ISyE 6661 (Fall 2016)

Linear Programming

Prof. S. Ahmed
Georgia Institute of Technology

LaTeXer: W. Kong
http://wwkong.github.io
Last Revision: December 12, 2016

Table of Contents

These notes are currently a work in progress, and as such may be incomplete or contain errors.

Acknowledgments:

Special thanks to *Michael Baker* and his LaTeX formatted notes. They were the inspiration for the structure of these notes.

Abstract

The purpose of these notes is to provide the reader with a secondary reference to the material covered in ISyE 6661.

1 Optimization Models

Example 1.1. Given a set of bonds $i = 1, ..., n$ and planning horizon $t = 1, ..., T$, define

$$\begin{aligned} C_{it} &= \text{payment of bond } i \text{ in year } t \\ L_t &= \text{liability in year } t \\ r_t &= \text{interest rate in year } t \\ p_i &= \text{price of bond } i \end{aligned}$$

How many units of bond i should I buy to pay my liabilities? Minimize my costs? For the first part, the constraints are

$$\begin{aligned} (1+r_t)Z_{t-1} + \sum_{i=1}^{n} C_{it}x_i = L_t + z_t, t = 1, ..., T \\ x_i \geq 0, i = 1, ..., n \\ z_t \geq 0, t = 1, ..., T \end{aligned}$$

and the objective is

$$\underset{x,z}{\text{minimize}}\ z_0 + \sum_{i=1}^{n} p_i x_i$$

where z_0 is the initial cash flow, z_t is the cash remaining at the end of year t, and x_i is the number of bond i to buy.

Definition 1.1. In general, the set up for an optimization problem is

$$\begin{aligned} (P)\ \min_x\ & f(x) \\ \text{s.t. } & x \in X \end{aligned}$$

where $X \subseteq \mathbb{R}^n$ is the set of allowed values (**constraints**), $x \in \mathbb{R}^n$ is a **decision vector**, and $f : \mathbb{R}^n \mapsto \mathbb{R}$ is called the **objective function**. In this class, we will only discuss

(i) finite dimensional decisions

(ii) single objectives

(iii) minimization problems (maximization is $-\max(-f(x))$)

There are several outcomes:

(i) Infeasible: $X = \emptyset$

(ii) Unbounded: $\exists \{x^i\} \subseteq X$ s.t. $f(x^i) \to -\infty$

(iii) Bounded but minimizer is not achieved (e.g. $\min\{x : x \in (0, \infty)\}$ d.n.e.)

(iv) An optimal solution exists

Example 1.2. Some examples of the forms of X are:

$$\begin{aligned} X_1 &= \{x \in \mathbb{R}^n : g_i(x) \leq 0, i = 1, ..., m\} \\ X_2 &= \{x \in \mathbb{R}^n : \exists y, h(x, y) \geq 0\} \end{aligned}$$

Definition 1.2. A **feasible solution** $\hat{x}$ is such that $\hat{x} \in X$; a **globally optimal solution** is a feasible solution such that $f(\hat{x}) \leq f(x)$ for all $x \in X$. A **locally optimal solution** is a feasible solution such that $\exists \epsilon > 0$ with

$$f(\hat{x}) \leq f(x), \forall x \in X \cap \mathbb{B}(\hat{x}, \epsilon)$$

where $\mathbb{B}(\hat{x}, \epsilon) := \{x : \|x - \hat{x}\| \leq \epsilon\}$. The **optimal value** is $\min\ f(x)$ s.t. $x \in X$.

Theorem 1.1. *In problem (P) if $f : \mathbb{R}^n \to \mathbb{R}$ is continuous and the set X is nonempty, closed, and bounded then (P) has an optimal solution. For this class, a set X is closed if for all convergent sequence sequences in X the limit points are contained in X.*

Proof. Suppose that instead $\exists\{x^i\}_{i\in\mathbb{N}} \subseteq X$ with $f(x^i) \leq -i$. Since f is bounded, then by the Bolzano-Weierstrass theorem let $\{x^{i_k}\}_{k\in\mathbb{N}}$ be a convergent subsequence in X. Then $\lim_{n\to\infty} x^{i_n} = x^* \in X$ by closure. Then $f(x^*) = \lim_{n\to\infty} f(x^{i_n}) \leq \lim_{n\to\infty} -i_k = -\infty$ which is impossible. So now $\exists l = \inf f(x)$ such that $x \in X$. For $\epsilon > 0$ define

$$S^k = \{x \in X : l \leq f(x) \leq l + \epsilon^k\} \neq \emptyset, k = 1, 2, ...$$

Pick $x^k \in S^k \implies \{x^k\} \subset X$ and hence by the Bolzano-Weierstrass, $\exists\{x^{k_i}\}$ which converges in X. By the Squeeze Theorem,

$$l \leq \lim_{i\to\infty} f(x^{k_i}) \leq l + \lim_{i\to\infty} \epsilon^{k_i} \implies f(x) = l$$

□

Definition 1.3. If I know a lower bound LB for $\min_x\{f(x) : x \in X\}$ and I have a solution $\hat{x} \in X$, define

$$0 \leq \text{gap}(\hat{x}) = f(\hat{x}) - v^* \leq f(\hat{x}) - LB$$

By convention, $v^* = \infty$ if (P) is infeasible, $v^* = -\infty$ if (P) is unbounded and a real number otherwise. Also define the **relaxation of (P)** as

$$\begin{aligned}(Q) \min f'(x)\\ s.t\ x \in X'\end{aligned}$$

if $f'(x) \leq f(x)$ for all $x \in X$ and $X' \supseteq X$.

Example 1.3. Consider the problem

$$\begin{aligned}(P)\ \min_x\ f(x)\\ \text{s.t. } g_i(x) \leq b_i,\ i = 1...m\end{aligned}$$

Let $\mu_i \geq 0, i = 1, ..., m$ and define

$$L(\mu) = \min_x f(x) + \sum_{i=1}^m \mu[g_i(x) - b_i]$$

which is called the **Lagrangian relaxation**. To find the best lower bound, we solve the problem

$$\sup_{\mu\geq 0} L(\mu) \leq v^*$$

which is called the **dual problem** and the above states a **weak duality**. Suppose we have a pair (x^*, μ^*) such that $L(\mu^*) = f(x^*)$ and $x^* \in X$. Then we have an optimal solution.

1.1 Convexity

Definition 1.4. Given a collection of vectors $x_1, ..., x_k \in \mathbb{R}^n$, an **affine combination of vectors** is $\sum_{i=1}^k \lambda_i x_i$ where $\sum_{i=1}^k \lambda_i = 1$, a **conic combination of vectors** is of the same form but $\lambda_i \geq 0$ for $i = 0, 1, ..., k$. A **convex combination of vectors** is both an affine and conic combination of vectors.

Definition 1.5. A **convex** function $f : \mathbb{R}^n \to \mathbb{R}$ is such that

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

for any $x, y \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. **Jensen's inequality** is a result of the above property:

$$f\left(\sum_{i=1}^k \lambda_i x_i\right) \leq \sum_{i=1}^k \lambda_i f(x_i)$$

A **concave** function is one where the convex inequality is flipped.

Proposition 1.1. *If f is differentiable, then*

$$f \text{ is convex} \iff f(y) \geq f(x) + \nabla f(x)^T (y-x)$$

for any $x, y \in \mathbb{R}^n$.

Proof. Define

$$f'(x;d) = \lim_{\epsilon \to 0^+} \frac{f(x+\epsilon d) - f(x)}{\epsilon} = \nabla f(x)^T d$$

($\implies$) Pick $x, y \in \mathbb{R}^n$ and $\lambda \in (0,1)$ and remark

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

Taking limits on $\lambda \to 0^+$ gives us the result. The converse is an assignment question. □

Remark 1.1. (Calculus of convex functions) The following are convex functions for given convex functions f_i:

(i) $\sum_i \lambda_i f_i(x), \lambda_i \geq 0$

(ii) $\max_i f_i(x)$

(ii)$h(f_1(x), f_2(x), ..., f_m(x))$ where h is convex and non-decreasing in each component where the second condition is not needed if f_i are affine functions

[See lecture 3 notes for more details]

Theorem 1.2. *(Separation theorem) Let $X \subseteq \mathbb{R}^n$ be a nonempty closed convex set. If $\hat{x} \notin X$ then there exists a hyperplane that separates $\hat{x}$ from X. That is, there exists $(\pi, \pi_0) \in \mathbb{R}^{n+1}$ such that*

$$\pi^T \hat{x} < \pi_0 \text{ and } \pi^T x \geq \pi_0, \; \forall x \in X$$

Proof. Since $X \neq \emptyset$ pick $w \in X$ and let $\beta = \|\hat{x} - w\|_2 > 0$ with $X' = X \cap B(\hat{x}, \beta) \neq \emptyset$ which is non-empty, bounded, closed. Consider the problem

$$\begin{aligned} (P) \min \; & \|x - \hat{x}\|_2 \\ s.t \; & x \in X' \end{aligned}$$

The Weierstrass theorem says that there is a minimizer $x^* \in X$ and hence

$$\|x^* - \hat{x}\|_2 \leq \|x - \hat{x}\|_2, \forall x \in X' \text{ and } \|x^* - \hat{x}\|_2 \leq \|x - \hat{x}\|_2, \forall x \in X$$

Since $\hat{x} \notin X$ then

$$\begin{aligned} \|\hat{x} - x^*\|_2 > 0 &\implies (\hat{x} - x^*)^T (\hat{x} - x^*) > 0 \\ &\implies (\hat{x} - x^*)^T \hat{x} > (\hat{x} - x^*)^T x^* \\ &\implies \underbrace{(x^* - \hat{x})^T}_{\pi^T} \hat{x} < \underbrace{(x^* - \hat{x})^T x^*}_{\pi_0} \\ &\implies \pi^T \hat{x} < \pi_0 \end{aligned}$$

Now pick $x \in X$ and consider $X(\lambda) = x^* + \lambda(x - x^*)$ for all $\lambda \in (0,1)$. By convexity, $X(\lambda) \in X, \forall \lambda \in (0,1)$ we have

$$\begin{aligned} \|x^* - \hat{x}\|_2^2 &\leq \|X(\lambda) - \hat{x}\|_2^2 \\ &= \|x^* + \lambda(x - x^*) - \hat{x}\|_2^2 \\ &= \|(x^* - \hat{x}) + \lambda(x - x^*)\|_2^2 \\ &= \|x^* - \hat{x}\|_2^2 + \lambda^2 \|x - \hat{x}\|_2^2 + 2\lambda(x - \hat{x})^T (x - x^*) \end{aligned}$$

and hence

$$
\begin{aligned}
&\implies \lambda^2\|x-\hat{x}\|_2^2 + 2\lambda(x-\hat{x})^T(x-x^*) \geq 0 \\
&\implies \frac{\lambda}{2}\|x-\hat{x}\|_2^2 + (x-\hat{x})^T(x-x^*) \geq 0 \\
&\implies \lim_{\lambda\to 0}\frac{\lambda}{2}\|x-\hat{x}\|_2^2 + (x-\hat{x})^T(x-x^*) \geq 0 \\
&\implies (x-\hat{x})^T(x-x^*) \geq 0 \\
&\implies (x^*-\hat{x})^T x \geq (x^*-\hat{x})^T x^* \\
&\implies \pi^T x \geq \pi_0
\end{aligned}
$$

□

[See lecture 4 notes for more details]

Example 1.4. Consider the airline problem of selling tickets where x_n is how many tickets to sell in each of the n fare classes, r_i is the revenue for a ticket solve in fare class i, c_i is the number of seats in fare class i, p_i is the penalty for each passenger $\geq$ capacity c_i, and a total of T tickets can be sold.

We also have m scenarios where in scenario k, α_{ik} is the proportion of passengers that show up and π_k is the probability of scenario k. We wish to maximize the expected profit. This can be formulated as

$$
\begin{aligned}
\text{maximize } & \sum_{i=1}^{n} x_i r_i - P \\
\text{s.t. } & P = \left(\sum_{k=1}^{m}\pi_k\left[\sum_{i=1}^{n} p_i \max(0, \alpha_{ik}x_i - c_i)\right]\right) \\
& \sum_{i-1}^{n} x_i \leq T \\
& x_i \geq 0,\ i = 1, 2, ..., n
\end{aligned}
$$

the first constraint and objective function are non-linear but they can be made linear through the following transformation:

$$
\begin{aligned}
\text{maximize } & \sum_{i=1}^{n}\left(x_i r_i - \sum_{k=1}^{m}\pi_k\left[\sum_{i=1}^{n} p_i y_{ik}\right]\right) \\
\text{s.t. } & y_{ik} \geq 0 \\
& y_{ik} \geq \alpha_{ik}x_i - c_i,\ i = 1, 2, ..., n,\ k = 1, 2, ..., m \\
& \sum_{i-1}^{n} x_i \leq T \\
& x_i \geq 0,\ i = 1, 2, ..., n
\end{aligned}
$$

In the notes, consider the 3 equivalences for epi g where $g = \|\cdot\|_1$. (1) to (2) comes from the fact that $w_j = \max\{x_j, -x_j\}$ and writing out the $\subseteq$ and $\supseteq$ proofs. For (1) to (3), we do something similar except using the transforms

$$
u_j' = \begin{cases} \hat{x}_j & \hat{x}_j \geq 0 \\ 0 & o/w \end{cases}
$$

$$
{v_j}' = \begin{cases} -\hat{x}_j & \hat{x}_j \leq 0 \\ 0 & o/w \end{cases}
$$

For the next (univariate) convex function $g : [b_0, b_k] \mapsto \mathbb{R}$ with $b_i < b_{i+1}$ and convexity implying $c_i \leq c_{i+1}$ for any i. Now,

$$g(x) = \max_{i=1...k} \{g(b_{i-1}) + c_i(x - b_{i-1})\}$$
$$\implies \text{epi } g = \{(y, x) : y \geq g(b_{i-1}) + c_i(x - b_{i-1}), \forall i = 1, ..., k, b_0 \leq x \leq b_k\}$$

Here is an alternate formulation (proved below):

$$\text{epi } g = \left\{(y, x) : y \geq g(b_0) + \sum_{i=1}^{k} c_i z_i, 0 \leq z_i \leq b_i - b_{i-1}, x = b_0 + \sum_{i=1}^{k} z_i, \forall i = 1, ..., k\right\}$$

Proof. ($\subseteq$) Call the set on the rhs$S$ in the space of (y, x, z). Let $(\hat{x}, \hat{y}) \in$ epi g in the original formulation. Then we create the transformation

$$\hat{z}_i = \begin{cases} b_i - b_{i-1} & b_i < \hat{x} \\ \hat{x} - b_{i-1} & b_{i-1} \leq \hat{x} \leq b_i \\ 0 & \hat{x} > b_{i-1} \end{cases}$$

Only the first set of inequalities in the alternate construction needs to be checked as the others are trivially true. Note that

$$g(b_0) + \sum_{i=1}^{k} c_i z_i = g(b_0) + \sum_{i:b_i<\hat{x}} (g(b_i) - g(b_{i-1})) + \sum_{i:b_{i-1}\leq\hat{x}\leq b_i} \left(\frac{g(b_i) - g(b_{i-1})}{b_i - b_{i-1}}\right)(\hat{x} - b_{i-1}) + \sum_{i:\hat{x}<b_{i-1}} c_i \cdot 0$$
$$= g(\hat{x}) \leq \hat{y}$$

which satisfies the first set of inequalities.

($\supseteq$) Let $(\hat{x}, \hat{y}, \hat{z}) \in S$ and suppose $b_{\hat{i}-1} \leq \hat{x} \leq b_{\hat{i}}$. Since $\hat{y} \geq g(b_0) + \sum_{i=1}^{k} c_i \hat{z}_i$, suppose that $\hat{z}_i < b_i - b_{i-1}$ and $\hat{z}_j > 0$ where $i < j$. Construct

$$\Delta_{ij} = \min\{(b_i - b_{i-1}) - \hat{z}_i, \hat{z}_j\}, \hat{\hat{z}}_i = \hat{z}_i + \Delta_{ij}, \hat{\hat{z}}_j = \hat{z}_j - \Delta_{ij}$$

and note that $(\hat{x}, \hat{y}, \hat{\hat{z}}) \in S$ with the last set of inequalities due to

$$g(b_0) + \sum_{i=1}^{k} c_i \hat{\hat{z}}_i \leq g(b_0) + \sum_{i=1}^{k} c_i \hat{z}_i \leq \hat{y}$$

from convexity. We can iterate this procedure (a finite amount of times) in order to re-align the z_i's such that the new formulation equals the original epi g. $\square$

Example 1.5. In the last part of the lecture package (fractional programming), if we are given

$$(P) \quad \min_x \frac{p^T x + p_0}{q^T x + q_0}$$
$$\text{s.t. } Ax \geq b$$

with $q^T x + q_0 > 0, \forall x : Ax \geq b$. Set

$$t = \frac{1}{q^T x + q_0} > 0 \iff q^T(tx) + q_0 t = 1, z = p^T(tx) + p_0 t, Atx \geq bt$$

If $y = tX$ then the original problem becomes

$$(Q) \quad \min_{y,t} p^T y + p_0 t$$
$$\text{s.t.} Ay \geq bt$$
$$q^T y + q_0 t = 1$$
$$t \geq 0$$

Theorem 1.3. *If (P) has an optimal solution x^* then we can construct $t^* = 1/(q^T x^* + q_0), y_j^* = tx_j^*$ then (y^*, t^*) is an optimal solution of (Q).*

Remark 1.2. For a given non-decreasing transform ϕ we have $\max f(x) = \max \phi(f(x))$ and for geometric programming, we generally use $\phi = \log$.

[See lecture 5 notes for more details]

Theorem 1.4. *If (P) is feasible and bounded, then it has an optimal solution.*

Proof. Consider the polyhedral set

$$S = \left\{(x, z) \in \mathbb{R}^n \times \mathbb{R}, c^T x \leq z, z \leq c^T x, Ax \geq b\right\}$$

use Fourier-Motzkin to project S to the space of z-variables and call the projected set $Z \subseteq \mathbb{R}$ which is a polyhedral set. There exists $\alpha \leq \inf\{z : z \in Z\}$ since z^* is closed. We map back to our original space to obtain a feasible optimal solution. ☐

Example 1.6. Consider the problem

$$\begin{aligned} \min\ & -x_1 - 4x_2 \\ s.t.\ & x_1 + x_2 \leq 2 \\ & x_2 \leq 1 \\ & x_1, x_2 \geq 0 \end{aligned}$$

If we set up $z \leq -x_1 - 4x_2$ and $z \geq -x_1 - 4x_2$ then by repeated application of Fourier-Motzkin,

$$\frac{1}{3}(-z-2) \leq -\frac{1}{4}z, 0 \leq -\frac{1}{4}z, \frac{1}{3}(z-2) \leq 1 \implies -8 \leq z, -5 \leq z, z \leq 0$$

and $z^* = -5$ which gives:

$$1 \leq x_2 \leq \frac{5}{4}, x_2 \leq 1 \implies x_2^* = 1$$

and so on for x_1.

[See lecture 6+7 notes for more details]

Proof. Consider the proposition of $\dim(X) = n - \text{rank}(A)$ for $X = \{x \in \mathbb{R}^n : Ax = b, Cx \leq d\}$ and there exists $\hat{x} \in X$ such that $C\hat{x} < d$. Here is the proof. Let $L = \{x : Ax = 0\}$ and pick $\{x^1, ..., x^p\}$ be linearly independent vectors in L where $p = n - \text{rank}(A)$. Consider the points$\hat{x}^i = \hat{x} + \epsilon \cdot x^i$ with $\epsilon > 0$ small enough so that $C\hat{x}^i \leq d$. The points $\{\hat{x}, \hat{x}^1, ..., \hat{x}^p\} \subseteq X$ where you can show that these points are affinely independent (by choice of $x^1, ..., x^p$) and hence

$$\dim(X) \geq p = n - \text{rank}(A)$$

If $\dim(X) = k$ then $\{x^1, ..., x^{k+1}\}$ are affinely independent points in X and they satisfy $Ax^i = b$ for $i = 1, ..., k+1$ and from the previous property $k + 1 \leq n + 1 - m \implies k \leq n - m$ where $k = \text{rank}(A)$. ☐

Proof. Here is a proof of Caratheodory's Theorem. Let $\hat{x} \in \text{conv}(X)$ and suppose that $\hat{x} = \sum_{i=1}^s \lambda_i x^i$ where this is a convex combination of elements in X. Assume that s is the smallest number that allows such a representation $\implies \lambda_i > 0, i = 1, 2, ..., s$. If $s \leq k + 1$ we are done, so instead if it is not consider the following system:

$$(*) \begin{cases} \sum_{i=1}^s \alpha_i x^i \\ \sum_{i=1}^s \alpha_i = 0 \end{cases}$$

where $\{x^i\}_{i=1}^s$ are vectors in the aff(X) whose dimension is k. These vectors cannot be affinely independent (a.i.) so the system has a non-trivial solution. Call such a solution $\bar{\alpha}_i, i = 1, ..., s$ and so that $\exists \hat{i} : \bar{\alpha}_{\hat{i}} \neq 0$ and define numbers $\mu_i(t) = \lambda_i + t\bar{\alpha}_i i = 1, ..., s$ where we have

$$\sum_{i=1}^s \mu_i(t) = \sum_{i=1}^s \lambda_i + t \sum_{i=1}^s \bar{\alpha}_i = 1$$

Note that I can choose $t \in \mathbb{R}$ small enough so that $\mu_i(t) \geq 0$. Choose a $t^* \in \mathbb{R}$ such that $\mu_i(t^*)$ such that $\mu_i(t^*) \geq 0$ for all i and $\mu_{\bar{i}}(t^*) = 0$ for sum $\bar{i}$. Then

$$\sum_{i=1}^{s} \mu_i(t^*)x^i = \sum_{i=1}^{s} \lambda_i x^i + t^* \sum_{i=1}^{s} \bar{\alpha}_i = \sum_{i=1}^{s} \lambda_i x^i = \hat{x}$$

and hence we have constructed a system of $(s-1)$ multipliers. Repeat until we get $s \leq k+1$. □

Proof. Here is a proof of Radon's theorem. Let $k \geq n+2$ and consider the system

$$(*) \begin{cases} \sum_{i=1}^{k} \alpha_i x^i = 0 \\ \sum_{i=1}^{k} \alpha_i = 0 \end{cases}$$

$(*)$ has a nontrivial solution $\bar{\alpha}_i$. Let

$$I = \{i : \bar{\alpha}_i > 0\}$$
$$J = \{i : \bar{\alpha}_i \leq 0\}$$

and clearly I, J are nonempty and constitute a partition of $\{1, ..., k\}$. Let

$$S = \sum_{i \in I} \bar{\alpha}_i = \sum_{i \in J} (-\bar{\alpha}_i) > 0$$

Consider

$$\begin{aligned} \hat{y} &= \sum_{i \in I} \left(\frac{\hat{\alpha}_i}{S}\right) x^i \in \text{conv}\left(\{x^i : i \in I\}\right) \\ &= \sum_{j \in I} \left(\frac{-\hat{\alpha}_i}{S}\right) x^i \in \text{conv}\left(\{x^i : i \in J\}\right) \end{aligned}$$

and we are done. □

Proof. Here is a proof of Helley's Theorem (note that there is a typo in the statement; we need $n = d$). In $\mathbb{R}^d$ assume that the claim holds for all collections of size $k-1$ and note that if $k \leq d+1$ the theorem holds trivially. Assume that instead $k \geq d+2$, and construct the sets

$$Y_i = \bigcap_{j=1, j \neq i}^{k} X_j \neq \emptyset$$

and pick $x^i \in Y_i, i = 1, ..., k \geq d+2$. By Radon's Theorem, we can partition these points into two sets whose convex hulls intersect. After re-indexing we have

$$\underbrace{x^1, x^2, ..., x^l}_{A} \underbrace{x^{l+1}, x^{l+2}, ..., x^k}_{B}$$

where $\text{conv}(A) \cap \text{conv}(B) \neq \emptyset$. Pick $\hat{y} \in \text{conv}(A) \cap \text{conv}(B)$ and we claim that $\hat{y} \in \bigcap_{i=1}^{k} X_i$. To see this, note that for $1 \leq i \leq l : x^i \in Y_i \subseteq \bigcap_{j=l+1}^{k} X_j$ then $\hat{y} \in \bigcap_{j=l+1}^{k} X_j$ and similarly $1 \leq i \leq l : x^i \in \bigcap_{j=l}^{l} X_j$ then $\hat{y} \in \bigcap_{j=1}^{l} X_j$ and so $\hat{y} \in \bigcap_{j=1}^{k} X_j$. □

1.2 Farkas' Lemma

[See lecture 8 notes for more details]

Proof. Here is a proof of Farkas' Lemma using the Separation Theorem. Suppose that $P \neq \emptyset$ and pick $\hat{x}$ such that $A\hat{x} = b, \hat{x} \geq 0$ and $\hat{y} \in Q \subseteq \mathbb{R}^m$ with $\hat{y}^T A \hat{x} = \hat{y}^T b < 0$ and note that

$$\underbrace{\hat{y}^T A}_{\geq 0} \underbrace{x}_{\geq 0} \geq 0$$

which is a contradiction. Therefore $P \neq \emptyset \implies Q = \emptyset$. We will show that $P = \emptyset \implies Q \neq \emptyset$. Let's define $U = \{u \in \mathbb{R}^n : Ax = u, x \geq 0\}$ and note that $P = \emptyset \implies b \notin U$. Since U is a non-empty closed, convex set, then by the separation theorem, there exists $(\pi, \pi_0) \in \mathbb{R}^{m+1}$ such that $\pi^T b < \pi_0$ and $\pi^T u \geq \pi_0$ for all $u \in U$. Since $\pi_0 \leq 0$ then $\pi^T b < 0$. Note that $a^j \in U, \forall i = 1, ..., n$ and $\lambda a^j \in U, \forall \lambda > 0$. Now

$$\pi^T(\lambda a^j) \geq \pi_0, j = 1, 2, ..., n$$

and rearranging

$$\pi^T a^j \geq \frac{1}{\lambda}\pi_0, \forall \lambda > 0, j = 1, ..., n \implies \lim_{\lambda \to \infty}\left(\pi^T a^j \geq \frac{1}{\lambda}\pi_0\right) \implies \pi^T A \geq 0^T \implies A^T \pi \geq 0$$

□

Remark 1.3. Given a system $(*)$ $Ax \geq b$ where $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$, a single step of Fourier-Motzkin (F-M) to eliminate a variable, say x_n, is equivalent to multiplying system $(*)$ by a non-negative matrix $M \in \mathbb{R}^{k \times m}$ to get $(**)$ $MAx \geq Mb$. such that the n^{th} column of MA has all zeroes. A vector $(x_1, ..., x_n)$ is solution of $(*)$ if $(x_1, ..., x_{n-1})$ is a solution of $(**)$ and $(x_1, ..., x_{n-1})$ is a solution of $(**)$ if $\exists x_n$ such that $(x_1, ..., x_n)$ is a solution of $(*)$.

Proof. Here is a proof of Farkas' Lemma using F-M. Suppose that $P = \emptyset$ and note that $Ax = b, x \geq 0 \iff Ax \geq b, -Ax \geq b, Ix \geq 0 \iff (*)$ $\tilde{A}x \geq \tilde{b}$ where $\tilde{A} = [A^T, -A^T, I^T]^T$ and $\tilde{b} = [b^T, -b^T, 0^T]^T$. Do F-M on $(*)$ to get $M\tilde{A}x \geq M\tilde{b}$ where $M \in \mathbb{R}^{k \times (2m+n)}$ and $M\tilde{A} = [0] \implies 0 \geq M\tilde{b} \in \mathbb{R}^{k \times 1}$. Since $(*)$ is infeasible ($P = \emptyset$) there must be at least one component i of $M\tilde{b}$ that is positive (i.e. $[M\tilde{b}]_i > 0 \implies M$ has a row m_i^T such that $m_i^T \tilde{b} > 0$). Now

$$\begin{aligned} 0 < m_i^T \tilde{b} &= [m_{i1}^T, m_{i2}^T, m_{i3}^T][b^T, -b^T, 0]^T \\ &= (m_{i1} - m_{i2})^T b \end{aligned}$$

and so $0 > (m_{i2} - m_{i1})^T b$. Next note that

$$\begin{aligned} &[0] = M\tilde{A} \\ \implies &0 = m_i^T \tilde{A} \\ &= m_i^T [A^T, -A^T, I]^T \\ &= (m_{i2} - m_{i1})^T A = m_{i3}^T I \geq 0^T \in \mathbb{R}^n \end{aligned}$$

and hence with $y = m_{i2} - m_{i1}$ we have $A^T y \geq 0$. □

Example 1.7. Consider

$$\begin{aligned} 3x_1 + 2x_2 &= 3 \ (y_1) \\ 2x_1 - x_2 &= 3 \ (y_2) \\ x_1, x_2 &\geq 0 \end{aligned}$$

then

$$\underbrace{(3y_1 + 2y_2)}_{=0}\underbrace{x_1}_{\geq 0} + \underbrace{(2y_1 - y_2)}_{=0}\underbrace{x_2}_{\geq 0} = 3y_1 - 2y_2$$

Remark 1.4. Note that $Ax \geq b$ has the equivalent form $\tilde{A}\tilde{x} = b, \tilde{x} \geq 0$ where $\tilde{A} = [A, -A, -I], \tilde{x} = [u^T, v^T, s^T]^T$ and the alternative formulation is

$$\tilde{y}^T \tilde{A} \geq 0^T, b^T \tilde{y} < 0 \implies \tilde{y}^T A = 0^T, \tilde{y} \leq 0, b^T \tilde{y} < 0$$

or if we set $y = \tilde{y}$ then $A^T y = 0, y \geq 0, b^T y > 0$.

[See lecture 9 notes for more details]

Proof. Here is a proof regarding valid inequalities on the system

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

Explicitly, $\pi^T x \geq \pi_0$ is valid for $P \iff \exists u \geq 0$ such that $A^T u = \pi$ and $b^T u \geq \pi_0$.

($\implies$) Given $\pi^T x \geq \pi_0$ for any $x \in P$, suppose that there is no $u \geq 0$ such that $A^T u = \pi, b^T u \geq \pi_0$. That is, the set

$$U = \{u \in \mathbb{R}^m : A^T u = \pi, b^T u \geq \pi_0, u \geq 0\} = \emptyset$$

By Farkas' Lemma, there exists $\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}, \gamma \in \mathbb{R}^m$ with $\beta, \gamma \geq 0$ such that

$$\begin{cases} \alpha^T A^T + \beta b^T + \gamma^T I = 0^T \\ \alpha^T \pi + \beta \pi_0 > 0 \end{cases}$$

Since $\gamma \geq 0$ then

$$-\alpha^T A^T - \beta b^T \geq 0 \implies -A\alpha \geq \beta b, -\pi^T \alpha < \beta \pi_0$$

<u>Case $\beta > 0$</u>

Let $\hat{x} = -\alpha/\beta$. Then $A\hat{x} \geq b \implies \hat{x} \in P$ and $\pi^T \hat{x} < \pi_0$ which contradicts the validity that $\pi^T x \geq \pi_0$.

<u>Case $\beta = 0$</u>

We then get $-A\alpha \geq 0$ and $-\pi^T \alpha < 0$. Pick $\hat{x} \in P$, since P is nonempty, and let $\pi^T \hat{x} = \pi_0 + \delta$ and $x(\lambda) = \hat{x} + \lambda(-\alpha)$ with $\delta, \lambda \geq 0$. Then

$$Ax(\lambda) = A\hat{x} + \lambda(-A\alpha) \geq b, \lambda \geq 0$$

and

$$\begin{aligned} \pi^T x(\lambda) &= \pi^T x(\lambda) + \lambda(-\pi^T \alpha) \\ &= \pi_0 + \delta + \lambda(-\epsilon) \end{aligned}$$

Hence, we may choose λ large enough so that $\pi^T x(\lambda) < \pi_0$ and we get the same contradiction as in the previous case.

($\impliedby$) Note that

$$\begin{aligned} Ax \geq b, \forall x \in P &\implies u^T A x \geq u^T b, \forall x \in P \\ &\implies \pi^T x \geq u^T b, \forall x \in P \\ &\implies \pi^T x \geq \pi_0, \forall x \in P \\ &\implies \pi^T x \geq \pi_0 \text{ is a valid inequality for } P \end{aligned}$$

□

Proposition 1.2. *(About Extreme Rays) Given x^* is an extreme point of $P, x^* \in P = \{x : Ax \geq b\}$, the following are equivalent:*

(A) $x^* = \frac{1}{2}x^1 + \frac{1}{2}x^2$ for $x^1, x^2 \in P \implies x^1 = x^2 = x^*$

(b) If $A^= x^* = b^=$ where $A^=, b^=$ that define the inequalities that are satisfied with equality $\implies$ rank$(A^=) = n$

Proof. $[(A) \implies (B)]$ Suppose x^* does not satisfy (B). That is,

$$\text{rank}(A^=) \leq n - 1 \implies \dim(\text{null}(A)) \geq 1 \implies \exists d \neq 0, A^= d = 0$$

Consider $x^1 = x^* + \lambda d, x^2 = x^* - \lambda d$. Note that $A^= x^1 = b^=$ and

$$A^> x^1 = \underbrace{A^> x^*}_{> b^>} + \lambda A^> d \geq b^>$$

with the proper choice of λ. Hence $Ax^1 \geq b$ and $x^1 \in P$. Similarly, from the above construction, $Ax^2 \geq b$ and $x^2 \in P$ which contradicts the statement $x^1 = x^2 = x^*$.

$[(B) \implies (A)]$ Suppose that $\exists x^1, x^2 \in P$ such that $x^* = \frac{1}{2}x^1 + \frac{1}{2}x^2$ and let $d = x^2 - x^1 \neq 0$. It can be seen that $d \in$ null$(A^=)$ which is a contradiction. □

Proof. Here is the proof that polyhedra will have lines if and only if they contain no extreme points.

($\implies$) P contains a line $\iff \exists x \in P$ and $d \neq 0$ such that $x + \lambda d \in P$ for all $\lambda \in \mathbb{R}$. Then,

$$A^{=}(x + \lambda d) \geq b^{=}, \lambda \in (-\infty, \infty) \implies A^{=}d = 0$$

($\impliedby$) P contains no lines $\implies$ we can crash into the boundary and keep moving until we hit an extreme point, picking up dimensions as we go along. □

Theorem 1.5. *The problem* $(LP) : \min_x\{c^T x : x \in P\}$ *is unbounded if and only if there exists an extreme ray* d *of* $P = \{x : Ax \geq b\}$ *such that* $c^T d < 0$*.*

Proof. ($\impliedby$) Start from $\bar{x} \in P \implies \bar{x} + \lambda d \in P, \forall \lambda \geq 0$. Hence $c^T(\bar{x} + \lambda d) = c^T\bar{x} + \lambda c^T d \to -\infty$ as $\lambda \to \infty$ by $c^T d < 0$.

($\implies$) Suppose that the LP is unbounded and select $\{x^i\} \in P$ such that $c^T x^i \leq -i$ for all $i \in \mathbb{N}$. We claim that $\exists d \in D = \{x : Ax \geq 0\}$ such that $c^T d = -1$. Suppose that the claim is not true. Then the following system is infeasible:

$$\begin{aligned} Ad &\geq 0 \\ c^T d &= -1 \end{aligned}$$

with alternative system

$$\begin{aligned} u^T A + vc^T &= 0^T \\ u^T 0 - v &> 0 \\ u &\geq 0 \end{aligned}$$

which tells us that $u^T A = -vc^T$. Let $\bar{u} = -\frac{1}{v}u \implies \bar{u}^T A = c^T$ and $\bar{u} \geq 0$. Then

$$\begin{aligned} u^T b &\leq \underbrace{\bar{u}^T}_{\geq 0}\underbrace{Ax^i}_{\geq b} = \underbrace{c^T x^i}_{\leq -i}, \forall i \\ &\implies \bar{u}^T b < -\infty \end{aligned}$$

which is a contradiction. So our claim is true. Now consider the polyhedron

$$D' = \{d : Ad \geq 0, c^T d = -1\} \neq \emptyset$$

which contains no lines since D does not have any lines. Therefore, from our previous theorem, it has an extreme point $\hat{d} \in D' \subseteq D$, which has n linearly independent constrains which are tight from D'. It then satisfies $(n-1)$ linearly independent constraints at equality from $D \implies \hat{d}$ is an extreme ray. □

Theorem 1.6. *If the problem* $(LP) : \min_x\{c^T x : x \in P\}$ *has an optimal solution then one of the extreme points of* P *must be an optimal solution.*

Proof. Suppose the LP has an optimal solution at $x^* \implies c^T d \geq 0, \forall d \in D = \{x : Ax \geq 0\}$. Suppose that x^* is not an extreme point. Then rank$(A^{=}) \leq n - 1$ and $\exists d \in$ null$(A^{=})$.

(i) if $d \in D$ then let $d' = -d$ where we have $c^T d' \leq 0$

(ii) if $d \notin D$ then let $d' = d$ or $-d$ such that $c^T d' \leq 0$

The dimensions will increase since every traversal adds another equality constraint. □

Proof. Here is the proof of the Representation Theorem.

$(Q \subseteq P)$ Pick $x \in Q$. Then $Ax = \sum \lambda_i Ax^i + \sum \mu_j Ad^j \geq b$.

$(P \subseteq Q)$ Pick $x^* \in P$ and suppose that $x^* \notin Q$. Then the alternative system, from Farkas' Lemma, is $\exists(u, v)$

$$\begin{aligned} u^T x^i + v &\geq 0, \forall i \\ u^T d^j &\geq 0, \forall j \\ u^T x^* + v &< 0 \end{aligned}$$

Then $u^T x^i \geq -v > u^T x^*$ and $u^T d^j \geq 0$ for all j. Consider the LP $u^T x^i + v \geq 0 u^T d^j \geq 0 u^T x^* + v < 0 \min_x \{u^T x : x \in P\}$ which is bounded and has x^* optimal and strictly less in objective value than all of the extreme points, which is impossible. □

2 Simplex Method

Consider the **standard form** (LP)

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

with $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, x \in \mathbb{R}^n$, and rank$(A) = m$. Define $P = \{x : Ax = b, x \geq 0\}$.

Definition 2.1. A vector $x \in \mathbb{R}^n$ is a **basic solution** to the system $Ax = b$ if there exists a non-singular $m \times m$ submatrix of A, call it A_B, such that $x_B = A_B^{-1} b, x_N = 0$ where $x = (x_B, x_N)^T$.

A **basis** of A is a $m \times m$ submatrix that is invertible (i.e. composed of m linearly independent columns of A) .

A **basic feasible solution** (bfs) is a basic solution such that $x \geq 0$ (i.e. $x_B = A_B^{-1} b \geq 0$).

A **degenerate bfs** is a bfs that has one or more of the basic variables equal to 0 $\iff$ more than n constraints are tight at the solution.

Theorem 2.1. *x^* is an extreme point of P $\iff$ x^* is a bfs.*

Proof. ($\Longrightarrow$) There has to be n linearly independent constraints that are binding at x^*. Suppose that k of the inequalities are binding. Then the system satisfies

$$\begin{pmatrix} A_{m\times(n-k)} & A_{m\times k} \\ 0_{k\times(n-k)} & I_{k\times k} \end{pmatrix} x^* = \begin{pmatrix} b \\ 0 \end{pmatrix}$$

where $A = A^=$ and $m + k = n$. □

Remark 2.1. Correspondence is not one to one. There may be multiple basis whose bfs correspond to the same extreme point.

Definition 2.2. Two **adjacent extreme points** are points on the polyhendron that share exactly $n-1$ active constraints. Two **bfs** are adjacent if their corresponding bases differ in exactly one column.

Remark 2.2. If there is no degeneracy, two adjacent bfs $\iff$ two adjacent extreme points.

Remark 2.3. Suppose that we are at a bfs $x = (x_B, x_N)^T$. Let B and N denote the index set of the basic and non-basic columns respectively.

Moving from x to an adjacent bfs corresponds to trying to increase one of the non-basic variables, say x_j, from 0. That is, we move in the direction $d^j = (d_B^j, d_N^j)^T$ where $d_N^j = e_j$. Then the new point after a step of $\lambda \geq 0$ along d^j will be

$$x + \lambda d^j = \begin{pmatrix} x_B + \lambda d_B^j \\ \lambda e_j \end{pmatrix}$$

Note that the new point must be feasible:

$$A(x + \lambda d^j) = b, \text{ small enough } \lambda \geq 0 \iff Ad^j = 0 \iff A_B d_B^j + A_N d_N^j = 0 \iff d_B^j = -A_B^{-1} A_N d_N^j$$

and since $d_N^j = e_j$ then $d_B^j = -A_B^{-1} A_j$. Our original objective value was $z^{old} = c^T x = c_B^T x_B + c_N^T x_N = c_B^T x_B$ and after moving in d^j for a step of $\lambda \geq 0$, we get

$$\begin{aligned} z^{new} &= c^T(x + \lambda d^j) = z^{old} + \underbrace{\lambda}_{\geq 0} \underbrace{c^T d^j}_{<0} \\ &= z^{old} + \theta c^T d^j \end{aligned}$$

where

$$c^T d^j = c_B^T d_B^j + c_N^T d_N^j = -c_B^T A_B^{-1} A_j + c_j = r_j = r^j$$

which we call the **reduced costs** of the j^{th} non-basic variable. A basic direction d^j is improving if $c^T d^j < 0 \iff r^j < 0$. We can define the vector of reduced costs as

$$r^T = c^T - c_B^T A_B^{-1} A$$

Suppose we have $j \in \mathbb{N}, r_j < 0$. Then any positive $\theta \geq 0$ will improve the objective. How large can θ be? As long as $y = x + \theta d^j$ is feasible. That we need

$$y = x + \theta d^j \implies \begin{cases} y = x + \theta d_i^j & i \in B \\ y = \theta d_i^j \geq 0 & i \in N \end{cases}$$

If $d_i^j \geq 0$ for all $i \in B \implies \theta^* + \infty$ and the problem is unbounded.

Proposition 2.1. *If x is bfs with basis A_B and $r_j < 0$ for some $j \in \mathbb{N}$ and $d^j \geq$ then LP is unbounded.*

Proof. Suppose $P = (x : Ax = b, x \geq 0\}$. The set of recession directions is $D = \{d : Ad = 0, d \geq 0\}$. For the given conditions $d^j \in D$ and $c^T d^j = r_j < 0 \implies$ the LP is unbounded. □

Remark 2.4. Suppose instead that $\exists i \in B, d_i^j < 0$ and so

$$x_i + \theta d_i^j \geq 0, \forall i \in B : d_i^j < 0 \implies \theta \leq -\frac{x_i}{d_i^j}, \forall i \in B : d_i^j < 0 \implies \theta^* = \min\left\{-\frac{x_i}{d_i^j} : i \in B, d_i^j < 0\right\}$$

which we call the **ratio test**. Let $l \in B$ be the index that achieves the above minimum. Given $y = x + \theta^* \cdot d^j$, we have

$$y_k = \begin{cases} 0 & k \in N \backslash \{j\} \\ \theta^* & k = j \\ x_k + \theta^* d_k^j & k \in B \backslash \{l\} \\ 0 & k = l \end{cases}$$

Proposition 2.2. *Let $\bar{B} = [B \backslash \{l\}] \cup \{j\}$, $\bar{N} = \{1, ..., n\} \backslash \bar{B}$. We claim that $A_{\bar{B}}$ is a basis of A and y is a bfs corresponding to $A_{\bar{B}}$.*

Proof. Define $A_B = [A_1, ..., A_{m-1}, A_l]$ and $A_{\bar{B}} = [A_1, ..., A_{m-1}, A_j]$. Suppose that the columns of $A_{\bar{B}}$ are not linearly independent. Then $\exists \lambda_1, ..., \lambda_m$ not all zero such that

$$\lambda_m A_j + \sum_{i=1}^{m-1} \lambda_i A_i = 0 \iff \lambda_m A_B^{-1} A_j + \sum_{i=1}^{m-1} \lambda_i A_B^{-1} A_i = 0$$
$$\iff -\lambda_m d_B^j + \sum_{i=1}^{m-1} \lambda_i e_i = 0$$

Consider the l^{th} equation (component) which says

$$-\lambda_m d_B^j = 0 \implies \lambda_m = 0 \implies \sum_{i=1}^{m-1} \lambda_i e_i = 0$$

which is impossible. Hence A_B is a basis. Now

$$Ay = A_{\bar{B}} y_{\bar{B}} + \underbrace{A_{\bar{N}} y_{\bar{N}}}_{=0} = A_{\bar{B}} y_{\bar{B}} = b \implies y \text{ is a bfs}$$

□

Algorithm 1. *(0) Find a bfs*

*(1) Check $r_j, \forall j \in N$ and if $r_j \geq 0, \forall n \in N$ then **stop**. Current bfs is optimal. Else pick $j \in N$ such that $r_j > 0$.*

(2) Compute $d^j = (d_B^j, d_N^j)$ where $d_B^j = -A_B^{-1} A_j, d_N^j = e_j$.

*(3) If $d_B^j \geq 0$ then **stop**. The problem is unbounded. Otherwise, compute*

$$\theta^* = \min\left\{-\frac{x_i}{d_i^j} : i \in B, d_i^j < 0\right\}$$

Let $l \in B$ be the minimized. Compute new solution (bfs)

$$y = x + \theta^* d^j$$

with new basis $\bar{B} = [B \backslash \{l\}] \cup \{j\}$. Go to step (1).

2.1 Degeneracy

A bfs is **degenerate** if $x_i = 0$ for some $i \in B$.

Theorem 2.2. *Assume:*

(i) we have a starting bfs

(ii) all bfs are non-degenerate

Then after a finite number of iterations, the simplex method will either find an optimal solution or detect the problem is unbounded.

Lemma 2.1. *If x is a bfs and $r \geq 0$*

(a) then x is optimal

(b) if x is an optimal bfs and not degenerate, then $r \geq 0$

Proof. Pick a solution $y \in P$ and let $d = y - x$. Note that

$$Ad = 0 \iff A_B d_B + A_N d_N = 0 \iff d_B = -A_B^{-1} A_N d_N$$

and

$$\begin{aligned} c^T(y-x) = c^T d &= c_B^T d_B + c_N^T d_N \\ &= c_B^T \left[-A_B^T A_N d_N\right] + c_N^T d_N \\ &= \underbrace{\left[c_N^T - c_B^T A_B^{-1} A_N\right]}_{r_N^T} d_N \end{aligned}$$

but since $r_N \geq 0, d_N \geq 0$ (since $x_N = 0$ and $y_N \geq 0$ by feasibility) then $c^T(y-x) \geq \implies c^T y \geq c^T x$. Since y was arbitrary then x is optimal. ☐

Remark 2.5. To get from one basis B to another $\bar{B}$, find Q such that it transforms the augmented form $[A_B^{-1}|u], u = d^j$ to $[A_{\bar{B}}^{-1}|e_l]$. That is $-Qd^j = QA_B^{-1}A_j = e_l$ and hence $QA_B^{-1} = A_{\bar{B}}^{-1}$.

2.2 Tableau Method

The tableau table has the form

$-z = 0$	c_1	$\cdots$	c_n
b_1			
$\vdots$		A	
b_m			

Given an initial basis B, this looks like:

$-z = -c_B^T A_B^{-1} b$	$r = c^T - c_B^T A_B^{-1} A$	$r_j : (r_j < 0)$	r
$A_B^{-1} b$	$A_B^{-1} A$	$A_B^{-1} A_j$	
		u_l	

If $r_j \geq 0$ for all j, the problem is optimal, if for j such that $r_j < 0$ we have $A_B^{-1} A_j \geq 0$, then the problem is unbounded. A basis change looks like:

$-z = -c_B^T A_B^{-1} b - r_j \theta^*$	$r = c^T - c_B^T Q A_B^{-1} A$	0	r
$Q A_B^{-1} b$	$Q A_B^{-1} A$	$Q A_B^{-1} A_j$	
		0	

where $-c_B^T A_B^{-1} b - r_j \theta^* = -(c_B^T A_B^{-1} b + r_j \theta^*)$.

2.3 Pivot Rules

For entering variables, you can:

(1) Pick the most negative reduced cost

(2) Pick a negative reduced cost with the smallest index (**Bland's Rule**)

(3) Find the smallest $r_j \theta^*$ (most negative)

(4) Steepest edge

For leaving variables, if there are multiple minimum ratios, you can:

(1) Pick the variable with smallest index (**Bland's Rule**)

Proposition 2.3. *Bland's rule ensures that there are no cycles.*

2.4 Initial BFS / Two-Phase Method

In the standard form problem, assume w.l.o.g that $b \geq 0$. We can first solve the Phase I problem

$$\begin{aligned} \min\ & e^T y \\ s.t.\ & Ax + Iy = b \\ & x, y \geq 0 \end{aligned}$$

where e is a vector of ones. We can immediately use y as the starting basis ($y = b$). This problem is bounded and feasible, so it must have an optimal solution. If the objective value is greater than 0, the original LP is infeasible.

Otherwise if the objective value is 0 and if there are no y variables in the optimal basis, we have a starting basis for the original LP. Otherwise, if the objective value is 0 and there are some y variables in the basis, the basis is degenerate and we will need to do some extra work.

An alternate formulation is the M method with the form

$$\begin{aligned} \min\ & c^T x + M \cdot e^T y \\ s.t.\ & Ax + Iy = b \\ & x, y \geq 0 \end{aligned}$$

where M is really **big**. If the objective value of this problem is unbounded, then the original LP is *either* unbounded or infeasible, but we have no way of telling which one it is.

2.5 Complexity

Empirical evidence suggests that the complexity of the Simplex algorithm is $O(m)$, where m is the number of constraints. The Klee-Minty cubes of order n have the vanilla Simplex algorithm taking 2^{n-1} iterations.

3 Duality

Remark that

$$\min_x \{f(x) : g(x) \leq 0)\} \leq \max_{\lambda \geq 0} \left[\min_x \{f(x) + \lambda g(x)\}\right]$$

where the right side is a relaxation of the left side. We will do something similarly for linear programming. Consider the standard primal problem

$$\begin{aligned} (P)\ \min\ & c^T x \\ \text{s.t. } & Ax = b \\ & x \geq 0 \end{aligned}$$

This has a relaxation

$$\phi(y) = \begin{bmatrix} \min\left[c^T x + y^T(b - Ax)\right] \\ \text{s.t. } x \geq 0 \end{bmatrix} = \begin{cases} y^T b & c^T - y^T A \geq 0 \\ -\infty & \text{o.w.} \end{cases}$$

Let v_P be the optimal value of (P) where it will be $-\infty$ if unbounded and $+\infty$ if it is infeasible. It is clear that

$$v_p \geq \phi(y), \forall y \in \mathbb{R}^m \implies v_p \geq \max_{y \in \mathbb{R}^m} \phi(y) = \begin{matrix} \max_y \phi(y) \\ y^T A \leq c^T \end{matrix} = \begin{matrix} (D)\ \max_y b^T y \\ y^T A \leq c^T \end{matrix}$$

The last expression is the **dual problem**, (D). So for every x feasible to (P) and every y feasible to (D), we have

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

which is called **weak duality**. We will also denote v_D as the optimal solution of (D).

Theorem 3.1. *(Weak Duality Theorem)* $v_P \geq v_D$. *In particular,*

$$\begin{aligned} \textit{primal unbounded} &\implies \textit{dual infeasible} \\ \textit{dual unbounded} &\implies \textit{primal infeasible} \end{aligned}$$

Remark 3.1. The dual problem of (D) is (P). To see this, a relaxation (upper bound) of (D) is

$$\begin{aligned} v_D \leq \max_y \{b^T y + x^T(c - A^T y) : x \geq 0\} &= \max_y \{b^T y + x^T(c - A^T y) : x \geq 0\} \\ &= c^T x + \max_y \{(b - x^T A^T) y^T : x \geq 0\} \\ &= \begin{cases} c^T x & Ax = b, x \geq 0 \\ -\infty & \text{o.w.} \end{cases} \\ &= \min\{c^T x : Ax = b, x \geq 0\} \end{aligned}$$

Here is a table (weakly) summarizing our observations:

Primal		Dual
min	$\geq$	max
# of (real) constraints	$\leftrightarrow$	# of variables
# of variables	$\leftrightarrow$	# of (real) constraints
obj. vector	$\leftrightarrow$	RHS vector
RHS vector	$\leftrightarrow$	obj. vector
≥ 0, free , ≤ 0 variables	$\leftrightarrow$	$\leq, =, \geq$ constraints
$\geq, =, \leq$ constraints	$\leftrightarrow$	≥ 0, free , ≤ 0 variables

Why Duality?

* Optimality certifying tool

* Algorithmic reasons

* Economic reasons

* Modeling tool

Theorem 3.2. *(Strong Duality) If (P) has an optimal solution then (D) has an optimal solution with the same objective value. That is, $v_P = v_D$.*

Proof. (Version 1) Let $P = \{x : Ax \geq b\}$. An inequality $\pi^T x \geq \pi_0$ is valid for $P \iff \exists u \in \mathbb{R}^m$ such that $u^T A = \pi^T, u^T b \geq \pi_0$. If (P) has an optimal solution then $c^T x \geq v_p$ is a valid inequality for

$$\{x : Ax \geq b, -Ax \geq -b, Ix \geq 0\}$$

So $c^T = (\alpha - \beta)^T A + \gamma^T I$ and $v_P \leq (\alpha - \beta)^T b + \gamma^T 0$ for some $\alpha, \beta, \gamma \geq 0$ and so $\hat{y}^T = (\alpha - \beta)^T$ is feasible to (D) with $y^T A \leq c^T$ and $v_P \leq b^T \hat{y} \leq v_D$. From weak duality, we know $v_D \leq v_P$ and hence $v_D = v_P$. ☐

Proof. (Version 2) If (P) has an optimal solution then there is an optimal bfs. Let A_B be the optimal basis $\iff$ (1) $A_B^{-1} b \geq 0$ and (2) $c^T - c_B^T A_B^{-1} A \geq 0 \implies c^T \geq c_B^T A_B^{-1} A$. Let $\hat{y}^T = c_B^T A_B^{-1}$. Then $\hat{y}$ is feasible to the dual. To see this, note that

$$c_B^T A_B^{-1} b = v_P \geq v_D \geq \hat{y}^T b = c_B^T A_B^{-1} b \implies v_P = v_D$$

and so we have strong duality. ☐

Theorem 3.3. *(Strong Duality v2) If one of (P) or (D) is feasible, then $v_P = v_D$.*

Here is a summary chart:

P\D	inf.	opt.	unb.
inf.	Y	N	Y
opt.	N	Y	N
unb.	Y	N	N

Example 3.1. (Multi-period bond cash flows) Consider the LP

$$\begin{aligned}
\min\ & z_0 + \sum_{i=1}^{n} p_i x_i \\
\text{s.t. } & (1 + r_t) z_{t-1} + \sum_{i=1}^{n} c_{it} x_i = L_t + z_t, \quad t = 1, 2, ..., T \\
& z_t \geq 0 \\
& x_i \geq 0
\end{aligned}$$

The dual is

$$\begin{aligned}
\max\ & \sum_{t=1}^{T} L_t^T y_t \\
\text{s.t. } & \sum_{i=1}^{T} c_{it} y_t \leq p_i & \forall i = 1, 2, ..., n (x_i) \\
& (1 + r_1) y_1 \leq 1 \\
& - y_t + (1 + r_{t+1}) y_{t+1} \leq 0 & \forall t = 1, ..., T - 1 (z_t) \\
& - y_T \leq 0 \\
& y_t \text{ unrestricted}
\end{aligned}$$

Example 3.2. (Minimum cost network flow problem) You are given a network, $G = (N, A)$ which are **N**odes and **A**rcs. Each arc has a cost and a capacity. We define $0 \le c_{ij}$ as the cost/unit flow on arc $(i, j) \in A$ and u_{ij} as the capacity on $(i, j) \in A$. Each node has a supply $b_i \in \mathbb{R}, \forall i \in N$ in a balanced network: $\sum_{i\in N} b_i = 0$. The primal problem is

$$\begin{aligned}(P)\min \sum_{(i,j)\in A} c_{ij}x_{ij} & \\ \text{s.t.} \sum_{(j,i)\in A} x_{ji} - \sum_{(i,j)\in A} x_{ij} = b_i \qquad & \forall i \in N(y_i) \\ 0 \le x_{ij} \le u_{ij} \qquad & \forall (i,j) \in A(w_{ij})\end{aligned}$$

The dual problem is

$$\begin{aligned}(D)\max \sum_{i\in N} y_i b_i + \sum_{(i,j)\in A} w_{ij}u_{ij} & \\ \text{s.t. } y_i - y_j + w_{ij} \le c_{ij} \qquad & \forall (i,j) \in A \\ y_i \text{ unrestricted} \qquad & \forall i \in N \\ w_{ij} \le 0 \qquad & \forall (i,j) \in A\end{aligned}$$

Theorem 3.4. *(Farkas' Lemma via Duality) Only one of the two systems is feasible:*

(I) $Ax = b, x \ge 0$

(II) $y^T A \le 0^T, y^T b > 0$

Proof. Consider the LP

$$\begin{aligned}(P)\ \min\ & 0^T x \\ \text{s.t. } & Ax = b \\ & x \ge 0\end{aligned}$$

which has the dual problem

$$\begin{aligned}(D)\ \max\ & y^T b \\ \text{s.t. } & y^T A \le 0\end{aligned}$$

If (I) is feasible, then $v_P = 0$ and from strong duality, $v_D = 0 \ge y^T b$ for all y such that $y^T A \le 0$. Thus (II) cannot be feasible. Similarly, if (II) is feasible then $v_D = y^T b > 0$ and from weak duality, then $v_P \ge v_D > 0$ which is only possible if (I) is infeasible. □

Theorem 3.5. *(Complementary Slackness Conditions) For the standard primal* $\min\{c^T x : Ax = b, x \ge 0\}$ *and dual* $\max\{b^T y : A^T y = c\}$, *if* (x, y) *are feasible solutions to* (P) *and* (D) *then* (x, y) *are optimal* $\iff$ $[b_i - (Ax)_i]\, y_i = 0, \forall i = 1, 2, ..., m$ *and* $\left[c_j - (y^T A)_j\right] x_j = 0, \forall j = 1, 2, ..., n.$

Proof. If (x, y) are optimal $\iff c^T x = y^T b \iff (c^T - y^T A)x = 0 \iff \sum_j \underbrace{[c_j - (Ax)_j]}_{\ge 0} \underbrace{x_j}_{\ge 0} = 0 \iff \left[c_j - (y^T A)_j\right] x_j = 0, \forall j = 1, 2, ..., n$. The former holds trivially from feasibility. □

Corollary 3.1. *If we have the canonical problems primal* $\min\{c^T x : Ax \ge b, x \ge 0\}$ *and dual* $\max\{b^T y : A^T y \le c, y \ge 0\}$ *then if* (x, y) *are feasible solutions to* (P) *and* (D) *then* (x, y) *are optimal* $\iff$ $[(Ax)_i - b_i]\, y_i = 0, \forall i = 1, 2, ..., m$ *and* $\left[c_j - (y^T A)_j\right] x_j = 0, \forall j = 1, 2, ..., n.$

Remark 3.2. (Optimality conditions for (P)): x is an optimal solution of $(P) \iff$

(1) Primal feasibility: $Ax \ge b, x \ge 0$

(2) Dual feasibility: $\exists y$ such that $y^T A \le c^T$, $y \ge 0$

(3) Complementary slackness

Remark 3.3. Given an optimal solution (x^*, y^*) to canonical $(P) \equiv \min\{c^T x : Ax \le b\}$ and $D \equiv \max\{y^T b : y^T A = c^T, y \ge 0\}$ with $a_i^T x^* = b_i, \forall i \in I$ and $a_i^T x^* > b_i, \forall i \notin I$, we have:

- $y_i^* \geq 0$, for all $i \in I$
- $y_i^* = 0$, for all $i \notin I$
- $c = \sum_{i \in I} y_i^* A_i \implies c$ lies in the cone generated by the active constraints (the $a_i' s$) on x^*

3.1 Dual Simplex

Remark 3.4. (Dual Simplex) Consider the standard LPs $(P) \equiv \min\{c^T x : Ax = b, x \geq 0\}$ and $D \equiv \{b^T y : y^T A \leq c^T\}$. A basis is **primal feasible** if $A_B^{-1} b \geq 0$ and is **dual feasible** if $c^T - c_B^T A_B^{-1} A \geq 0$. The **dual simplex** uses this in the following (high level) way:

- Start from a dual feasible basis
- Iterate to get a a primal feasible basis (while maintaining dual feasibility)

Remark 3.5. This is useful for integer programming branching since the parent node will produce a dual feasible basis which is not not affected (i.e. the dual feasible basis will always remain feasible) by tightening of the bounds in the primal problem.

Algorithm 2. *(Dual Simplex in Detail)*

0. Start from a dual feasible basis

1. Find l such that $[A_B^{-1} b]_l < 0$. If none exists, we are done and the current basis is optimal.

2. Check $v^T = [A_B^{-1} A]_l = [A_B^{-1}]_l A$. If $v^T \geq 0$ then the dual is unbounded and the primal is infeasible and STOP. (*)

3. Conduct the minimum ratio test of finding j such that

$$j \in \textit{argmin} \left\{ \frac{r_k}{|v_k|} : v_k < 0 \right\}$$

4. l is the pivot row and j is the pivot column. Add a multiple of the pivot row the all rows to all rows so that all elements of the pivot columns except the pivot element is reduced to 0, and the pivot element is 1.

5. Set $B \leftarrow (B \backslash \{l\}) \cup \{j\}$

Proof. [of (*)] Consider $d^T = -[A_B^{-1}]_l$ and recall that $[A_B^{-1} b]_l < 0 \implies d^T b > 0 \implies d \neq 0$. Since d^T is in the recession cone of the dual, $\{d : d^T A \leq 0\}$, with $-[A_B^{-1}]_l A = -v^T \leq 0$, then (D) is unbounded. ☐

Remark 3.6. We have $y_{old}^T = c_B^T A_B^{-1}$ and $y_{new}^T = y_{old}^T + \theta[-A_B^{-1}]$ with necessary condition

$$\begin{aligned} r_{new}^T = c^T - y_{new}^T A &= c^T - (y_{old}^T A) + \theta [A_B^{-1}]_l A \\ &= r_{old}^T + \theta v^T \geq 0 \end{aligned}$$

and $\theta \leq r_k / |v_k| \implies \theta = \min \left\{ \frac{r_k}{|v_k|} : v_k < 0 \right\}$.

3.2 Applications of Duality

Consider the standard LPs $(P) \equiv \min\{v(b) = c^T x : Ax = b, x \geq 0\}$ and $D \equiv \{b^T y : y^T A \leq c^T\}$.

- We have, under the Farkas' Lemma:

Outcome	Certificate
(P) is infeasible	$y : y^T A \leq 0$ and $y^T b > 0$
(P) is unbounded	$x : Ax = 0, x \geq 0$ and $c^T x < 0$
(P) has an optimal solution	$(x, y) :$ $Ax = b, x \geq 0$; $y^T A \leq c^T$; $c^T x = y^T b \iff x_j [c_j - (y^T A)_j] = 0, j = 1, 2, ..., n$

- Suppose we are given a $\bar{b}$ and have an optimal bfs with basis A_B. The corresponding dual solution is $y^T = c_B^T A_B^{-1}$. Assume that the optimal bfs is non-degenerate with $x_B = A_B^{-1}\bar{b} > 0$. Suppose the rhs $\bar{b}$ is changed by a small amount to $\bar{b} + \Delta b$. If Δb is small enough, then A_B remains primal and dual feasible. Then

$$v(b + \Delta b) = y^T(b + \Delta b) = v(b) + y^T \Delta b$$

  So semantically, $\left.\frac{\partial v(b)}{\partial b_i}\right|_{b=\bar{b}} = y_i$.

- (Core of a co-operative game) Consider firms K where each can produce products J from resources I.
  - Define:
    * x_j as the number of units of product j produced
    * a_{ij} as the units of resource i per unit of product j
    * r_j as the revenue / unit of product j
    * b_{ik} as the units of resource i available to firm k
  - A coalition of firms $S \subseteq K$ can pool their resources. The value of this coalition is

$$\begin{aligned} v(S) = \max \ & \sum_j r_j x_j \\ \text{s.t.} \ & \sum_j a_{ij} x_j \le \sum_{k \in S} b_{ik}, \quad \forall i \in I \\ & x_j \ge 0 \qquad\qquad\qquad \forall j \in I \end{aligned}$$

    A grand coalition has value $v(K)$. How can we allocate $v(K)$ to the firms in a "fair" way? A **core** is an allocation $\{z_k\}_{k \in K}$ such that it is
    (a) $\sum_{k \in K} z_k = v(K)$
    (b) $\sum_{k \in S} z_k \ge v(S), \forall S \subseteq K$ [Rationality]
  - Claim: Let y^* be an optimal dual solution of the LP defining $v(K)$. Then $z_k = (y^*)^T b_k = \sum_{i \in I} y_i^* b_{ik}$ for all $k \in K$ forms a core.
    * Proof:
      (a) $\sum_{k \in K} y^* b_k = v(K)$ by strong duality.
      (b) This follows from dual feasibility:

$$\begin{aligned} v(S) = \quad & \max r^T x & = \quad & \min y^T \left(\sum_{k \in S} b_k \right) \quad \le \quad y^* \left(\sum_{k \in S} b_k \right) = \sum_{k \in S} z_k \\ & \text{s.t. } Ax \le \sum_{k \in S} b_k & & \text{s.t } A^T y \ge r \\ & x \ge 0 & & y \ge 0 \end{aligned}$$

[Other duality applications have been posted in the professor's notes; one of the exercises will be on the FINAL EXAM!]

Example 3.3. Consider the program

$$\begin{aligned} \min_x \ & p^T x \\ \text{s.t.} \ & \begin{pmatrix} \min_r & r^T x \\ \text{s.t.} & Ar \le b \end{pmatrix} \ge R \\ & e^T x = 1 \\ & x \ge 0 \end{aligned}$$

The dual of the inner LP is

$$\begin{pmatrix} \max & y^T b \\ \text{s.t} & y^T A = x^T \\ & y \le 0 \end{pmatrix} \ge R \iff \begin{array}{rc} \exists y: & y^T A = x^T \\ & y \le 0 \\ & y^T b \ge B \end{array}$$

where you *may* want to prove this to yourself formally. This gives the equivalent formulation to the first LP:

$$\begin{aligned}\min_{x,y}\ & p^T x \\ \text{s.t. } & y^T A = x^T \\ & y \leq 0 \\ & y^T b \geq R \\ & e^T x = 1 \\ & x \geq 0\end{aligned}$$

[Other notes on sensitivity analysis in lecture notes]

Definition 3.1. Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be a convex function. A vector $S \in \mathbb{R}^m$ is a subgradient of f at x^0 if

$$f(x) \geq f(x^0) + s^T(x - x^0)$$

4 Large Scale Optimization

4.1 Bender's Decomposition

Example 4.1. Suppose we have n assets, $\tilde{r}_j$ as the random return on asset j for $j = 1, 2, ..., n$, x_j as the investment in asset j, and B as the budget. The randomized optimization program is

$$\begin{aligned}\max\ & \tilde{R} = \sum_{j=1}^{n} \tilde{r}_j x_j \\ \text{s.t. } & \sum_{j=1}^{n} x_j = B \\ & x_j \geq 0\end{aligned}$$

The max expected return (with utility function U) program is

$$\begin{aligned}\max\ & E[U(\tilde{r}^T x)] \\ \text{s.t. } & e^T x \\ & x \geq 0\end{aligned}$$

where e^T is a vector of all ones. In the case where $\tilde{r}$ is a a discrete distribution with $\{(x, f(x))\} = \{(r_k, p_k)\}_{k=1}^K$ and $U(s) = \min\{s, T\}$, we have a program of

$$\begin{aligned}\max\ & \sum_{k=1}^{K} p_k \\ \text{s.t. } & e^T x = b \\ & x \geq 0\end{aligned}$$

This has an LP formulation of

$$\begin{aligned}\max\ & \sum_{k=1}^{K} p_k z_k \\ \text{s.t. } & z_k \leq r_k^T x,\ \forall k = 1, 2, ..., K \\ & z_k \leq T,\ \forall k = 1, 2, ..., K \\ & e^T x = B \\ & x \geq 0\end{aligned}$$

Example 4.2. Suppose that we have warehouses $i \in I$ with supply x_i and we observe some demand $\tilde{d}_j$. We want to move the supplies at a minimal cost y_{ij} to meet demand. The LP to do this is

$$\begin{aligned} Q(\tilde{d}, x) = \min_y \ & \sum_i \sum_j c_{ij} y_{ij} \\ \text{s.t.} \ & \sum_i y_{ij} \geq \tilde{d}_i, \forall j \in I \\ & \sum_j y_{ij} \leq x_i, \forall i \in I \\ & y_{ij} \geq 0, \forall i, j \in I \end{aligned}$$

To minimize the expected costs as function of x, we construct the following LP:

$$\begin{aligned} \min \ & \sum_i p_i x_i + E_{\tilde{d}}[Q(x, \tilde{d})] \\ \text{s.t.} \ & x \geq 0, \forall i \in I \end{aligned}$$

[See lecture 21 for info on further decompositions]

Proof. (D-W Bounding) We need to show that $z + \sum_i (z_i - \beta_i) \leq z^*$. If $z_i = -\infty$ for any subproblem then the inequality holds trivially. Assume that $z_i > -\infty$ for all $i = 1, 2, ..., m$. The dual of the D-W reformulation is

$$\begin{aligned} \max \ & b^T \alpha + \sum_i \beta_i \\ \text{s.t.} \ & \alpha^T (D_i u^k) + \beta_i \leq (c^i)^T u^k, \forall k \in K_i, \forall i \\ & \alpha^T (D_i v^l) \leq (c^i)^T v^l, \forall l \in L_i, \forall i \end{aligned}$$

We can construct (prove this) a solution $(\alpha, z_1, ..., z_m)$ that is dual feasible. The result will follow. □

5 Network Flows

Definition 5.1. An **undirected graph** $G = (N, A)$ is a collection of nodes N and arcs/edges A. In contrast, a **directed graph** is an undirected graph where the arcs are ordered pairs of nodes.

Definition 5.2. A **network flow problem** is a problem in a directed graph with flow costs c_{ij} and capacities u_{ij} at each arc and supplies b_i at each node.

Definition 5.3. Define $O(i) = \{j : (i, j) \in A\}$, $I(i) = \{j : (j, i) \in A\}$. The **minimum cost problem** is

$$\begin{aligned} \min_f \ & \sum_{(i,j) \in A} c_{ij} f_{ij} \\ \text{s.t.} \ & \sum_{j \in O(i)} f_{ij} - \sum_{j \in I(i)} f_{ji} = b_i \qquad , \forall i \in N \\ & 0 \leq f_{ij} \leq u_{ij} \qquad , \forall (i, j) \in A \end{aligned}$$

We assume that:

- The underlying graph is **connected**.
- We have a balanced system: $\sum_{i \in N} b_i = 0$.

Definition 5.4. The **node-arc incidence matrix** is an $|N| \times |A|$ matrix A such that

$$a_{ik} = \begin{cases} 1 & \text{if edge } k \text{ leaves } i \\ -1 & \text{if edge } k \text{ enters } i \\ 0 & \text{otherwise} \end{cases}$$

The minimum cost problem can be re-posed as

$$\begin{aligned} \min_{f \in \mathbb{R}^{|A|}} \ & c^T f \\ \text{s.t. } & Af = b \\ & 0 \le f \le u \end{aligned}$$

Since A is not full rank, then we must change $A \mapsto \tilde{A}, b \mapsto \tilde{b}$ by dropping one arbitrary row to make the constraints linearly independent.

Definition 5.5. An $m \times n$ matrix A is **total unimodular** (TU) if every square submatrix of A has determinant $-1, 0$, or $+1$.

Theorem 5.1. *If A is TU then the polyhedron*

$$X = \{x \in \mathbb{R}^n : Ax = b, x \ge 0\}$$

with $b \in \mathbb{Z}^m$ (if non-empty) has integer extreme points.

Proof. Suppose that X is an extreme point of X which is a bfs with basis B. Then

$$A_B x_B = b \iff (x_B)_j = \frac{\underbrace{\det(A_B^j)}_{\in \mathbb{Z}}}{\underbrace{\det(A_B)}_{\in \{1,-1\}}} \in \mathbb{Z}$$

□

Theorem 5.2. *A node-arc-incidence matrix is TU.*

Proof. Suppose A is not TU and pick the smallest submatrix B such that $\det(B) \notin$ {-1,0,1}. Each column of B has at most two non-zero entries.

Each column of B has at most two nonzero entries.

If there are no nonzero entries, $\det(B) = 0$.

If there is one nonzero entry, B is not the smallest such submatrix.

So every column has two nonzero entries and hence $\det(B) = 0$ as summing up the rows will yield a zero vector. □

Corollary 5.1. *A is TU $\implies$ $\tilde{A}$ is TU $\implies$ $\begin{bmatrix} \tilde{A} & 0 \\ I & I \end{bmatrix}$ is TU*

Corollary 5.2. *The standard from of the minimum cost problem:*

$$\begin{aligned} \min \ & c^T f \\ \textit{s.t. } & \tilde{A} f = b \\ & If + Is = u \\ & f, s \ge 0 \end{aligned}$$

is a TU system and hence has integer optimal basic feasible solutions. Also bfs $c_B^T A_B^{-1}$ to the dual are also integral.

Definition 5.6. The **shortest path problem** is posed where you are in a directed graph $G = (N, A)$, a start node $s \in N$, an end node $t \in N$, lengths $c_{ij} \ge 0$ of each arch $(i, j) \in A$. The goal is to find the minimum length path from $s \to t$. This can be written as

$$\begin{aligned} \min \sum_{(i,j)\in A} & c_{ij}x_{ij} \\ \text{s.t.} \sum_{j\in O(s)} & x_{sj} = 1 \\ \sum_{j\in I(t)} & x_{jt} = 1 \\ \sum_{j\in I(i)} & x_{ji} = \sum_{j\in O(j)} x_{ij}, \forall i \in N\backslash\{s,t\} \end{aligned}$$

You can show that this system is TU. You can solve this efficiently with Dijkstra's algorithm and use this set-up to find solutions to alternative (more complicated) formulations.

Definition 5.7. The **assignment problem** is a matching problem between two sets of nodes which forms a bipartite graph. There will be n initial nodes in one group, m initial nodes in the other group (with an additional $n-m$ dummy nodes with zero flow if necessary).

Definition 5.8. The **max flow problem** is a problem where you wish to find the maximum amount that we can push from a node s to another node t. This can be posed a minimum cost problem (circulation problem).

6 Ellipsoid Method

Definition 6.1. An **LP standard form instance** is encoded as a triple (c, A, b) where all entries are upper bounded by a large number U. The size of the problem is roughly $n\log_2 U + nm\log_2 U + m\log_2 U \sim O(mn\log_2 U)$ in binary.

Definition 6.2. A **running time** $T_I(n)$ of an algorithm A on an instance I of a problem of size n is **polynomial time** if there exists k, independent of n, such that $T(n) = O(n^k)$. The **running time of a family of instances** P is $T(n) = \sup_{I\in P} T_I(n)$.

Remark 6.1. The Simplex algorithm has exponential ($\sim O(n^m)$) run-time in the worst case for the standard Dantzig pivoting rule.

Definition 6.3. The **linear feasibility problem** is to decide whether or not a set $X = \{x \in \mathbb{R}^n : Ax \geq b\}$ is empty or there exists an element $\hat{x} \in X$.

Here is the general algorithm. Assume that $A \in \mathbb{Z}^{m\times n}, b \in \mathbb{Z}^m$ and the size is $O(mn\log_2 U)$. If $X \neq \emptyset$ then let $\underline{v}, \overline{v} > 0$ such that $\underline{v} \leq vol(X) \leq \overline{v}$. The idea is to generate a set of regions $E_0, E_1, ..., E_T$ and try to determine if the center of one of the regions is in X (i.e. $X \neq \emptyset$). If for some $0 \leq k \leq T$ we have $vol(E_k) < \underline{v}$ then $X = \emptyset$. We have to ensure that k is a "reasonable" number of iterations. For the 1D case with intervals, halving at each iteration, we have $T \geq \log \overline{v} - \log \underline{v}$.

Definition 6.4. Given $u \in \mathbb{R}^n$, an **affine transformation** $T_{L,z}(u)$ is determined by a square invertible matrix $L \in \mathbb{R}^{n\times m}$ and a vector $z \in \mathbb{R}^n$ where $T_{L,z}(u) = Lu + z$.

Definition 6.5. An **ellipsoid** is $T_{L,z}(S_n)$ where $S_n = \{x \in \mathbb{R}^n : x^T x \leq 1\}$. Explicitly, x is an element of the ellipsoid if

$$x = L\cdot u + z \iff u = L^{-1}(x-z) \iff (x-z)^T(L^{-1})^T(L^{-1})(x-z) \leq 1 \iff (x-z)^T D^{-1}(x-z) \leq 1$$

where $D = LL^T$ is a positive definite matrix. So alternatively, an ellipsoid can be defined via a center z and positive definite matrix D via

$$E(z, D) = \{x \in \mathbb{R}^n : (x-z)^T D^{-1}(x-z) \leq 1\}$$

Consider $X = \{x \in \mathbb{R}^n : Ax \geq b\}$ and assume that:

(A1) If $X \neq \emptyset$ then $\exists$ ellipsoids $\underline{E}, \overline{E}$ such that $\underline{E} \subseteq X \subseteq \overline{E}$ and $\underline{v} = vol(\underline{E}) > 0, \overline{v} = vol(\overline{E}) > 0$

(A2) Given $\hat{x} \notin X$ we can identify (separate), in polynomial time, an inequality such that $\pi^T\hat{x} < \pi_0$ and $\pi^T x \geq \pi_0$ for all $x \in X$ [trivially true]

(A3) Given an ellipsoid $E(z, D)$ and a halfspace $H = \{x : \pi^T x \geq \pi_0^T z\}$, we can find another ellipsoid E' such that $E' \supseteq E \cap H$ and

$$\frac{Vol(E')}{Vol(E)} \leq e^{-\frac{1}{2(n+1)}}$$

which we call the **ellipsoid property**.

Note that we stop when

$$\frac{Vol(E^T)}{Vol(\overline{E})} \leq e^{-\frac{T}{2(n+1)}} \leq \frac{\underline{v}}{\overline{v}} \implies T \geq \lceil 2(n+1)\left[\log \overline{v} - \log \underline{v}\right] \rceil$$

Theorem 6.1. *Given ellipsoid $E(z, D)$ and $H = \{x \in \mathbb{R}^n : \pi^T x \geq \pi^T z\}$, let*

$$D' = \frac{n^2}{n^2 - 1}\left(D - \frac{2}{n+1} \cdot \frac{D\pi\pi^T D}{\pi^T D\pi}\right)$$
$$z' = z + \frac{1}{n+1} \cdot \frac{D\pi}{\sqrt{\pi^T D\pi}}$$

then

$$\frac{Vol[E(z, D)]}{Vol[E(z', D')]} \leq e^{-\frac{1}{2(n+1)}}$$

and $E(z', D') \supseteq E(z, D) \cap H$.

Lemma 6.1. *Every extreme point of $Ax \geq b$ satisfies $-(nU)^n \leq x_j \leq (nU)^n$.*

Proof. If x is an extreme point, then x is a solution of $\tilde{A}x = \tilde{b}$. By Cramer's rule,

$$|x_j| = \left|\frac{\det(\tilde{A}^j)}{\det(\tilde{A})}\right| \leq \left|\det(\tilde{A}^j)\right|, \forall j = 1, 2, ..., n$$

Since $\det(A) = \sum_{\sigma \in S_n} (-1)^{|\sigma|} \prod_{i=1}^n a_{i\sigma(i)} \leq (nU)^n$ then $|x_j| \leq (nU)^n$ and $\log(|x_j|) = n(\log n + \log U)$. □

Note 1. We will convert our feasible set X to $X' = \{x \in \mathbb{R}^n : Ax \geq b, -K \leq x \leq K\}$ to keep our set bounded using bounds above.

Lemma 6.2. *Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}$ and $P_\varepsilon = \{x \in \mathbb{R}^n : Ax \geq b - \varepsilon e\}$ where*

$$\varepsilon = \frac{1}{2(n+1)}\left[(n+1)U\right]^{-(n+1)}$$

and e is a vector of all ones.

(a) $P = \emptyset \implies P_\varepsilon = \emptyset$

(b) $P \neq \emptyset \implies P_\varepsilon \neq \emptyset$ and full dimensional (non-zero volume)

Remark 6.2. We start with $P \to P_\varepsilon \to P'_\varepsilon$ and the starting ellipsoid will be centered at the origin with radius $(nU)^n$.

Theorem 6.2. *Given ellipsoid $E(z, D)$ and $H = \{x \in \mathbb{R}^n : \pi^T x \geq \pi^T z\}$, let*

$$D' = \frac{n^2}{n^2 - 1}\left(D - \frac{2}{n+1} \cdot \frac{D\pi\pi^T D}{\pi^T D\pi}\right)$$
$$z' = z + \frac{1}{n+1} \cdot \frac{D\pi}{\sqrt{\pi^T D\pi}}$$

We can approximate D', z' within enough precision to still get

(a) $P = \emptyset \implies P_\varepsilon = \emptyset$

(b) $P \neq \emptyset \implies P_\varepsilon \neq \emptyset$ and full dimensional (non-zero volume)

from the previous lemma.

Remark 6.3. (An optimization problem is a feasibility problem) The optimization problem $\min\{c^T x : Ax \geq b\}$ is equivalent to finding a feasible solution to $\{(x, y) : Ax \geq b, y^T A = c^T y \geq 0, c^T x = y^T b\}$.

Remark 6.4. Given $x_0 \in X$, define $X^{t+1} = X^t \cap \{x : c^T x \leq c^T x_t - \epsilon\}$. X^t converges to a "small" region which contains the optimal solution.

7 Interior Point Methods

Proof. (Of duality gap) Let $\gamma = \frac{\sqrt{\beta}-\beta}{\sqrt{\beta}+\sqrt{n}}$ and $\alpha = 1 - \gamma$. For all k, we have

$$\begin{aligned}
&\sum_{j=1}^{n}\left(\frac{x_j^k s_j^k}{\mu_k} - 1\right)^2 \leq \beta^2 \\
\iff & -\beta \leq \frac{x_j^k s_j^k}{\mu_k} - 1 \leq \beta, \forall j \\
\iff & (1-\beta)\mu_k \leq x_j^k s_j^k \leq (1+\beta)\mu_k, \forall j \\
\iff & \mu_k n(1-\beta) \leq (s^k)^T x^k \leq \mu_k n(1+\beta)
\end{aligned}$$

Now since $\mu_0 n(1-\beta) \leq \epsilon_0 \iff \mu_0 \leq \epsilon_0/(n(1-\beta))$ and $\mu_k = (1-\gamma)^k \mu_0 \leq \mu_0 e^{-k\gamma}$ then

$$k^* \geq \frac{1}{\gamma} \ln \frac{\epsilon_0(1+\beta)}{\epsilon(1-\beta)} \implies \mu_k \leq \mu_0 \frac{\epsilon(1-\beta)}{\epsilon_0(1+\beta)}$$

and hence $(s^{k*})^T(x^{k*}) \leq \mu_k n(1+\beta) = \epsilon$ □