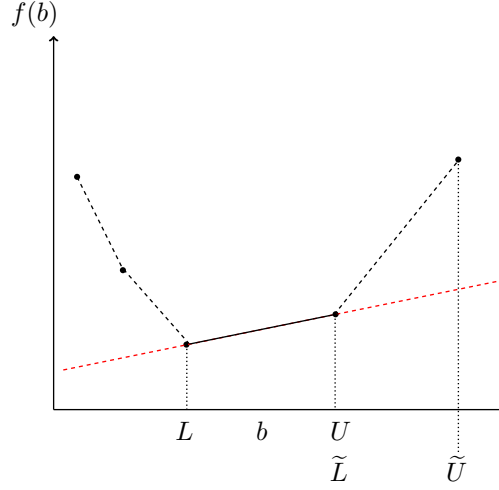


Figure 6.1: Local Analysis

## 6.4 More on Global Analysis

Still, consider the [primal](#) and [dual](#) pair

$$\begin{aligned} f(b) = \min_{\substack{Ax = b \\ x \geq 0}} c^\top x & \quad \max_{\substack{y^\top A \leq c^\top}} y^\top b \\ (P_b) & \quad (D_b) \end{aligned}$$

A [basis](#)  $\beta$  is feasible for  $D_b$  is independent of  $b$ . (recall that  $\bar{y}^\top := c_\beta^\top A_\beta^{-1}$ ) Then we have

$$f(b) := \max \{ (c^\top A^{-1})_\beta b : \beta \text{ is a dual feasible basis} \}.$$

Consider

$$\begin{aligned} g(c) = \min_{\substack{Ax = b \\ x \geq 0}} c^\top x \\ (P_c) \end{aligned}$$

where  $g$  is a piece-wise linear concave function (contrast to [Definition 6.2.3](#)) in  $c$ . We see that  $D_b$  is equivalence to

$$\begin{aligned} - \min & -(y^+ - y^-)^\top b \\ & (y^+ - y^-)^\top A + IS^\top = c^\top \\ & y^+ \geq 0, y^- \geq 0. \end{aligned}$$



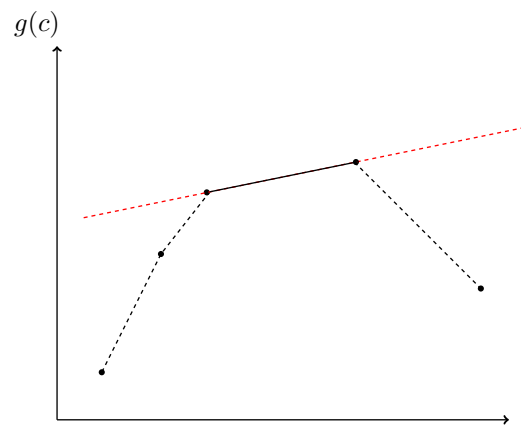


Figure 6.2: The dual version

# Chapter 7

## Large-Scale Linear Optimization

### Lecture 16: Large-Scale Linear Optimization

Let's first look at an example.

1 Nov. 08:00

**Example.** Consider a constraint matrix

$$\begin{pmatrix} \begin{bmatrix} \phantom{0} \end{bmatrix} & & & & \\ \begin{bmatrix} \phantom{0} \end{bmatrix} & & & & \\ & \begin{bmatrix} \phantom{0} \end{bmatrix} & & & \\ & & \begin{bmatrix} \phantom{0} \end{bmatrix} & & \\ & & & \begin{bmatrix} \phantom{0} \end{bmatrix} & \\ & 0 & & & \ddots & \\ & & & & & \begin{bmatrix} \phantom{0} \end{bmatrix} \end{pmatrix}$$

We see that if the first constraint (the first row) doesn't exist, then the problem decomposes to those small block matrix corresponds to some smaller, easier linear optimization problems, and we can solve it very quickly.

**Note.** We called the above matrix as **Nearly Separates**.

There is something we need in order to solve the above problem.

### 7.1 Decomposition Algorithm

In this section we describe what is usually known as **Dantzig-Wolfe Decomposition**. We need

1. [Simplex Algorithm](#).
2. Geometry of [basic feasible solutions](#) and [directions](#).
3. Duality.

We first see a useful theorem.

#### 7.1.1 Representation Theorem

Let  $(P)$  be

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0. \end{aligned}$$

**Theorem 7.1.1 (Representation theorem).** Suppose that  $(P)$  is feasible. Then let  $\mathcal{X}$  be

$$\mathcal{X} := \{\hat{x}^j : j \in \mathcal{J}\}$$

be the set of **basic feasible solutionss** of  $(P)$ . Also, let  $\mathcal{Z}$  be

$$\mathcal{Z} := \{\hat{z}^k : k \in \mathcal{K}\}$$

be the set of basic feasible **rays** of  $(P)$ .

Then the **feasible region** of  $(P)$  is equal to

$$S' := \left\{ \sum_{j \in \mathcal{J}} \lambda_j \hat{x}^j + \sum_{k \in \mathcal{K}} \mu_k \hat{z}^k : \sum_{j \in \mathcal{J}} \lambda_j = 1; \lambda_j \geq 0, j \in \mathcal{J}; \mu^k \geq 0, k \in \mathcal{K} \right\}.$$

**Proof.** Let  $S$  be the **feasible region** of  $(P)$ . We show that  $S = S'$  by showing  $S' \subseteq S$  and  $S' \supseteq S$ .

1.  $S' \subseteq S$ . Since

$$A \left( \sum_j \lambda_j \hat{x}^j + \sum_K \mu_K \hat{z}^K \right) = \sum_j \lambda_j \underbrace{(A\hat{x}^j)}_{=b} + \sum_K \mu^K \underbrace{(A\hat{z}^K)}_{=0} = b.$$

Moreover, since everything in the sum is non-negative, we see that  $S' \subseteq S$ .

2.  $S \subseteq S'$ . Assume  $\hat{x} \in S$ . Then consider the following system

$$\begin{aligned} \begin{matrix} n+1 \\ \text{equations} \end{matrix} \quad & \begin{cases} \sum_{j \in \mathcal{J}} \lambda_j \hat{x}^j + \sum_{k \in \mathcal{K}} \mu_k \hat{z}^k & = \hat{x} \\ \sum_j \lambda_j & = 1 \end{cases} \\ (I) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}; \mu^k \geq 0 \text{ for } k \in \mathcal{K}. \end{aligned}$$

**Note.** Keep in mind that in the above system,  $\hat{x}$  and  $\hat{z}$  are fixed, the variables are the  $\lambda_j$  and  $\mu_k$ .

Now, instead of directly constructing a solution, we use **Farkas' Lemma**. Namely, we write down a system such that if this system is infeasible, by **Farkas' Lemma**, our original system is feasible. Firstly, in **Farkas' Lemma**, we have

$$A = \begin{pmatrix} \hat{x}^1 & \hat{x}^2 & \dots & \hat{z}^1 & \hat{z}^2 & \hat{z}^3 & \dots \\ 1 & 1 & \dots & 0 & 0 & \dots & 0 \end{pmatrix}, \quad b = \begin{pmatrix} \hat{x} \\ 1 \end{pmatrix}$$

in (I). Now, denote the **dual** variables with  $w, t$ , then we have

$$\begin{aligned} (w^\top \quad t) \begin{pmatrix} \hat{x} \\ 1 \end{pmatrix} &> 0 \\ (w^\top \quad t) \begin{pmatrix} \hat{x}^j \\ 1 \end{pmatrix} &\leq 0 \text{ for } j \in \mathcal{J} \\ (w^\top \quad t) \begin{pmatrix} \hat{z}^k \\ 0 \end{pmatrix} &\leq 0 \text{ for } k \in \mathcal{K} \end{aligned}$$

for (II). We only need to show that (II) cannot have a solution. This is easy to show. Firstly,

we see that the above inequalities are equivalent to

$$\begin{aligned} w^\top \hat{x} + t &> 0 & -w^\top \hat{x} &< t \\ w^\top \hat{x}^j + t &\leq 0 \Leftrightarrow -w^\top \hat{x}^j &\geq \hat{t} &\text{ for } j \in \mathcal{J} \\ w^\top \hat{z}^k &\leq \vec{0} & -w^\top \hat{z}^k &\geq 0 \text{ for } k \in \mathcal{K}. \end{aligned}$$

Now, suppose this does have a solution  $\hat{w}, \hat{t}$ . Then, consider

$$\begin{aligned} \min \quad & -\hat{w}^\top x (< \hat{t}) \\ \text{subject to} \quad & Ax = b \\ & x \geq 0. \end{aligned}$$

Notice that the objective value of  $\hat{x}$  here is less than  $\hat{t}$  by (II). Since we know that  $Ax = b$ , hence this linear programming is feasible. Moreover, from  $-\hat{w}^\top \hat{x} \leq \hat{t}$  and  $-\hat{w}^\top \hat{x}^j \geq \hat{t}$ , we see that we have a better solution with respect to the **objective function** among the linear combination of *extreme points*  $\hat{x}^j$ . But this is only possible for unbounded linear programming problem, which needs the positive dot product between rays and the objective vector. But from  $-\hat{w}^\top \hat{z}^k \geq 0$ , we see that this will never happen, hence the theorem is proved.

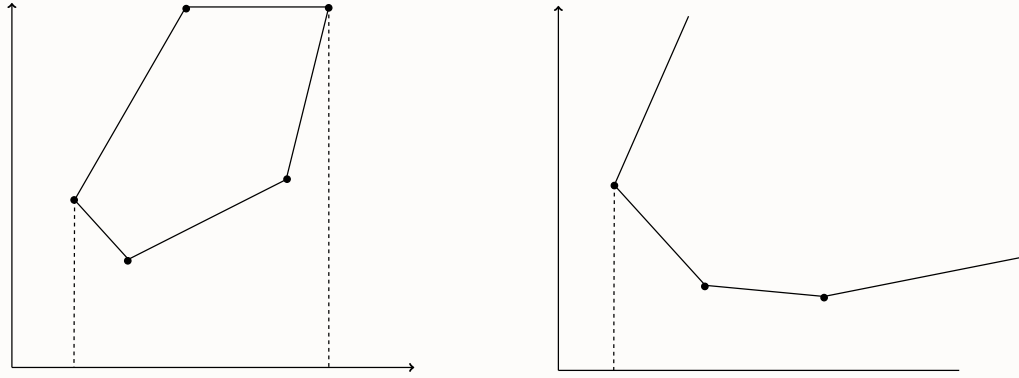


Figure 7.1: Bounded and unbounded case in **Simplex Algorithm**

With this **representation theorem**, consider

$$\begin{aligned} \min \quad & c^\top x \\ \text{subject to} \quad & Ex \geq h \\ \text{"easy"} \quad & \begin{cases} Ax = b \\ x \geq 0. \end{cases} \end{aligned}$$

Then by

$$\left\{ \underbrace{\sum_{j \in \mathcal{J}} \lambda_j \hat{x}^j + \sum_{k \in \mathcal{K}} \mu_k \hat{z}^k}_{= \{x \in \mathbb{R}^n : Ax = b, x \geq \vec{0}\}} : \sum_{j \in \mathcal{J}} \lambda_j = 1, \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu^k \geq 0 \text{ for } k \in \mathcal{K} \right\},$$

we turn the linear problem into

$$\begin{aligned}
 \min \quad & c^\top \left( \sum_{j \in \mathcal{J}} \lambda_j \hat{x}^j + \sum_{k \in \mathcal{K}} \mu_k \hat{z}^k \right) \\
 & E \left( \sum_{j \in \mathcal{J}} \lambda_j \hat{x}^j + \sum_{k \in \mathcal{K}} \mu_k \hat{z}^k \right) \geq h \\
 & \sum_{j \in \mathcal{J}} \lambda_j = 1 \\
 & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}.
 \end{aligned}$$

Furthermore, this is equivalent to

$$\begin{aligned}
 \min \quad & \sum_{j \in \mathcal{J}} (c^\top \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (c^\top \hat{z}^k) \mu_k \\
 & \sum_{j \in \mathcal{J}} (E \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (E \hat{z}^k) \mu_k \geq h \\
 & \sum_{j \in \mathcal{J}} \lambda_j = 1 \\
 (M) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}.
 \end{aligned}$$

The system is now extremely reduced, but the cost is that we now have huge amount of variables. We call this as the [master problem](#).

**Definition 7.1.1 (Master problem).** Given a linear programming problem

$$\begin{aligned}
 \min \quad & c^\top x \\
 & Ex \geq h \\
 \text{"easy"} \quad & \begin{cases} Ax = b \\ x \geq 0, \end{cases}
 \end{aligned}$$

the so-called *master problem* is defined as

$$\begin{aligned}
 \min \quad & \sum_{j \in \mathcal{J}} (c^\top \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (c^\top \hat{z}^k) \mu_k \\
 & \sum_{j \in \mathcal{J}} (E \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (E \hat{z}^k) \mu_k \geq h \\
 & \sum_{j \in \mathcal{J}} \lambda_j = 1 \\
 (M) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}.
 \end{aligned}$$

We formalize the above result as so-called [Decomposition Theorem](#).

**Theorem 7.1.2 (Decomposition theorem).** Let

$$\begin{aligned}
 \min \quad & c^\top x \\
 & Ex \geq h \\
 & Ax = b \\
 (Q) \quad & x \geq 0
 \end{aligned}$$

Let  $S := \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ ,  $\mathcal{X} := \{\hat{x}^j : j \in \mathcal{J}\}$  be the set of [basic feasible solutions](#)  $S$  and  $\mathcal{Z} := \{\hat{z}^k : k \in \mathcal{K}\}$  be the set of basic feasible [rays](#) of  $S$ . Then  $Q$  is equivalent to the [master problem](#)

(M)

$$\begin{aligned}
\min \quad & \sum_{j \in \mathcal{J}} (c^\top \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (c^\top \hat{z}^k) \mu_k \\
& \sum_{j \in \mathcal{J}} (E \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (E \hat{z}^k) \mu_k \geq h \\
& \sum_{j \in \mathcal{J}} \lambda_j = 1 \\
(M) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}.
\end{aligned}$$

**Remark.** We think of  $E$  being a *complicated* constraint matrix, while  $A$  is much easier. Further, the reason why we choose  $\leq$  for  $E$  and  $=$  for  $A$  is not because this makes them complicated or easy, but only for our convenience. In deed, we will soon see that we can turn  $M$  into a [standard form](#) problem without increasing complexity.

## 7.2 Solution of the Master Problem via the Simplex Algorithm

We now want to solve  $M$ . And since we can't write out  $M$  explicitly since there are too many variables. But instead, we can reasonably *maintain* a [basic solution](#) of  $\bar{M}$ , the [standard form](#) of  $M$ . Furthermore, the only part of the [simplex algorithm](#) that is sensitive to the total number of variables is when we check for variables with negative [reduced cost](#). So we now try to find an indirect way to check this rather than find it one by one.

Denotes the [dual](#) variable of  $M$  as  $y$  and  $\sigma$  with  $y \geq \bar{0}$  and  $\sigma$  unrestricted. We further turn  $M$  into the [standard form](#) problem, which is just

$$\begin{aligned}
\min \quad & \sum_{j \in \mathcal{J}} (c^\top \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (c^\top \hat{z}^k) \mu_k \\
& \sum_{j \in \mathcal{J}} (E \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (E \hat{z}^k) \mu_k - Is = h \\
& \sum_{j \in \mathcal{J}} \lambda_j = 1 \\
(\bar{M}) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}, s \geq 0.
\end{aligned}$$

Suppose that  $\bar{y}, \bar{\sigma}$  forms a [basic dual solution](#). The [reduced cost](#) of  $\lambda_j$  associated with  $\hat{x}^j$  is

$$(c^\top \hat{x}^j) - (\bar{y}^\top \quad \bar{\sigma}) \begin{pmatrix} E \hat{x}^j \\ 1 \end{pmatrix} = c^\top \hat{x}^j - \bar{y}^\top E \hat{x}^j - \bar{\sigma} = (c^\top - \bar{y}^\top E) \hat{x}^j - \bar{\sigma}$$

since  $\bar{c}_{\eta_j} = c_{\eta_j} - \bar{y}^\top A_{\eta_j}$ .

**Problem 7.2.1.** Is there a  $\lambda_j$  with this [reduced cost](#) negative?

**Answer.** Consider

$$\begin{aligned}
-\sigma + \min \quad & (c^\top - \bar{y}^\top E)x \\
& Ax = b \\
& x \geq 0.
\end{aligned}$$

⊛

## Lecture 17: Large-Scale Linear Optimization

**As previously seen.** We now focus on one particular problems: What's the conditions for a variable to enter the [basis](#)?

3 Nov. 08:00

1. What's the **reduced cost** of  $s_i$ ?

$$0 - (\bar{y}^\top \quad \bar{\sigma}) \begin{pmatrix} -e_i \\ 0 \end{pmatrix} = \bar{y}_i.$$

If  $\bar{y}_i < 0$ , then  $s_i$  can enter the **basis**.

2. What's the **reduced cost** of  $\lambda_j$ ?

$$(c^\top \hat{x}^j) - (\bar{y}^\top \quad \bar{\sigma}) \begin{pmatrix} E\hat{x}^j \\ 1 \end{pmatrix} = c^\top \hat{x}^j - \bar{y}^\top E\hat{x}^j - \bar{\sigma} = (c^\top - \bar{y}^\top E)\hat{x}^j - \bar{\sigma}.$$

We consider a sub problem

$$\begin{aligned} -\sigma + \min \quad & (c^\top - \bar{y}^\top E)x \\ & Ax = b \\ \text{(SUB)} \quad & x \geq 0. \end{aligned}$$

If the **optimal** value  $< 0$ , then the **optimal basic solution**  $\hat{x}^j$  has an associated  $\lambda_j$  with negative **reduced cost**, so  $\lambda_j$  can enter the **basis** of  $(M)$ . Else if the **optimal** value  $\geq 0$ , then no  $\lambda_j$  can enter the **basis**.

**Note.** We need to include  $-\sigma$  for evaluating the **optimal** values.

**Problem 7.2.2.** What if the **optimal** value is unbounded?

3. What's the **reduced cost** of  $\mu^k$ ?

$$(c^\top \hat{z}^k) - (\bar{y}^\top \quad \bar{\sigma}) \begin{pmatrix} E\hat{z}^k \\ 0 \end{pmatrix} = (c^\top - \bar{y}^\top E)\hat{z}^k.$$

Again, consider a sub problem

$$\begin{aligned} \min \quad & (c^\top - \bar{y}^\top E)z \\ & Az = \vec{0} \\ & z \geq 0 \end{aligned}$$

**Remark.** Compare this problem to the previous sub problem SUB.

- (a) Notice that the objective value of this problem will always be 0 or unbounded. Since 0 is always a **feasible solution**, or if once it's negative, we can multiply it by a positive number and make the **optimal** values smaller.
- (b) When solving SUB, the **optimal** values of SUB is
  - i. negative  $\Rightarrow \lambda_j$  to enter the **basis**.
  - ii. non-negative  $\Rightarrow$  no  $\lambda_j$  can enter the **basis**.
  - iii. unbounded  $\Rightarrow$  we get a  $\bar{z}$  that is a basic **ray** with  $c^\top \bar{z} < 0$ , which implies for some  $\hat{z}^k, \mu^k$  with negative **reduced cost**.

**Note.** We stop when SUB has the **optimal** values being 0.

Now, we know what variable can enter the **basis**, but we have not yet consider what variable can leave. Recall that the basic matrix  $B$  for  $\bar{M}$  will have the following columns

$$s_i = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, \quad \lambda_j = \begin{pmatrix} E\hat{x}^j \\ 1 \end{pmatrix}, \quad \mu^k = \begin{pmatrix} E\hat{z}^k \\ 0 \end{pmatrix},$$

Why the **optimal** values of SUB will always be non-positive?

where we see that the last entries of  $\lambda_j$  will always be 1, and at least one of  $\lambda_j$  will be in the **basis** due to the fact that  $B$  is invertible. For simplicity, we just consider

$$B = \begin{pmatrix} & -I & & E\hat{x}^1 \\ 0 & \dots & 0 & 1 \\ & s_1 & s_2 & \dots & s_k & \lambda_1 \end{pmatrix}$$

where we get  $\hat{x}^1$  by solving

$$\begin{aligned} \min \quad & e^\top \hat{x} \\ \text{s.t.} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

If  $E\hat{x}^1 \geq h$ , then  $\bar{s} \geq \bar{0} \Rightarrow$  directly go to Phase II. Then,

$$(\bar{y}^\top \quad \bar{\sigma}) = ((\bar{c}\hat{x}^j) \quad (c^\top \hat{z}^k) \quad 0) B^{-1},$$

where  $\bar{c}\hat{x}^j$  initially is

$$(0 \quad \dots \quad 0 \quad c^\top \hat{x}^1).$$

Recall the ratio test for determining what entry should enter the **basis** and what should leave. Namely,

$$\bar{y}^\top = c_\beta^\top A_\beta^{-1}, \quad \bar{x}_\beta = A_\beta^{-1}b = \begin{pmatrix} \bar{x}_{\beta_1} \\ \vdots \\ \bar{x}_{\beta_m} \end{pmatrix}, \quad \bar{A}_{\eta_j} = A_\beta^{-1}A_{\eta_j} = \begin{pmatrix} \bar{a}_{1,\eta_j} \\ \vdots \\ \bar{a}_{m,\eta_j} \end{pmatrix}$$

with the ratio being

$$\min_{i: \bar{a}_{i,\eta_j} > 0} \left\{ \frac{\bar{x}_{\beta_i}}{\bar{a}_{i,\eta_j}} \right\}.$$

Now, in our situation, we carry out the ratio test by noting that the basic variable values is just

$$B^{-1} \begin{pmatrix} h \\ 1 \end{pmatrix},$$

and the updated entering column is

$$B^{-1} \begin{pmatrix} -e_i \\ 0 \end{pmatrix} \text{ or } B^{-1} \begin{pmatrix} E\hat{x}^j \\ 1 \end{pmatrix} \text{ or } B^{-1} \begin{pmatrix} E\hat{z}^k \\ 0 \end{pmatrix},$$

which corresponds to  $\lambda_j$ ,  $\mu_k$ ,  $s_i$  is entering the **basis**, respectively.

Then we just do the ratio test. If  $B^{-1} \begin{pmatrix} h \\ 1 \end{pmatrix} \geq \bar{0} \Rightarrow$  go to Phase II. If not we create an artificial column

$$\begin{pmatrix} E\hat{x}^1 \\ 1 \end{pmatrix}.$$

## Lecture 18: Lagrangian Relaxation

**As previously seen.** The **Simplex Algorithm**.

8 Nov. 08:00

1. Initialization (Phase I). Find an initial **basic feasible partition**  $\beta, \eta$
2. Is there a **non-basic** variable with negative **reduced cost**?

$$\bar{c}_j := c_j - \bar{y}^\top A_{\eta_j} < 0.$$

If not, then we have an **optimal solution**.

3. Find the leaving variable.

$$i^* := \arg \max_{\bar{a}_{i,\eta_j} > 0} \left\{ \frac{\bar{x}_{\beta_i}}{\bar{a}_{i,\eta_j}} \right\}.$$

If  $i^*$  is undefined, then problem is unbounded.



4. Swap  $\beta_i$  and  $\eta_j$  and **GOTO 2**.

Then the decomposition algorithm can be written as follows. We change the step 0. and 2. of the above [simplex algorithm](#) into the following.

0. Reformulate  $Q$  as  $M$  and apply [simplex algorithm](#) to  $M$ , where

$$\begin{aligned} \min \quad & c^\top x \\ & Ex \geq h \\ & Ax = b \\ (Q) \quad & x \geq 0 \end{aligned}$$

and

$$\begin{aligned} \min \quad & \sum_{j \in \mathcal{J}} (c^\top \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (c^\top \hat{z}^k) \mu_k \\ & \sum_{j \in \mathcal{J}} (E \hat{x}^j) \lambda_j + \sum_{k \in \mathcal{K}} (E \hat{z}^k) \mu_k \geq h \\ & \sum_{j \in \mathcal{J}} \lambda_j = 1 \\ (M) \quad & \lambda_j \geq 0 \text{ for } j \in \mathcal{J}, \mu_k \geq 0 \text{ for } k \in \mathcal{K}. \end{aligned}$$

2. Solve the sub-problem

$$\begin{aligned} -\bar{\sigma} + \min \quad & \bar{c} - \bar{y}^\top E \\ & Ax = b \\ & x \geq 0 \end{aligned}$$

- [optimal](#) & Objective value  $< 0 \Rightarrow$  a  $\lambda$  variable can enter the [basis](#).
- [optimal](#) & Objective value  $> 0 \Rightarrow$  have an optimal for  $(M)$ .
- Unbounded  $\Rightarrow$  a  $\mu_k$  variable can enter the [basis](#).

**Note.** Compare 2. here and 2. in the [simplex algorithm](#).

**Remark.** For the real implementation in step 2., we

1. Keep all generated columns.
2. First check [reduced costs](#) of columns already generated. **Repeat.** Only solve for sub-problem when needed.

**Note.** We see that we are solving  $(M)$  over the known columns. So instead, we can pass  $(M)$  to a solver (Gurobi). And since it will give us the [dual](#) variable  $\bar{y}$  and  $\bar{\sigma}$ , we can continue to solve the sub-problem without problems. Furthermore, we solve the sub-problem and append new column to known ones and go solve the sub-problem again. In short, let the solver keep track of the [basis](#).

## 7.3 Lagrangian Relaxation

The motivation is to get a good lower bound of optimal objective value for

$$\begin{aligned} z := \min \quad & c^\top x \\ & Ex \geq h \\ & Ax = b \\ (Q) \quad & x \geq 0 \end{aligned}$$

Since the problem is large, hence we want to exit the algorithm whenever we get a *good enough* solution such that it's not far away from the objective value. But the problem is, when should we stop? Do we stop at plateaus? What if there is a second drop in terms of objective value?

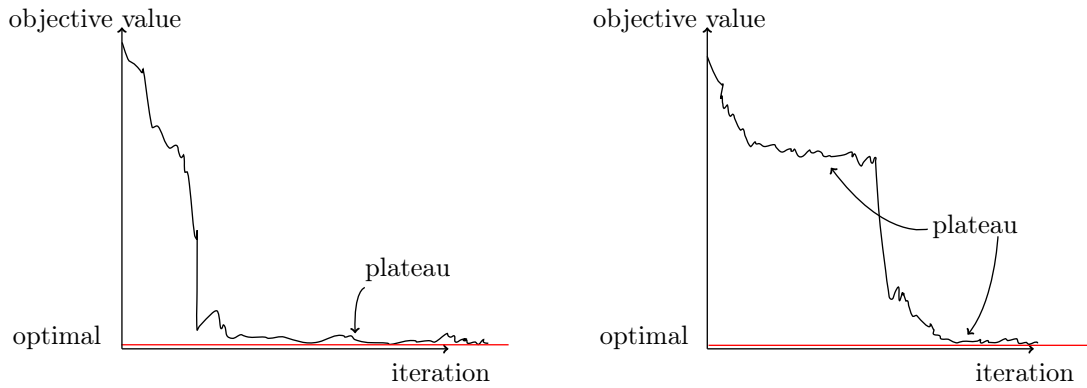


Figure 7.2: Early Arrival, can we?

### 7.3.1 Lagrangian Bounds

We first start with a specific problem which is important in our analysis.

**Definition 7.3.1** (Lagrangian subproblem). We choose  $\hat{y} \geq \vec{0}$ , and the corresponding *Lagrangian subproblem*  $L_{\hat{y}}$  is defined as

$$\begin{aligned} v(\hat{y}) := \hat{y}^\top h + \min & (c^\top - \hat{y}^\top E)x \\ & Ax = b \\ (L_{\hat{y}}) \quad & x \geq 0 \end{aligned}$$

where  $L$  stands for Lagrange<sup>a</sup>.

<sup>a</sup>[https://en.wikipedia.org/wiki/Joseph-Louis\\_Lagrange](https://en.wikipedia.org/wiki/Joseph-Louis_Lagrange)

**Intuition.** We are trying to *bring* the complex constraint  $Ex \geq h$  into the objective function.

To characterize how good will this approximation be, we first see a simple result.

**Lemma 7.3.1.** For any  $\hat{y} \geq \vec{0}$ ,  $v(\hat{y}) \leq z$ .

**Proof.** Let  $x^*$  be an **optimal solution** for  $Q$ . Then  $x^*$  is **feasible** for  $L_{\hat{y}}$ . Then we see

$$v(\hat{y}) \leq \hat{y}^\top h + (c^\top - \hat{y}^\top E)x^* = \underbrace{c^\top x^*}_z + \underbrace{\hat{y}^\top}_{\geq \vec{0}} \underbrace{(h - Ex^*)}_{\leq \vec{0}} \leq z$$

■

Denote the **dual** variables of  $Q$  as  $y$  and  $\pi$ . The **dual** of  $Q$  is

$$\begin{aligned} \max \quad & y^\top h + \pi^\top b \\ & y^\top E + \pi^\top A \leq c^\top \\ & y \geq 0 \end{aligned}$$

and the **dual** of  $L_{\hat{y}}$  is

$$\begin{aligned} \hat{y}^\top h + \max \quad & \pi^\top b \\ & \pi^\top A \leq c^\top - \hat{y}^\top E \end{aligned}$$

**Note.**  $\hat{y}$  is not the variable.

**Theorem 7.3.1** (Lagrangian dual theorem). Suppose  $x^*$  is optimal for  $Q$ . Further, suppose  $\hat{y}$  and  $\hat{\pi}$  are optimal for the dual of  $Q$ . Then

- $x^*$  is optimal for  $L_{\hat{y}}$
- $\hat{\pi}$  is optimal for the dual of  $L_{\hat{y}}$
- $\hat{y}$  is a maximizer of  $v(y)$  over  $y \geq \vec{0}$
- The maximum value of  $v(y)$  over  $y \geq \vec{0}$  is  $z$ .

**Proof.** We first make a note.

**Note.** In above, we want to find

$$z = \max_{y \geq 0} v(y).$$

$x^*$  is feasible for  $L_{\hat{y}}$  and  $\hat{y}$  and  $\hat{\pi}$  is feasible for the dual of  $Q$ . Then

$$\hat{y}^\top E + \hat{\pi}^\top A \leq c^\top$$

with  $\hat{y}^\top \geq \vec{0}$ . But we see that this is equivalent to

$$\hat{\pi}^\top A \leq c^\top - \hat{y}^\top E,$$

which implies  $\hat{\pi}$  is feasible for the dual of  $L_{\hat{y}}$ .

From strong duality theorem for  $Q$ ,

$$c^\top x^* = \hat{y}^\top h + \hat{\pi}^\top b.$$

Then, by using  $E\hat{x}^* \geq h$ , we see that

$$(c^\top - \hat{y}^\top E)x^* \leq \hat{\pi}^\top b.$$

Moreover, recall  $\hat{\pi}$  is feasible for the dual of  $L_{\hat{y}}$  with  $\hat{\pi}^\top A \leq c^\top - \hat{y}^\top E$ , then since  $x^* \geq 0$ , we have

$$\hat{\pi}^\top \underbrace{Ax^*}_b \leq (c^\top - \hat{y}^\top E)x^* \Leftrightarrow \hat{\pi}^\top b \leq (c^\top - \hat{y}^\top E)x^*.$$

We conclude

$$\hat{\pi}^\top b = (c^\top - \hat{y}^\top E)x^*.$$

Now, we claim that  $x^*$  is optimal for  $L_{\hat{y}}$  and  $\hat{\pi}$  is optimal for the dual of  $L_{\hat{y}}$ . Indeed, recall that the objective function of  $L_{\hat{y}}$  is

$$\hat{y}^\top h + \min_{x \geq 0} (c^\top - \hat{y}^\top E)x,$$

with  $(c^\top - \hat{y}^\top E)x^* \geq \hat{\pi}^\top b$ , we see that the objective value of  $x^*$  in  $L_{\hat{y}}$  is equal to  $\hat{y}^\top h + \hat{\pi}^\top b$ , which implies that  $x^*$  is optimal in  $L_{\hat{y}}$  from weak duality theorem.

Lastly, since  $x^*$  is optimal for  $L_{\hat{y}}$ ,

$$z \geq v(\hat{y}) = \hat{y}^\top h + (c^\top - \hat{y}^\top E)x^* = \hat{y}^\top h + \hat{\pi}^\top b = z$$

by optimality for  $\hat{y}$  and  $\hat{\pi}$ , hence we see that  $\hat{y}$  solves

$$\max_{y \geq 0} v(y)$$

and  $v(\hat{y}) = z$ . ■

**Theorem 7.3.2** (Converse Lagrangian dual theorem). Suppose that  $\hat{y}$  is a maximizer of  $v(y)$  over  $y \geq \vec{0}$ . Suppose  $\hat{\pi}$  solves the dual of  $L_{\hat{y}}$ . Then  $\hat{\pi}$  and  $\hat{y}$  solve the dual of  $Q$  and the optimal value of  $Q$  is  $v(\hat{y})$ .

**Intuition.** We see that compare to ??, we now try to say something *backwards*. But we immediately see that it suffers from the situations depicts as follows.

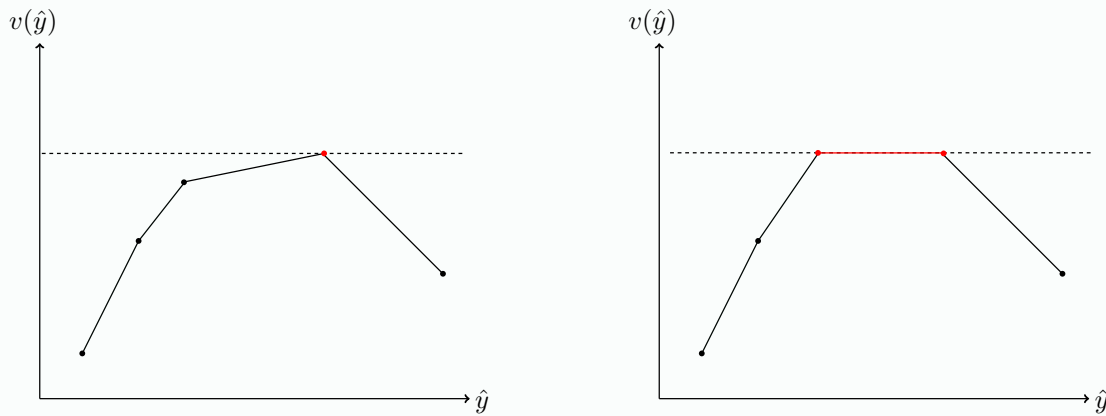


Figure 7.3: There may exist several  $\hat{y}$ !

## Lecture 19: Lagrangian Relaxation

We are building to prove Theorem 7.3.2.

10 Nov. 08:00

As previously seen. We have

$$\begin{aligned} z &:= \min c^\top x \\ &\quad Ex \geq h \\ &\quad Ax = b \\ (Q) \quad &x \geq 0 \end{aligned}$$

and by choosing  $\hat{y} \geq \vec{0}$ , we have the Lagrangian subproblem

$$\begin{aligned} v(\hat{y}) &:= \hat{y}^\top h + \min (c^\top - \hat{y}^\top E)x \\ &\quad Ax = b \\ (L_{\hat{y}}) \quad &x \geq 0. \end{aligned}$$

Now, we introduce another problem.

**Definition 7.3.2** (Lagrangian dual). The *Lagrangian dual* problem is defined as

$$\max_{y \geq \vec{0}} v(y).$$

**Note.** We see that for  $\hat{y} \geq \vec{0}$ ,  $v(\hat{y}) \leq z$ . Now, the goal is to solve the Lagrangian dual to get a lower bound for the original problem. (Notice that this is the maximum of the dual!)

Now, we try to proof Theorem 7.3.2, which is the *partial* converse of Theorem 7.3.1.

**Proof of Theorem 7.3.2.** Recall that

$$\begin{aligned}
 v(\hat{y}) &:= \max_{y \geq \vec{0}} v(y) \\
 &= \max_{y \geq \vec{0}} \left\{ y^\top h + \min_x \underbrace{\left\{ (c^\top - y^\top E)x : Ax = b, x \geq \vec{0} \right\}}_{\text{just a linear program}} \right\} \\
 &= \max_{y \geq \vec{0}} \left\{ y^\top h + \max_{\Pi} \left\{ \Pi^\top b : \Pi^\top A \leq c^\top - y^\top E \right\} \right\} \\
 &= \max_{y \geq \vec{0}, \Pi} \left\{ y^\top h + \Pi^\top b : \Pi^\top A + y^\top E \leq c^\top \right\} \\
 &= z.
 \end{aligned}$$

The last equality is derived from the fact that it's just the **dual** of the  $Q$ . ■

### 7.3.2 Solving the Lagrangian Dual

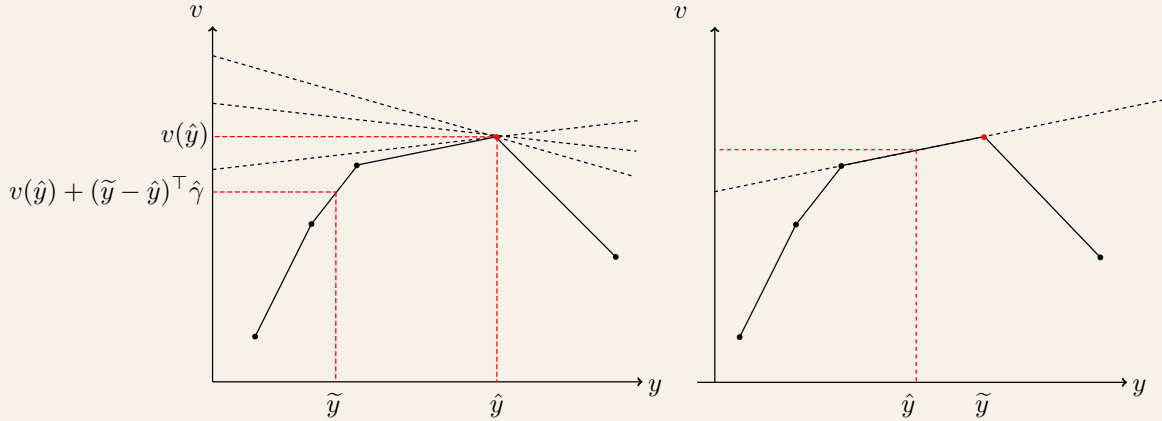
**Intuition.** Theorem 7.3.2 provides a simple way to calculate a lower bound on  $z$  by solving a potentially easier linear optimization problem. But we see that the bound depends on the choice of  $\hat{y} \geq 0$ . This pushes us to find the best such  $\hat{y}$ , and we indeed can solve this by solving the so-called **Lagrangian dual** problem of *maximizing*  $v(y)$  over all  $y \geq 0$  in the domain of  $v$ .

One may want to use some calculus technique to solve for such maximizing problem, but since  $v$  is not a smooth function, rather a piece-wise linear function, hence we need to introduce the concept of **subgradient**. Before we formally introduce it, we first see a theorem.

**Theorem 7.3.3.** Suppose we fix  $\hat{y} \geq \vec{0}$  and solve for  $v(\hat{y})$ . Let  $\hat{x}$  be the **optimal solution** of  $L_{\hat{y}}$ . Denote  $\hat{\gamma} := h - E\hat{x}$ , then

$$v(\tilde{y}) \leq v(\hat{y}) + (\tilde{y} - \hat{y})\hat{\gamma}$$

for all  $\tilde{y}$  in the domain of  $v$ .



**Proof.** We see that since

$$\begin{aligned}
 v(\hat{y}) + (\hat{y} - \tilde{y})\hat{\gamma} &= v(\hat{y}) + (\hat{y} - \tilde{y})(h - E\hat{x}) \\
 &= \hat{y}^\top h + (c^\top - \hat{y}^\top E)\hat{x} + (\tilde{y} - \hat{y})^\top (h - E\hat{x}) \\
 &= \tilde{y}^\top h + (c^\top - \tilde{y}^\top E)\hat{x} \\
 &\geq v(\tilde{y}).
 \end{aligned}$$

The last inequality follows from the fact that  $\hat{x}$  is only **optimal** for  $L_{\hat{y}}$ , not  $L_{\tilde{y}}$ .  $\hat{x}$  may just be **feasible** for  $L_{\tilde{y}}$ . ■

In the theorem,  $\hat{\gamma}$  is the so-called **subgradient**. Given  $\tilde{y}$  and  $\hat{y}$ , we choose  $\hat{\gamma}$  such that the linear estimation  $v(\hat{y}) + (\tilde{y} - \hat{y})^\top \hat{\gamma}$  is always an upper bound on the value  $v(\tilde{y})$  of the function for all  $\tilde{y}$  in the domain of  $f$ . This  $\hat{\gamma}$  is then a **subgradient** of (the concave function<sup>1</sup>)  $v$  at  $\hat{y}$ .

Mathematically, we have the following

**Definition 7.3.3 (Subgradient).** For a concave function  $f$  such that

$$f: I \rightarrow \mathbb{R},$$

the *subgradient* (also known as *subderivative*) at point  $x_0$  is a  $c \in \mathbb{R}$  such that

$$f(x) - f(x_0) \geq c(x - x_0)$$

for every  $x \in I$ .

With this **Theorem 7.3.3** about **subgradient**, we can then develop an algorithm to utilize this.

### 7.3.3 Projected Subgradient Optimization Algorithm

**Intuition.** We iteratively move in the direction of a **subgradient** to maximize  $v$ .

---

**Algorithm 7.1:** Projected subgradient optimization algorithm

---

**Data:** **Objective function** of **Lagrangian dual**  $v(\hat{y}) = \hat{y}^\top h + \min(c^\top - \hat{y}^\top E)x$ , maximum iteration  $K$

**Result:** Estimated maximum value of  $v(\hat{y})$

```

1 Initialize a random  $\mathbb{R}^m$  vector  $\hat{y}^\top \geq \vec{0}$ 
2 for  $k = 1, \dots, K$  do
3   Solve  $L_{\hat{y}^k}$  to get  $\hat{x}^k$ 
4    $\hat{\gamma}^k \leftarrow h - E\hat{x}^k$ 
5    $\hat{y}^{k+1} \leftarrow \text{Proj}_{\mathbb{R}_+^m}(\hat{y}^k + \lambda_k \hat{\gamma}^k)$ 
6 return  $v(\hat{y}^K)$ 
```

---

**Remark.** There are a few remarks we want to make.

- The projection  $\text{Proj}_{\mathbb{R}_+^m}$  is just used to set any negative entries equal to 0.
- The key is in the **line 5**. We want to choose  $\lambda_k > 0$  and satisfying something, which will make this algorithm converges.
  - **Harmonic step size:** Define the step size as  $\lambda_k := \frac{1}{k}$ , which will converge in theory, but it is slow. Notice that this choice of step size is *independent* of the current value of the subgradient.
  - **Polyak step size:** Define the step size as

$$\lambda_k := \frac{\text{GUESS} - v}{\|\hat{\gamma}^k\|^2},$$

where we need an initial GUESS (we get this by literally *guessing*) to let the algorithm behaves reasonable.

## Lecture 20: Convergence of Projected Subgradient Optimization Algorithm

17 Nov. 08:00

---

<sup>1</sup>Contrast to **Definition 6.2.1**.

**As previously seen.** We have already shown the [algorithm of projected subgradient optimization](#), and the key is to choose an adequate step size  $\lambda_k$ . So we now try to give some conditions about how we can choose  $\lambda_k$  such that the algorithm converges.

**Lemma 7.3.2.** Let  $y^*$  be any maximizer of  $v$  over  $y \geq \vec{0}$ . Suppose  $\lambda_k > 0$  for all  $k$ . Then

$$\|y^* - \hat{y}^{k+1}\|^2 - \|y^* - \hat{y}^1\|^2 \leq \sum_{i=1}^k \lambda_i^2 \|\hat{\gamma}^i\|^2 - 2 \sum_{i=1}^k \lambda_i (v(y^*) - v(\hat{y}^i)).$$

**Proof.** Let  $w^{k+1} := \hat{y}^k + \lambda_k \hat{\gamma}^k$ . Then for  $k \geq 1$ ,

$$\begin{aligned} \|y^* - \hat{y}^{k+1}\|^2 - \|y^* - \hat{y}^k\|^2 &\leq \|y^* - w^{k+1}\|^2 - \|\hat{y}^k - \hat{y}^k\|^2 \\ &= \|(y^* - \hat{y}^k) - \lambda_k \hat{\gamma}^k\|^2 - \|y^* - \hat{y}^k\|^2 \\ &= \lambda_k^2 \|\hat{\gamma}^k\|^2 - 2\lambda_k (y^* - \hat{y}^k)^\top \hat{\gamma}^k \leq \lambda_k^2 \|\hat{\gamma}^k\|^2 - 2\lambda_k (v(y^*) - v(\hat{y}^k)), \end{aligned}$$

where the first inequality follows from the triangle inequality, and the last inequality follows from the definition of [subgradient](#). We then do some *telescoping* and see that the result follows.

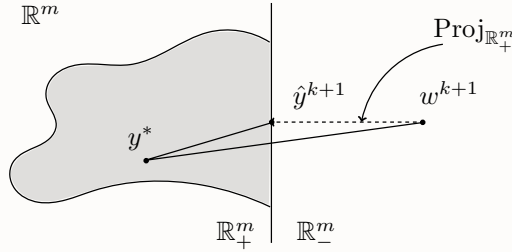


Figure 7.4: Triangle Inequality for  $\text{Proj}_{\mathbb{R}^m_+} w^{k+1} = \hat{y}^{k+1}$ .

Now, denotes  $v_k^* := \max_{i=1, \dots, k} \{v(\hat{y}^i)\}$ , which is just the best function value up to iteration  $k$ . Then we have the following result.

**Theorem 7.3.4.** Let  $y^*$  be any maximizer of  $v$  over  $y \geq \vec{0}$ . Assume that we take a [basic optimal solution](#) of  $L_{\hat{y}^k}$ . We further suppose  $\lambda_k > 0$ ,  $\sum_{k=1}^{\infty} \lambda_k = +\infty$ ,  $\sum_{k=1}^{\infty} \lambda_k^2 < +\infty$ . Then

$$\lim_{k \rightarrow \infty} v_k^* = v(y^*).$$

**Proof.** From the [Lemma 7.3.2](#), the first term of the left-hand side is non-negative, hence we have

$$-\|y^* - \hat{y}^1\|^2 \leq \sum_{i=1}^k \lambda_i^2 \|\hat{\gamma}^i\|^2 - 2 \sum_{i=1}^k \lambda_i (v(y^*) - v(\hat{y}^i)),$$

after rearrangement,

$$2 \sum_{i=1}^k \lambda_i (v(y^*) - v(\hat{y}^i)) \leq \sum_{i=1}^k \lambda_i^2 \|\hat{\gamma}^i\|^2 + \|y^* - \hat{y}^1\|^2.$$

From the definition of  $v_k^*$ , we have

$$2 \sum_{i=1}^k \lambda_i (v(y^*) - v_k^*) \leq \sum_{i=1}^k \lambda_i^2 \|\hat{\gamma}^i\|^2 + \|y^* - \hat{y}^1\|^2.$$

And since the  $(v(y^*) - v_k^*)$  doesn't depend on  $i$  anymore, we can take it out of the summation. We further have

$$0 \leq v(y^*) - v_k^* \leq \frac{\sum_{i=1}^k \lambda_i^2 \|\hat{\gamma}^i\|^2 + \|y^* - \hat{y}^1\|^2}{2 \sum_{i=1}^k \lambda_i}.$$

We observe that  $\|y^* - \hat{y}^1\|^2$  is a constant, denotes it by  $c$ . Further, for all  $i$ ,  $\|\hat{\gamma}^i\|^2$  is bounded, so we can define

$$\Gamma := \max \left\{ \|h - Ex\|^2 : x \text{ is a basic feasible solution of } Ax = b, x \geq 0 \right\}.$$

With  $\Gamma$ , the inequality becomes

$$0 \leq v(y^*) - v_k^* \leq \frac{\Gamma \sum_{i=1}^k \lambda_i^2 + c}{2 \sum_{i=1}^k \lambda_i} \rightarrow 0 \text{ as } k \rightarrow \infty$$

since we assume  $\sum \lambda_i \rightarrow +\infty$  and  $\sum \lambda_i^2 < +\infty$ . Then we see  $v_k^* = v(y^*)$  as  $k \rightarrow \infty$  by squeeze theorem. ■

**Remark.** Suppose we instead choose

$$\lambda_k = s \in \mathbb{R}^+$$

being just a constant. Then the inequality in the above proof becomes

$$\frac{c + s^2 k \Gamma}{2ks} \rightarrow \frac{s\Gamma}{2}.$$

We see that with different choice  $\lambda_k$ , we can simply derive the upper-bound of  $v(y^*) - v_k^*$ .

## 7.4 Cutting-Stock Problem

So far we are talking about constraints being just positive, what about in other domain, like in  $\mathbb{N}^+$ ?

Consider

$$\begin{aligned} \max \quad & \sum_{i=1}^n x_i \\ & x_i + x_j \leq 1, \text{ for all } 1 \leq i < j \leq n \\ & 0 \leq x_i \leq 1. \end{aligned}$$

This linear programming solution is  $x_1 = x_2 = \dots = x_n = \frac{1}{2}$  with the objective value being  $\frac{n}{2}$ . Denotes  $y$  as the dual variables. The dual is

$$\begin{aligned} \min \quad & \sum_{i < j} y_{ij} \\ & \sum_{j: i \neq j} y_{ij} \geq 1 \text{ for all } i = 1, \dots, n \\ & y_{ij} \geq 0 \end{aligned}$$



By setting  $y_{ij} = \frac{1}{n-1}$ , then the objective value is

$$\binom{n}{2} \frac{1}{n-1} = \frac{n}{2},$$

hence we confirm that  $x_i = \frac{1}{2}$  is really the **optimal solution**. One can see that if now we let  $x_i \in \mathbb{N}$ , then the objective solution will be only one of  $x_i = 1$ , and the other  $x_j = 0, j \neq i$ . This leads to an **optimal** value being 1. This just shows how bad if we just **round down** the **optimal solution** when we consider so-called *integer programming*.

## Lecture 21: Cutting-Stock Problem

It's now a good timing to introduce an application of what we have been discussing, namely the **cutting-stock problem**. We'll see that it naturally utilize the idea of column generation. It's an integer programming problem, though contrarily, it can be nicely approximated by **rounding down**.

22 Nov. 08:00

**Problem 7.4.1 (Cutting-stock problem).** Consider we have rolls of paper of width  $W$ , with the demand widths being  $w_1, w_2, \dots, w_m < W$  and demands being  $(D)$ , which is usually pretty big. The goal is to use as few stock rolls as possible.

**Answer.** One may try to define

$$x_{ij} := \# \text{ of rolls of width } w_i \text{ to cut from stock roll } j.$$

But we immediately see that the number of variables is huge for an integer programming, hence this doesn't work. Instead, we denote a *pattern* as a vector  $a$  being

$$a := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix},$$

where  $a_i = \#$  of pieces of width  $w_i$  to cut using this pattern. Then the constraints for a pattern is

$$\sum_{i=1}^m w_i a_i \leq W$$

$$\mathbb{N} \ni a_i \geq 0 \text{ for } i = 1, \dots, m.$$

Moreover, denotes  $(D)$  as

$$d := \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_m \end{pmatrix},$$

then we formulate the cutting-stock problem as

$$z := \min \sum_j x_j$$

$$\sum_j A_{.j} x_j \geq d$$

$$(\text{CSP}) \quad x_j \geq 0 \text{ integer for all } j,$$

where

$$A_{\cdot j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}.$$

Turning CSP into a **standard form** problem and **drop** the integer constraint, we have

$$\begin{aligned} \min \quad & \sum_j x_j \\ & \sum_j A_{\cdot j} x_j - t = d \\ (\underline{\text{CSP}}) \quad & x_j \geq 0 \text{ for all } j, \quad t_i \geq 0 \text{ for all } i = 1, \dots, m. \end{aligned}$$

**Note.**  $\underline{\text{CSP}}$  gives a lower bound on CSP. Moreover, the constraint of  $x_j \in \mathbb{N}$  is now gone.

We now want to solve  $\underline{\text{CSP}}$  exactly to get **optimum**  $\bar{x}, \bar{t}$  with value  $\underline{z} = \sum_{i=1}^m \bar{x}_i$ .

Firstly, if we round up  $\bar{x}$  to  $\lceil \bar{x} \rceil$ , then it is **feasible** for CSP. We immediately see

$$\sum_{i=1}^m \bar{x}_i = \underline{z} \leq z \leq \sum_{i=1}^m \lceil \bar{x}_i \rceil.$$

Since we also have

$$\left\lceil \sum_{i=1}^m \bar{x}_i \right\rceil \leq z,$$

hence we see that the rounding up solution  $\lceil \bar{x} \rceil$  is within  $m - 1$  of **optimum**.

Now we consider how to solve  $\underline{\text{CSP}}$  exactly. Denotes the **dual** variables of  $\underline{\text{CSP}}$  being  $y$  such that

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}.$$

Suppose  $\bar{y}$  is a **basic dual solution**. Then the **reduced cost** of a variable is:

- $t_i$ :

$$0 - \bar{y}^\top (-e_i) = \bar{y}_i.$$

Hence, if  $\bar{y}_i < 0$ ,  $t$  can enter the **basis**.

- $x_j$ :

$$1 - \bar{y}^\top A_{\cdot j} = 1 - \sum_{i=1}^m \bar{y}_i a_{ij}.$$

If this is  $< 0$ , then  $x_j$  can enter the **basis**. To drive some quantity negative, we simply set up a minimization problem. Specifically, we set up a linear program such that

$$\begin{aligned} \min \quad & 1 - \sum_{i=1}^m \bar{y}_i a_{ij} \\ & \sum_{i=1}^m w_i a_{ij} \leq W \\ & a_{ij} \geq 0 \text{ integers.} \end{aligned}$$

Equivalently,

$$\begin{aligned} 1 - \max \quad & \sum_{i=1}^m \bar{y}_i a_i \\ \quad & \sum_{i=1}^m w_i a_i \leq W \\ \quad & a_i \geq 0 \text{ integers for } i = 1, \dots, m. \end{aligned}$$

This is known as the *Knapsack problem*<sup>a</sup>. Now, let  $f(S)$  being the optimal value for knapsack of capacity  $S$  such that  $S = 0, 1, \dots, W$ . We see that

$$\begin{aligned} f(0) &= 0 \\ &\vdots \\ f(S) &= \max_{i: w_i \leq S} \{\bar{y}_i + f(S - w_i)\} \\ &\vdots \\ f(W) &= \text{solution}. \end{aligned}$$

The running time is  $\Theta(Wm)$ .<sup>b</sup>

Notice that the above only gives  $f(W)$ , which is the objective value, but without information for variables. We can retrieve the information by keeping tracking of the argument of maximum in each step, namely we record

$$\begin{aligned} i_0^* &\rightarrow f(0) = 0 \\ &\vdots \\ i_S^* &\rightarrow f(S) = \max_{i: w_i \leq S} \{\bar{y}_i + f(S - w_i)\} \\ &\vdots \\ i_W^* &\rightarrow f(W) = \text{solution}. \end{aligned}$$

Then we simply *back-track* every  $i^*$  from  $i_W^*$ , and then the next one is simply  $i_{W-w_{i_W^*}}^*$ , and so on.

⊛

<sup>a</sup>[https://en.wikipedia.org/wiki/Knapsack\\_problem](https://en.wikipedia.org/wiki/Knapsack_problem)

<sup>b</sup>Note that we assume  $W$  and  $w_i$  to be integers.

## Chapter 8

# Integer-Linear Optimization

### Lecture 22: Optimization of Integer Variables

Let first see some common pitfalls of integer programming.

29 Nov. 08:00

- If  $A$  has big entries and small entries, then these two constraints is like parallel to each other, which will lead the intersection be very far away. Then, if we simply round down the variable, the **optimal** value will drop significantly.

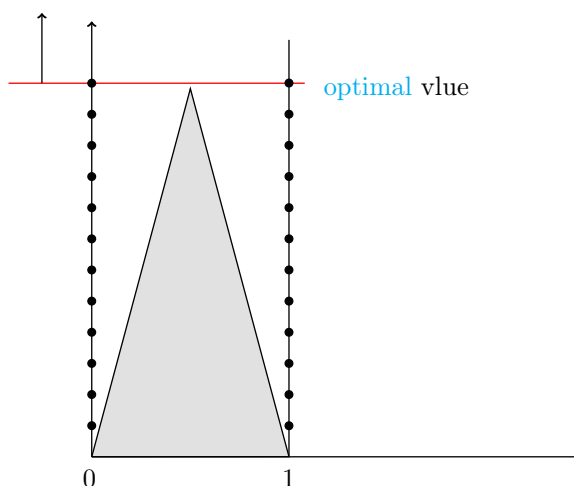


Figure 8.1: Pitfall of Integer Programming

But as one can see, we can often avoid this situation by carefully design our model and the problem is solved.

- Another possibility is the following. Consider an integer with the following constraints.

$$\begin{aligned} \forall_{1 \leq i < j \leq n} \quad & x_i + x_j \leq 1 \\ \forall_{i=1, \dots, n} \quad & x_i \geq 0 \text{ integer.} \end{aligned}$$

Then, there are two **feasible solutions** one can observe immediately, namely

$$x_1 = x_2 = \dots = x_n = \frac{1}{2};$$

and

$$x_1 = 1, x_2 = \dots = x_n = 0.$$

It's then really hard to tell which is better. But again, if the right-hand side is 2 for the first constraint, then the problem is gone.

**Note.** We see that this is totally opposite to the linear programming. The modern integer programming solver can easily solve a programming with like one hundred of variables, but in practice, we're often facing more than thousands of variables. The one need to carefully design his model in terms of number of variables.

## 8.1 Modeling Techniques

We now introduce the so-called **Big-M Method**. Consider the following constraints.

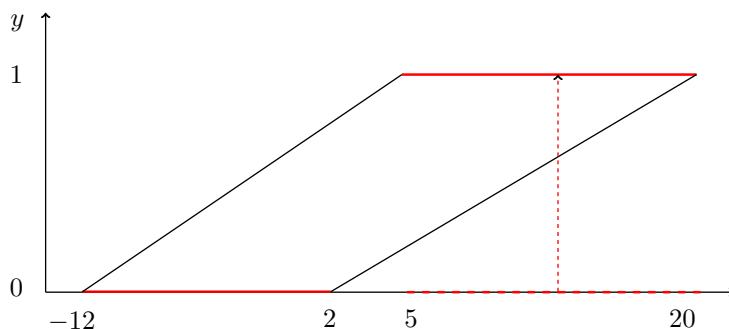
$$-12 \leq x \leq 2 \vee 5 \leq x \leq 20.$$



Then we need to find the smallest **convex set** which contains all **feasible points**. It's just

$$-12 \leq x \leq 20.$$

But that empty space between 2 and 5 causes the problem. To solve this, we simply introduce a new *indicator variable*, denotes it as  $y$ .  $y$  will be 0 if we are in  $[-12, 2]$ , and 1 if we are in  $[5, 20]$ . Then the smallest **convex feasible region** becomes the following quadrilateral.



Put it in the constraints, we have

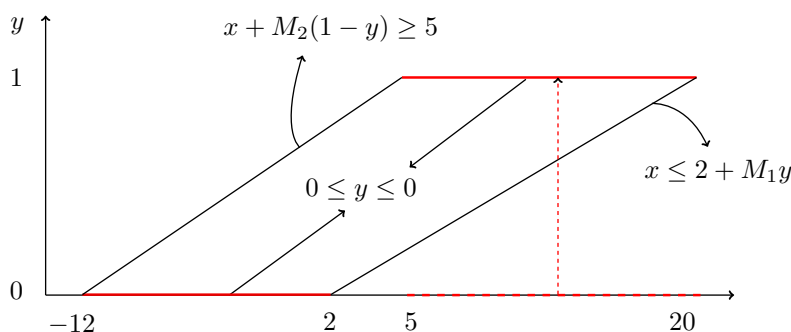
$$\begin{aligned} -12 &\leq x \leq 20 \\ 0 &\leq y \leq 1, \text{ integer} \\ x &\leq 2 + M_1 y \\ x + M_2(1 - y) &\geq 5, \end{aligned}$$

where we let  $M_1$  be big enough to let the constraints always be satisfied when  $x$  is in  $[5, 20]$ . For example, we can let  $M_1 := 18$ , then

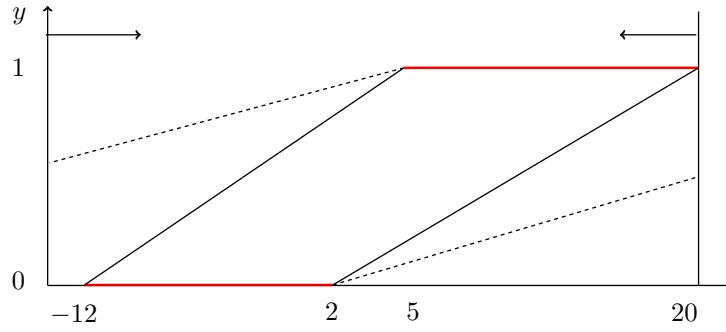
$$\begin{cases} y = 0, & x \leq 2 \\ y = 1, & x \leq 20. \end{cases}$$

Analogously, we use  $M_2$  to help us to model the case that when  $y = 1$ ,  $x \geq 5$  and when  $y = 0$ ,  $x \geq -12$ . For example, we can let  $M_2 := 17$ .

The last three constraints exactly corresponds to the line segment in the graph:



We further see that if we make the constant  $M_i$  too large, we will have



In terms of integer programming, this doesn't affect the integer feasible region. We call this **Big-M Method**. Although this is fine mathematically, but this is unfriendly to the solver.

### 8.1.1 Uncapacitated Facility-location Problem

Assume that there are  $m$  facilities with the fixed costs  $f_i$ ,  $i = 1, \dots, m$ . And assume there are  $n$  customers, denote by  $j = 1, \dots, n$ . Now, let  $c_{ij}$  be the cost of satisfying all demand of customer  $j$  from facility  $i$ . The goal is to minimize the total cost of satisfying all customer's demand. We then define our variables as  $x_{ij}$  such that

$$x_{ij} := \text{proportion of customer } j \text{ demand satisfied facility } i,$$

where  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ . Furthermore, we need indicator variables  $y_i$  such that

$$y_i := \begin{cases} 1, & \text{if facility } i \text{ operates} \\ 0, & \text{if not} \end{cases}$$

for all  $i = 1, \dots, m$ .

The optimization problem can now be modeled as

$$\begin{aligned} \min \quad & \sum_{i=1}^m f_i y_i + \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \\ & \sum_{i=1}^m x_{ij} = 1, & \text{for } j = 1, \dots, n \\ & -y_i + x_{ij} \leq 0, & \text{for } i = 1, \dots, m, j = 1, \dots, n \\ & 0 \leq y_i \leq 1 \text{ integers}, & \text{for } i = 1, \dots, m \\ & x_{ij} \geq 0, & \text{for } i = 1, \dots, m, j = 1, \dots, n. \end{aligned}$$

**Note.** The third constraint

$$-y_i + x_{ij} \leq 0, \quad \text{for } i = 1, \dots, m, j = 1, \dots, n$$

is for the following reason. For any  $i$ , if  $x_{ij}$  is positive for any  $j$ , then we need  $y_i = 1$  to *force* us to pay the fixed cost to operate the facility if anything is shipped out of facility  $i$ . To get this constraint, we first see that we want

$$x_{ij} > 0 \Rightarrow y_i = 1$$

for any  $j$ . It is equivalent to say

$$\sum_{j=1}^n x_{ij} > 0 \Rightarrow y_i = 1.$$

We see that from the first expression, the constraint immediately follows. As for the second con-

straint, we start from considering

$$\sum_{j=1}^n x_{ij} \leq y_i$$

for every  $i = 1, \dots, m$ . But this causes some problem. If the facility is really cheap, then the sum may exceed 1. To solve this problem, we simply make  $y$  become  $n \cdot y$ , namely

$$\sum_{j=1}^n x_{ij} \leq n \cdot y_i$$

for  $i = 1, \dots, m$ , where  $n$  is just the **Big-M** in the big-M Method.

We now have two equivalent constraints, namely

$$\forall_{i,j} \quad -y_i + x_{ij} \leq 0 \quad \text{and} \quad \forall_i \quad \sum_{j=1}^n x_{ij} \leq n \cdot y_i.$$

Now the problem is which to use? The answer is the first one. We call the first model as the *strong model*, while the second model as the *weak model*.

**Intuition.** The second model has the big-M constant. As we just discuss, we prefer  $M$  to be as small as possible. But in the first model, we don't have that big-M coefficient. And since

$$\sum_{j=1}^n (x_{ij} \leq y_i) \Leftrightarrow \sum_{j=1}^n x_{ij} \leq ny_i,$$

we see that the weak constraint is just the sum over all strong constraint. In other words, we have

$$x_{ij} \leq y_i \Rightarrow \sum_{j=1}^n x_{ij} \leq ny_i.$$

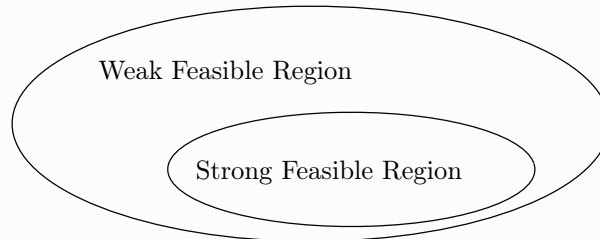
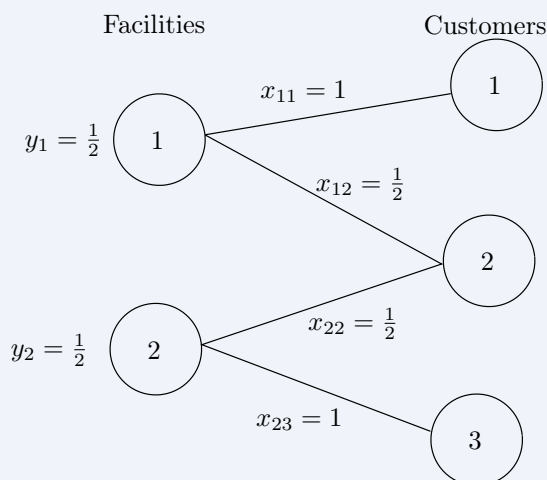


Figure 8.2: Venn diagram of strong and weak [feasible region](#)

**Example.** For  $m = 2$ ,  $n = 3$ , find  $x, y$  where *weak* constraints are satisfied while *strong* constraints are not.



It's easy to check that  $x_{11} \not\leq y_1$ , but

$$x_{11} + x_{12} \leq 3y_1$$

and

$$x_{22} + x_{23} \leq 3y_2.$$

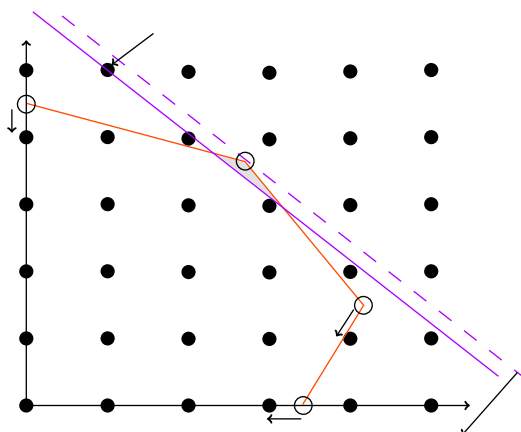
**Remark.** It's important to see that although we said we should keep the number of variables down when setting up the integer programming, but in this case, few is not always better!

**Note (Disaggregation).** It's worth noting that the process of un-summing from the weak constraint to the strong constraint is called *disaggregation*.

## 8.2 Algorithmically Solving Integer-Programming Problem

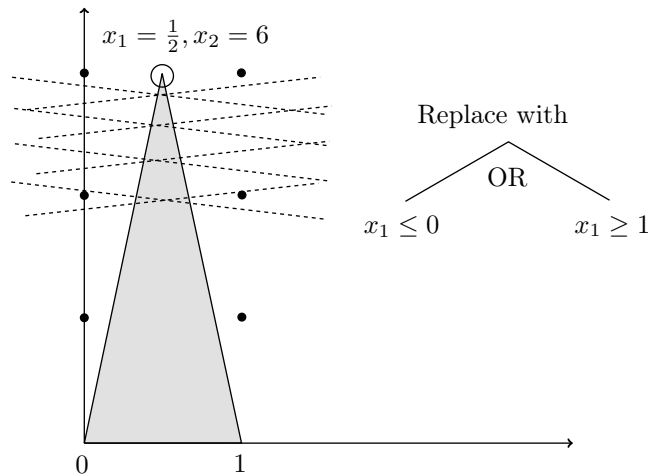
We now see some potential algorithm to solve the integer-programming problem.

- **Cutting-Plane** algorithm. If we have the following feasible region, then the *cutting-plane algorithm* suggests that we should use a plane at a corner (corresponds to an optimal solution to the linear version of this programming) and *reduce* the feasible region by a little until we touch an integer point.





- **Branch-and-Bound** algorithm. We first consider the following **feasible region** and try to use **cutting-plane algorithm**.



We see that if we simply start from **cutting-plane algorithm**, it takes forever to get to the answer. More generally, when the integer solution is far from the linear solution, the **cutting-plane algorithm** performs poorly.

Instead, we consider so-called *Branch-and-Bound algorithm*. It essentially just goes from  $n$  variables to two  $n - 1$  variables programming problem, and until we get to the bottom (1 variable). We see that we are doing exactly the opposite with what we have introduced, namely we are not modeling the **or**, but bring it into the algorithm. In this example, right after we branch, we solve the problem instantly since there in both branches, we only have one point to consider.

**Note.** Every modern solver which solves the integer programming exactly, will first go for **branch-and-bound algorithm**, and then on top of that, solve the remaining problem by **cutting-plane algorithm**.

## Lecture 23: Branch and Bound Algorithm

### 8.2.1 Branch and Bound Algorithm

01 Dec. 08:00

We first dive into **branch and bound algorithm** as we mentioned.

**As previously seen.** The worst case in terms of time complexity for **simplex algorithm** is

$$\Theta(2^n - 1)$$

for  $n$  variables, but it's efficient in practice. And this is similar to the **branch and bound algorithm** for the integral programming problem.

We now focus on the following integer programming,

$$\begin{aligned} \max \quad & y^\top b (=z) \\ & y^\top A \leq c^\top \\ (D_{\mathcal{I}}) \quad & y \in \mathbb{R}^m (y_i \in \mathbb{Z} \text{ for } i \in \mathcal{I}), \end{aligned}$$

where  $\mathcal{I} \subseteq \{1, 2, \dots, m\}$ . By taking the **dual**, we have

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0. \end{aligned}$$

We'll see that the branch and bound algorithm maintains the following:

- $\mathcal{L}$ : A list  $\mathcal{L}$  of *subproblems* that have the form of  $(D_{\mathcal{I}'})$  where  $\mathcal{I}' \subset \mathcal{I}$
- LB: The current best lower bound on  $z$  such that  $\text{LB} \leq z$ .
- $\bar{y}_{\text{LB}}$ : The  $\bar{y}$  corresponds to LB.

**Note.** LB is the objective value of the best *feasible solution* to the original problem seen so far. And we'll set

$$\text{LB} := -\infty$$

if there is no known *feasible solution*.

**Remark.** The key property of  $\mathcal{L}$  is that if there is a *feasible solution* to the original problem that is better than LB, it should be *feasible* for some subproblem on  $\mathcal{L}$ .

Initially, we have

$$\mathcal{L} := \{D_{\mathcal{I}'}\}.$$

And we stop if

$$\mathcal{L} = \emptyset,$$

since this implies  $z = \text{LB}$ .

The general procedure is to take some problem  $(\tilde{D}_{\mathcal{I}'})$  from  $\mathcal{L}$  and remove it, and then solve its continuous relaxation  $(\tilde{D})$  and proceed. Rigorously, we have the following pseudocode.

---

**Algorithm 8.1:** Branch and Bound Algorithm

---

**Data:** (Mixed) Integer programming  $(D_{\mathcal{I}})$  with *feasible*  $(P)^a$

**Result:** *optimal solution*  $\bar{y}$  and *optimal* value, or report  $(D_{\mathcal{I}})$  is infeasible

---

```

1   $\mathcal{L} \leftarrow \{D_{\mathcal{I}'} : \mathcal{I}' \subseteq \mathcal{I}\}$ 
2   $\text{LB} \leftarrow -\infty$ 
3   $\bar{y}_{\text{LB}} \leftarrow$  random vector
4
5  while  $\mathcal{L} \neq \emptyset$  do
6       $D_{\mathcal{I}'} \leftarrow$  an element in  $\mathcal{L}$                                 // See subsection 8.2.3
7       $\mathcal{L} \leftarrow \mathcal{L} \setminus \{D_{\mathcal{I}'}\}$ 
8       $\text{result} \leftarrow \text{SolveContRelax}(D_{\mathcal{I}'})$ 
9      if  $\text{result} \neq \text{infeasible}$  then
10          $\bar{y} \leftarrow \text{result}$                                         // Retrieve basic optimal solution
11         if  $\bar{y}^\top b > \text{LB}$  then
12             if  $\bar{y}_i^\top \in \mathbb{Z}$  for  $i \in \mathcal{I}$  then                                //  $\bar{y}$  solves  $D_{\mathcal{I}}$ 
13                  $\text{LB} \leftarrow \bar{y}^\top b$                                 // Record optimal value
14                  $\bar{y}_{\text{LB}} \leftarrow \bar{y}$                                 // Record optimal solution
15             else                                                //  $\bar{y}$  doesn't solve  $D_{\mathcal{I}}$ ,  $\bar{y}_i^\top \notin \mathbb{Z}$  for some  $i \in \mathcal{I}$ 
16                  $i^* \leftarrow$  an  $i \in \mathcal{I}$  such that  $\bar{y}_i \notin \mathbb{Z}$                                 // See subsection 8.2.4
17                  $D_{\mathcal{I}'}^u \leftarrow D_{\mathcal{I}'}$  with  $y_{i^*} \geq \lceil \bar{y}_{i^*} \rceil$                                 // Up branch
18                  $D_{\mathcal{I}'}^d \leftarrow D_{\mathcal{I}'}$  with  $y_{i^*} \leq \lfloor \bar{y}_{i^*} \rfloor^b$                                 // Down branch
19                  $\mathcal{L} \leftarrow \mathcal{L} \cup \{D_{\mathcal{I}'}^u, D_{\mathcal{I}'}^d\}$ 
20
21 if  $\text{LB} = -\infty$  then
22     return infeasible
23 else
24     return  $\bar{y}_{\text{LB}}, \text{LB}$ 

```

---

<sup>a</sup>This makes sure that the *feasible region* of the continuous relaxation  $(\tilde{D})$  of any  $(\tilde{D}_{\mathcal{I}'})$  is a bounded set, so we can guarantee finite termination.

<sup>b</sup>To match the form, we use  $-y_i \leq -\lceil \bar{y}_i \rceil$ .

**Remark (Finite termination).** We see that there are only finitely many (though exponential)  $D_{\mathcal{I}'}$  can be added into  $\mathcal{L}$ , hence **branch and bound algorithm** is finitely terminating.

Because the above algorithm maintains the key invariant for **branch and bound**, i.e., every **feasible solution** of  $(D_{\mathcal{I}'})$  with greater objective value than LB is **feasible** for a problem on the list  $\mathcal{L}$  we have the following result.

**Theorem 8.2.1.** Suppose that  $(P)$  is feasible. Then at termination of **branch and bound**, we have  $\text{LB} = -\infty$  if  $(\tilde{D}_{\mathcal{I}'})$  is infeasible, or with  $\bar{y}_{\text{LB}}$  being an **optimal solution** of  $(D_{\mathcal{I}'})$ .

**Remark (Detail of the algorithm).** We make some remark on **branch and bound algorithm** we have above.

- When calling **line 8**, we first obtain the continuous relaxation  $(\tilde{D})$  of  $(D_{\mathcal{I}'})$  and its **primal**  $(\tilde{P})$  as follows.

$$\begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top \\ (\tilde{D}) & \end{array} \quad \begin{array}{ll} \min & c^\top x \\ & Ax = b. \\ (\tilde{P}) & x \geq 0 \end{array}$$

Then what we're really doing is solving  $(\tilde{P})$  instead of  $(\tilde{D})$  by **simplex algorithm** and get an **optimal** basis  $\beta$ , and this give us an **optimal dual solution**  $\bar{y}^\top := c_\beta^\top A_\beta^{-1}$ .

- When calling **line 16**, after choosing  $i^*$ , adding a constraint to  $(D_{\mathcal{I}'})$  effectively adds a variable to the corresponding continuous relaxation  $(\tilde{D})$ , hence adds a variable to the **standard form** problem  $(\tilde{P})$ . So, a **basis** for  $(\tilde{P})$  remains **feasible** after we introduce such a variable.
  - Down branch: The constraint  $y_{i^*} \leq \lfloor \bar{y}_{i^*} \rfloor$  dualize to a new variable  $x_{\text{down}}$  in  $(\tilde{P})$ , which has a new column  $A_{\text{down}} := e_{i^*}$  and a cost coefficient  $c_{\text{down}} := \lfloor \bar{y}_{i^*} \rfloor$ .

$$\begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top \\ (\tilde{D}) & y_{i^*} \leq \lfloor \bar{y}_{i^*} \rfloor \end{array} \quad \begin{array}{ll} \min & c^\top x + \lfloor \bar{y}_{i^*} \rfloor x_{\text{down}} \\ & Ax + e_{i^*} x_{\text{down}} = b \\ (\tilde{P}) & x \geq 0, x_{\text{down}} \geq 0. \end{array}$$

The **reduced cost** of  $x_{\text{down}}$  is

$$\bar{c}_{\text{down}} = c_{\text{down}} - \bar{y}^\top A_{\text{down}} = \lfloor \bar{y}_{i^*} \rfloor - \bar{y}^\top e_{i^*} = \lfloor \bar{y}_{i^*} \rfloor - \bar{y}_{i^*} < 0$$

since  $\bar{y}_{i^*}$  is not an integer. Hence,  $x_{\text{down}}$  is eligible to enter the **basis**.

- Up branch: Similarly, we have a new variable  $x_{\text{up}}$  in  $(\tilde{P})$  for the new constraint  $y_{i^*} \geq \lceil \bar{y}_{i^*} \rceil$ .

$$\begin{array}{ll} \max & y^\top b \\ & y^\top A \leq c^\top \\ (\tilde{D}) & y_{i^*} \geq \lceil \bar{y}_{i^*} \rceil \end{array} \quad \begin{array}{ll} \min & c^\top x - \lceil \bar{y}_{i^*} \rceil x_{\text{up}} \\ & Ax - e_{i^*} x_{\text{up}} = b \\ (\tilde{P}) & x \geq 0, x_{\text{up}} \geq 0. \end{array}$$

The **reduced cost** of  $x_{\text{up}}$  is

$$-\lceil \bar{y}_{i^*} \rceil - \bar{y}^\top (-e_{i^*}) = \bar{y}_{i^*} - \lceil \bar{y}_{i^*} \rceil < 0,$$

since  $\bar{y}_{i^*}$  is not an integer. Hence,  $x_{\text{up}}$  is eligible to enter the **basis**.

**Remark (Partially solving  $\tilde{P}$ ).** In practice, when solving  $(\tilde{P})$  (induced from any  $(D_{\mathcal{I}'})$ ) via **simplex algorithm**, we are generating a sequence of *decreasing objective values* of  $(\tilde{P})$ , each one of which is an upper-bound on the **optimal** value of its parent, and it's also a *potential new* LB. We see that when the **optimal** value of  $(\tilde{P}) \leq \text{LB}$ , we terminate immediately since solving this  $(\tilde{P})$  will not improve LB.

### 8.2.2 Global Upper Bound

Since in practice, there are many errors in the data, so we may just want to solve it approximately, which means we only want to get a global upper bound. Conceptually,

$$UB := \max \{LB, \max \{LP \text{ relaxation values for all problems on } \mathcal{L} \}\}$$

To calculate the set in the max, whenever children are created, solve their LP relaxation upon insertion into list. And we stop if

$$UB - LB < \text{absolute tolerance.}$$

**Remark.** Apparently, we see that we can do this by reordering the algorithm. But for the [original algorithm](#), we don't care about UB.

### 8.2.3 Node Selection

Node Selection means which problem to select from  $\mathcal{L}$  to process. There are several ways to do this.

1. FIFO (First In First Out)  $\cong$  BFS (Breadth First Search)<sup>1</sup>  
New problems go at the end of the list, select from the front. We see that this strategy will **maximize memory usage**.
2. LIFO (Last In First Out)  $\cong$  DFS (Depth First Search)<sup>2</sup>  
New problems go to the first of the list, select from the front. We see that this strategy will **increase LB quickly**.
3. Best Bound.  
Need the LP upper bound for all problems on the list. We see that this strategy will **decrease UB quickly**.

**Remark.** For any reasonable solver, it will first do the second strategy for several times, and they exclusively do the third strategy.

### 8.2.4 Branching Variable selection

1. Random: Choose randomly among  $y_i$  such that  $\bar{y}_i \notin \mathbb{Z}$ .
2. Biggest Cost: Choose based on the biggest  $c_i$ .
3. Most Fractional: Choose  $i$  with  $\bar{y}_i$  *most fractional*.
4. **Pseudo Cost Branching**

**Note.** Someone argues that the *most fractional* rules is as bad as choosing randomly.

## Lecture 24: Cutting Planes Algorithm

### 8.2.5 Cutting Planes Algorithm

06 Dec. 08:00

We focus on the problem in the form of

$$\begin{aligned} \max \quad & y^\top b \\ & y^\top A \leq c^\top \\ (D_{\mathfrak{X}}) \quad & y_i \text{ integer for } i = 1, \dots, m. \end{aligned}$$

<sup>1</sup>[https://en.wikipedia.org/wiki/Breadth-first\\_search](https://en.wikipedia.org/wiki/Breadth-first_search)

<sup>2</sup>[https://en.wikipedia.org/wiki/Depth-first\\_search](https://en.wikipedia.org/wiki/Depth-first_search)

**Note.** Compare to [subsection 8.2.1](#), now we require all  $y_i$  be integer, namely  $\mathcal{I} = \{1, \dots, m\}$  if we refer to the used notation.

Further, we also let  $(P)$  be

$$\begin{aligned} \min \quad & c^\top x \\ & Ax = b \\ (P) \quad & x \geq 0. \end{aligned}$$

**Remark.** We assume that the data  $A, c$  are all integer.

Now, we choose  $w \in \mathbb{R}^n$ ,  $w \geq 0$ . Then the constraint in  $(D_{\mathfrak{X}})$  becomes

$$y^\top (Aw) \leq c^\top w.$$

**Remark.** This valid for all  $y$  such that

$$y^\top A \leq c^\top,$$

no matter it's integer or not.

Suppose  $Aw \in \mathbb{Z}^m$ . With the fact that  $y \in \mathbb{Z}^m$ , then for

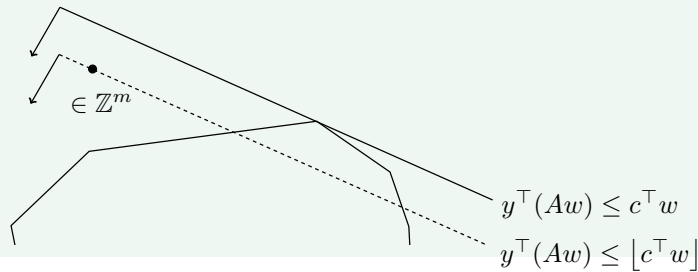
$$y^\top (Aw) \leq c^\top w,$$

we can actually get

$$y^\top (Aw) \leq \lfloor c^\top w \rfloor$$

for all integer solutions of  $y^\top A \leq c^\top$ .

**Definition 8.2.1 (Chvátal-Gomory cut).** The inequality  $y^\top (Aw) \leq \lfloor c^\top w \rfloor$  is called a *Chvátal-Gomory cut*.



**Remark.** To get a [Chvátal-Gomory cut](#), we need to choose  $w \in \mathbb{Z}^n$ , and we would like to develop a concrete algorithmic scheme for generating [Chvátal-Gomory cuts](#). We'll do this via [basic solutions](#).

We now solve  $(P)$  and get an [optimal basis](#)  $\beta$ . Consider the associated [dual basic solution](#)

$$\bar{y}^\top := c_\beta^\top A_\beta^{-1}$$

for the continuous relaxation of  $(D_{\mathfrak{X}})$ .

Notice that if  $\bar{y} \in \mathbb{Z}^m$ , then  $\bar{y}$  solves  $(D_{\mathfrak{X}})$ . Otherwise, suppose  $\bar{y}_i \notin \mathbb{Z}$ , then let

$$\tilde{b} := e_i + A_\beta r \in \mathbb{Z}^m,$$

where  $r \in \mathbb{Z}^m$ , our goal is to derive a valid [cut](#) for  $(D_{\mathfrak{X}})$  that is violated by  $\bar{y}$ .

**Theorem 8.2.2.**  $\bar{y}^\top \tilde{b} \notin \mathbb{Z}$ , and so

$$y^\top \tilde{b} \leq \lfloor \bar{y}^\top \tilde{b} \rfloor$$

cuts off  $\bar{y}$ .

**Proof.** Since

$$\bar{y}^\top \tilde{b} = \bar{y}^\top (e_i + A_\beta r) = \bar{y}_i + \bar{y}^\top A_\beta r = \bar{y}_i + c_\beta^\top A_\beta^{-1} A_\beta r = \bar{y}_i + c_\beta^\top r$$

We see that  $\bar{y}_i \notin \mathbb{Z}$ ,  $c_\beta^\top, r \in \mathbb{Z}$ , hence we have

$$\bar{y}^\top \tilde{b} = \bar{y}_i + c_\beta^\top r \notin \mathbb{Z}.$$

Now, we need to check that  $y^\top \tilde{b} \geq \lfloor \bar{y}^\top \tilde{b} \rfloor$  is satisfied by  $\bar{y}$ .

**Intuition.** Consider if the inequality is

$$\vec{0}^\top y \leq -1,$$

then it makes no sense.

Hence, since  $y^\top \tilde{b} \geq \lfloor \bar{y}^\top \tilde{b} \rfloor$ ,  $y^\top \tilde{b} \leq \lfloor \bar{y}^\top \tilde{b} \rfloor$  indeed cuts off  $\bar{y}$ . ■

We now need to show that the inequality  $y^\top \tilde{b} \leq \lfloor \bar{y}^\top \tilde{b} \rfloor$  is valid for  $(D_{\mathfrak{X}})$ . Let  $H := A_\beta^{-1}$ , then  $H_{\cdot i} = A_\beta^{-1} e_i$ . Further, we let  $w := H_{\cdot i} + r$ . Since we need  $w \geq \vec{0}$ , we can always choose  $r \in \mathbb{Z}^m$  so that  $w \geq \vec{0}$ . Specifically, we choose

$$r_k \geq -\lfloor h_{ki} \rfloor$$

for  $k = 1, \dots, m$ . Then, we have the following theorem.

**Theorem 8.2.3.** Choosing  $r \in \mathbb{Z}^m$  satisfying  $r_k \geq -\lfloor h_{ki} \rfloor$  for all  $k$ , we have

$$y^\top \tilde{b} \leq \lfloor \bar{y}^\top \tilde{b} \rfloor$$

is valid for  $(D_{\mathfrak{X}})$ .

**Proof.** Because  $y^\top A \leq c^\top$ , we have  $y^\top A_\beta \leq c_\beta^\top$ . Then we have

$$(y^\top A_\beta) (H_{\cdot i} + r) \leq c_\beta^\top (H_{\cdot i} + r).$$

This is equivalence to

$$(y^\top A_\beta) (A_\beta^{-1} e_i + r) \leq c_\beta^\top (A_\beta^{-1} e_i + r).$$

After expanding, we have

$$y_i + y^\top A_\beta r \leq \bar{y}_i + c_\beta^\top r,$$

which can be written as

$$y^\top \underbrace{(e_i + A_\beta r)}_{\tilde{b}} \leq \bar{y}^\top \underbrace{(e_i + A_\beta r)}_{\tilde{b}}$$

since  $\bar{y}^\top = c_\beta^\top A_\beta^{-1}$ . Then, since  $\tilde{b} \in \mathbb{Z}^m$  and  $y$  is constrained to be in  $\mathbb{Z}^m$  for  $(D_{\mathfrak{X}})$ , we see

$$y^\top \tilde{b} \leq \lfloor \bar{y}^\top \tilde{b} \rfloor.$$

■

Now, given any non-integer [basic dual solution](#)  $\bar{y}$ , we produce a valid inequality for  $(D_{\mathfrak{X}})$  from [Theorem 8.2.2](#) that cuts it off.

**Note.** This [cut](#) for  $(D_{\mathfrak{X}})$  is used as a column for  $(P)$ : the column is  $\tilde{b}$  with objective coefficient  $\lfloor \bar{y}^\top \tilde{b} \rfloor$ . Taking  $\beta$  to be an [optimal basis](#) for  $(P)$ , the new variable corresponding to this column is the unique variable eligible to enter the [basis](#) in the context of the [simplex algorithm](#) applied to  $(P)$  since the [reduced cost](#) is precisely

$$\lfloor \bar{y}^\top \tilde{b} \rfloor - \bar{y}^\top \tilde{b} < 0.$$

Also, the new column for  $A$  is  $\tilde{b}$  which is integer, and the new objective coefficient for  $c$  is  $\lfloor \bar{y}^\top \tilde{b} \rfloor$ , which is also an integer. So the original assumption that  $A$  and  $c$  are integer is maintained. Lastly, we need  $Aw$  are all integers. This is true since

$$A_\beta w = A_\beta \left( A_\beta^{-1} e_i + r \right) = e_i + A_\beta r \in \mathbb{Z}^m.$$

We can then repeatedly add new **cuts**.<sup>a</sup>

<sup>a</sup>Though we do our computations as column generation with respect to  $(P)$ .

We note that there is clearly a lot of flexibility in how to choose  $r$ . The next (last!) theorem shows that it's always best to choose a minimal  $r \in \mathbb{Z}^m$  satisfying  $r_k \geq -\lfloor h_{ki} \rfloor$  for all  $k = 1, \dots, m$ .

**Theorem 8.2.4.** Let  $r \in \mathbb{Z}^m$  be defined by

$$r_k = -\lfloor h_{ki} \rfloor, \text{ for } k = 1, \dots, m.$$

Suppose that  $\hat{r} \in \mathbb{Z}^m$  satisfies  $\hat{r}_k \geq -\lfloor h_{ki} \rfloor$  for all  $k = 1, \dots, m$ , then the **cut** determined by  $r$  dominates the **cut** determined by  $\hat{r}$ .

**Proof.** Obviously,  $r \leq \hat{r}$ . It's then easy to check that the **cut** can be re-expressed as

$$y_i \leq \lfloor \bar{y}_i \rfloor + (c_\beta^\top - y^\top A_\beta) r.$$

Noting that  $c_\beta^\top - y^\top A_\beta \geq \vec{0}$  for all  $y$  that are feasible for the continuous relaxation of  $(D_{\mathcal{X}})$ , we see that the strongest inequality is obtained by choosing  $r \in \mathbb{Z}^m$  to be minimal. ■

Revisiting the **example**. Now we see that the cutting plane algorithm will need at least  $2k$  steps for such a triangle with height  $k$ , since it can only cut off one point at a time.

**Example.** Now we see some bad examples for Branch and Bound. Consider the following integer programming problem.

$$\begin{aligned} \min \quad & y_{n+1} \\ & 2y_1 + 2y_2 + \dots + 2y_n + y_{n+1} = \underbrace{n}_{\text{odd}} \\ & 0 \leq y_i \leq 1 \text{ for } i = 1, \dots, n+1, \text{ integer.} \end{aligned}$$

We see that the optimum has  $y_{n+1} = 1$ .

If  $n = 17$ . Then we can let

$$y_{18} = 0, \quad y_1 = y_2 = \dots = y_8 = 1, \quad y_9 = \frac{1}{2}, \quad y_{10} = \dots = y_{17} = 0.$$

We immediately see there are lots of solutions like this, namely there are lots of symmetric groups going on such that half of the variables are 1, and another half of the variables are 0. This is pretty bad for the branch and bound algorithm since it will look at all of them. Analytically, we see that this will go into  $\frac{n}{2}$  depth in the recursion tree, hence it's clearly exponential.

**Remark (Pure v.s. Mixed).** We only consider pure integer programming under the Gomory cutting plane scheme here, for mixed case, please refer to [Lee22].

**Remark (Finite termination).** To prove that Gomory cutting plane scheme is finitely terminating is rather technical in either pure or mixed case.<sup>a</sup>

We need to treat the **objective function** value as an additional variable (numbered first), employ the **simplex algorithm** adapted to the  $\epsilon$ -perturbed problem, always choose the least-index  $i \in \mathcal{I}$  having  $\bar{y}_i \notin \mathbb{Z}$  and choose  $r$  via **Theorem 8.2.4**<sup>b</sup> as appropriate to generate the **cuts**.

<sup>a</sup>Though it's done in essentially the same manner.

<sup>b</sup>Or another rule for mixed case, see [Lee22]

---

**Remark (Branch-and-Cut).** SOTA algorithms for (mixed-)integer linear optimization (like **Gurobi**, **Cplex**) combine cuts with branch and bound, though it's too technical to make this work, so we'll skip it.



# Appendix

---

This note is completed in  $\text{\LaTeX}$  with Inkscape, in case of anyone is interested, please check out this blog<sup>3</sup> together with my configuration<sup>4</sup>.

---

<sup>3</sup><https://castel.dev/>

<sup>4</sup><https://github.com/sleepymalc/VSCoDe-LaTeX-Inkscape>

# Bibliography

- [Lee22] Jon Lee. *A First Course in Linear Optimization*". Reex Press, 2022. URL: [https://github.com/jon77lee/JLee\\_LinearOptimizationBook](https://github.com/jon77lee/JLee_LinearOptimizationBook).