CO255: Introduction to Optimization (Advanced Level)

Felix Zhou

Fall 2019, University of Waterloo
from Professor Levent Tuncel's Lectures

Contents

1 Introduction

1.1 Optimization

Definition 1.1.1 (Optimization Problem)
$f : S \to \mathbb{R}$, called the **objective function**, with S the **feasible region**, and x our **variable**.
We wish to maximize or minimize $f(x), x \in S$.

note that $\{\max f(x) : x \in S\} = \{\min -f(x) : x \in S\}$, so it suffices to minimize a function.

1.1.1 Outcomes

- $S = \emptyset$, the problem is **infeasible**, otherwise we say the problem is **feasible**
- $S \neq \emptyset, \exists x' \in S$ such that $f(x') \leq f(x), \forall x \in S$ so x' is an **optimal solution** and $f(x')$ is an **optimal value**
- $S \neq \emptyset$ but the problem is unbounded, so there are reasible solutions of arbitaryily small objective values
- $S \neq \emptyset$ and the problem is bounded, but there are no optimal solutions

1.2 Classes of Optimization Problems

1.2.1 Linear Programs

Definition 1.2.1 (Linear Programs)
$f(x) = c^T x = \sum c_i x_i + z_0$
$S = \{x \in \mathbb{R}^n : Ax \leq b \iff a_i^T x \leq b_i, A \in M(\mathbb{R})_{m\times n}, b \in \mathbb{R}^m\}$
We minimize $c^T x$ such that $Ax \leq b, x \in \mathbb{R}^n$
A **feasible solution** is a point $x' \in S$

Definition 1.2.2 (Integer (Linear) Programs)
Minimize $c^T x$ where $Ax \leq b, x \in \mathbb{Z}^n$
We can also only require *some* x_i to be integers.

1.2.2 Convex Programs

Definition 1.2.3 (Convex Set)
$S \subseteq \mathbb{R}^n$ is convex if

$$\forall x, y \in S, \forall \lambda \in [0,1], \lambda x + (1-\lambda)y \in S$$

Definition 1.2.4 (Convex Function)
$f : S \to \mathbb{R}$ is convex if

$$\forall x, y \in S, \forall \lambda \in [0,1], f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y)$$

Definition 1.2.5 (Concave Function)
f as above is concave if $-f$ is convex

Definition 1.2.6 (Convex Programs)
S is a convex set, $f : S \to \mathbb{R}$ is a convex function.

1.3 Examples

1.3.1 Transportaion Problem

Example 1.3.1
A company has a set F of distribution centers, set C of clients, each $i \in F$ can supply at most u_i units and each $j \in C$ demands d_j units. Shipping items cost $i \to j$ costs c_{ij}/unit. Find the minimum cost way of satisfying client demands.

Let us use variables x_{ij} to represent the number of units went from $i \to j$. We want to find the minimum $\sum_{i \in F} \sum_{j \in C} c_{ij} x_{ij}$ where

$$\forall i \in F, \forall j \in C, \sum_{j \in C} x_{ij} \leq u_i, \sum_{i \in F} x_{ij} = d_j, x_{ij} \geq 0$$

1.3.2 2-Player Game

Example 1.3.2
$A \in \mathbb{R}^{m\times n} = (a_{ij})$. Players R, C choose a row $i \in \{1, \ldots, m\}$, a column $j \in \{1, \ldots, n\}$ respectively. R pays C a_{ij}, what is C's best randomized strategy?

(C-LP) Maximize $\min_{i\in\{1,\ldots,m\}} \sum_{j=1}^{n} a_{ij}p_j$ where $\sum_{j=1}^{n} p_j = 1, p_j \geq 0$. So we find

$$\max v, v \leq \sum_{j=1}^{n} a_{ij}p_j, \forall i = 1, \ldots, m, \sum_{j=1}^{n} p_j = 1, p_j \geq 0$$

Note that v is simply another variable introduced as a method to formulate the problem of maximizing a minimum.

(R-LP) R's best strategy is similar. Minimize $\max_{i\in\{1,\ldots,n\}} \sum_{i=1}^{m} a_{ij}q_i$ where $\sum_{i=1}^{m} q_i = 1, q_i \geq 0$. So we find

$$\min w, w \geq \sum_{i=1}^{m} a_{ij}q_i, \forall j = 1, \ldots, n, \sum_{i=1}^{m} q_i = 1, q_i \geq 0$$

Note that both programs are feasible, not unbounded, and have optimal solutions.

Suppose (v, p) is a feasible solution to (C-LP), (w, q) is a feasible solution to (R-LP), we have

$$w \geq \max_j \sum_{i=1}^{m} a_{ij}q_i \geq \min_i \sum_{j=1}^{n} a_{ij}p_j \geq v$$

So optimal values are equal, which is a consequence of LP-duality.

1.3.3 General 2-Player Game

In general, we wish for $A, B \in \mathbb{R}^{m\times n}$, the payoff matrices for C, R respectively. Before, we had $B = -A$ (zero-sum game). Now, R, C play i, j respectively and receive a_{ij}, b_{ij} respectively.

Example 1.3.3
We wish to find an equilibrium such that each player has no incentive to deviate even if the other player's strategy is revealed. Given p, to not deviate from q, R wants $q^t Bp = \max_i (Bp)_i$. Given q, to not deviate from p, C wants $q^t Ap = \max_j (q^T A)_j$.

To find an equilibrium (p, q) means to find an equilibrium which maximizes total payoff:

$$\max p^T (A + B)q$$

1.3.4 Fair Assignment

Definition 1.3.1 (Fair Division)
We have n players $i = 1, \ldots, n$, m items $j = 1, \ldots, m$. Assigning a fraction x of $j \to i$ gives player i utility $u_{ij} \cdot x$.

Example 1.3.4 (Fair Assignment)
We wish to maximize

$$\prod_{i=1}^{n} u_{ij} \cdot x$$

Use variables x_{ij} to represent the fraction of j given to i. Determine

$$\max \prod_{i=1}^{n} \left(\sum_{j=1}^{m} u_{ij} x_{ij} \right)$$

where $\sum_{i=1}^{n} x_{ij} \leq 1, x \geq 0$

Note that the above is equivalent to determining

$$\max \sum_{i=1}^{n} \ln \left(\sum_{j=1}^{m} u_{ij} x_{ij} \right)$$

where $\sum_{i=1}^{n} x_{ij} \leq 1, x \geq 0$.

This would be a convex programing problem.

1.3.5 Job Assignment

Example 1.3.5
We have n worker, n jobs, and must match a different worker to each job. Assigning job j to worker i costs c_{ij}. We wish to find the minimum cost assignment.

Let

$$x_{ij} = \begin{cases} 1, & \text{worker } i \text{ assigned to job } j \\ 0, & \text{otherwise} \end{cases}$$

Determine

$$\min \sum_{i,j} c_{ij} x_{ij}$$

where $\sum_{j=1}^{n} x_{ij} = 1, i = 1, \ldots, n, \sum_{i=1}^{n} x_{ij} = 1, j = 1, \ldots, n, x_{ij} \in \mathbb{N}$

Note that there is an optimal solution $\overline{w}, x_{ij}^*$ integral to the LP obtained by dropping the integer constraint.

1.3.6 Fermat's Last Theorem

To demonstrate the power of non-linear programming, we model the infamous Fermat's Last Theorem.

Example 1.3.6
Consider the following optimization problem:

$$(\min x_1^{x_4} + x_2^{x_4} - x_3^{x_4})^2 + \sum_{i=1}^{4} \sin^2(\pi x_i)$$

The sine terms force the variables to be integers.

such that

$$x_1 \geq 1, x_2 \geq 1, x_3 \geq 1, x_4 \geq 3$$

Note that the infimum which is 0 is not attained if and only if Fermat's Last Theorem holds.

2 Linear Programming

2.1 Definitions

Definition 2.1.1 (Affine Function)
$f : \mathbb{R}^n \to \mathbb{R}$ if $f(x) = \alpha^T x + b$ for some $\alpha \in \mathbb{R}^n, \beta \in \mathbb{R}$.

Definition 2.1.2 (Linear Function)
An affine function with $\beta = 0$.

Definition 2.1.3 (Linear Constraint)
Inequalities in the form of $f(x) \leq g(x)$ or $f(x) = g(x)$ or $f(x) \geq g(x)$.

Definition 2.1.4 (Linear Program - LP)
An optimization problem such that

- the objective function being affine
- variables $x \in \mathbb{R}^n$
- the variables are subject to finite linear constraints

In this context,

- a feasible solution is $\bar{x} \in \mathbb{R}^n$ satisfying the linear contraints.
- an optimal solution is a feasible solution with maximal objective value
- the feasible region is the set of feasible solutions

Definition 2.1.5 (Standard Inequality Form)
$\max c^T x$ where

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

Definition 2.1.6 (Standard Equality Form)
$\max c^T x$ where

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

Proposition 2.1.1
The Standard Inequality Form and Standard Equality Form are general enough to express any linear program.

Proof
We can turn an LP in SIF to SEF by introducing slack variables.

We can turn any LP to SIF by

1. reversing $\geq$ constraints by multiplying both sides by -1.
2. splitting equality constraints by taking $\leq, - \geq -$.

2.2 Feasibility

2.2.1 Fourier-Motzkin Elimination

This is a procedure to determine the feasibility of $Ax \leq b$.

Example 2.2.1

$$\begin{aligned} 2x_1 + x_2 + x_3 &\leq 5 && (1) \\ -x_1 + 3x_2 + 2x_3 &\leq 6 && (2) \\ 2x_1 - x_2 &\leq 0 && (3) \\ x_1 - 2x_2 - x_3 &\leq -2 && (4) \end{aligned}$$

From (1), (2), and (4) we have that

$$\begin{aligned} x_3 &\leq \min\left\{5 - 2x_1 - x_2, \frac{6 + x_1 - 3x_2}{2}\right\} \\ x_3 &\geq x_1 - 2x_2 + 2 \end{aligned}$$

So we arrive at a new system of equations

$$\begin{aligned} x_1 - 2x_2 + 2 &\leq 5 - 2x_1 - x_2 && (5) \\ x_1 - 2x_2 + 2 &\leq \frac{6 + x_1 - 3x_2}{2} && (6) \\ 3x_1 - x_2 &\leq 0 && (7) \end{aligned}$$

Note how we completely eliminated x_3!

Theorem 2.2.2 (Fourier-Motzkin Elimination)
Let $Ax \leq b$ be the given system where $A \in \mathbb{R}^{m\times n}$.
Let

$$I_+ = \{i \in \{1,\dots,m\} : a_{in} > 0\}$$
$$I_i = \{i \in \{1,\dots,m\} : a_{in} < 0\}$$
$$I_0 = \{i \in \{1,\dots,m\} : a_{in} = 0\}$$

For all $k \in I_+, l \in I_-$, consider

$$x_n \leq \left(b_k - \sum_{j=1}^{n-1} a_{kj}x_j\right)\frac{1}{a_{kn}}$$
$$x_n \geq \left(b_l - \sum_{j=1}^{n-1} a_{lj}x_j\right)\frac{1}{a_{ln}}$$

We can then generate a new system $A'x \leq b'$ in terms of

(i) $a_i^T x \leq b, \forall i \in I_0$

(ii)

$$\left(b_l - \sum_{j=1}^{n-1} a_{lj}x_j\right)\frac{1}{a_{ln}} \leq \left(b_k - \sum_{j=1}^{n-1} a_{kj}x_j\right)\frac{1}{a_{kn}}$$

for all $k \in I_+, l \in I_-$.

By repeating this procedure to eliminate all variables, then we get the final system $A^0x \leq b^0$, where A^0 is a zero matrix. This is feasible if and only if $b^0 \geq 0$.

The idea behind the Elimination algorithm is to produce a lowest upper bound and a greatest lower bound for x_n, from which we can extract values of x_n given feasible $x_0,\dots,x_{n-1}$ to produce a feasible solution $x = (x_1,\dots,x_n)$.

However, to stay within the rules of LP, we generate all possible inequalities and know that one of the generated inequalities will be between the LB and UB. If LB $\leq$ UB, we are done, else the LP is infeasible.

Proposition 2.2.3

(1) If either I_+, I_- is the emptyset, there are no (k,l)-inequalities

(2) Every inequality of $A'x \leq b'$ is a nonnegative linear combination of inequalities from $Ax \leq b$

(3) $A'x \leq b'$ has $n-1$ variables but can have as many as $\frac{n_2}{4}$ constraints (not very efficient)

Lemma 2.2.4

$Ax \leq b$ is feasible $\iff$ $A'x \leq b'$ is feasible

Proof

$\implies$ This follows from (2) of the proposition above

$\impliedby$ When I_+ or I_- is $\emptyset$, we can choose x_n large or small enough to satisfy the inequalties. Elsewise, since $A'x \leq b'$ is feasible, we must have

$$\left[\max_{l \in I_-} \frac{b_l - \sum_{j=1}^{n-1} a_{lj}x_j}{a_{ln}}, \min_{k \in I_+} \frac{b_k - \sum_{j=1}^{n-1} a_{kj}x_j}{a_{kn}}\right] \neq \emptyset$$

We can then simply choose x_n in the interval above.

So when we say we have "eliminated" a variable, we actualy mean that we can easily find feasible values for it given feasible values of the reduced system of linear inequalities.

Note that if $Ax \leq b$ is feasible, Fourier-Motzkin is guaranteed to produce a feasible solution.

Next note that if $Ax \leq b$ is infeasible, Elimination gives nonnegative linear combinations of the system of inequality given by $u \in \mathbb{R}^m_+$ so

$$u^T = 0, u^T b = \gamma < 0$$

this is basically when we eliminate the very last variable.

Also note the similarities between Fourier-Motzkin and Gaussian Elimination. However, there is much less "flexibility" in the choices (pivots, linear constraints generated).

2.2.2 Farkas' Lemma

Theorem 2.2.5 (Fundamental Theorem of Linear Algebra)
Let $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ one and only one of the following holds

(I) $\exists x \in \mathbb{R}^n, Ax = b$

(II) $\exists y \in \mathbb{R}^m, A^T y = 0 \wedge b^T y = -1$

Proof
Suppose that (II) does not hold. This means that the rows of A are linearly independent.

By gaussian elimination, we will arrive at something of the form

$$\begin{pmatrix} 1 & 0 & \dots & 0 & \dots \\ 0 & 1 & \dots & 0 & \dots \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & \dots \end{pmatrix}$$

which enables us to get the x for the b.

Suppose (II) holds, and for a contradiction that (I) holds as well, we have

$$\begin{aligned} x^T A^T &= b^T \\ x^T (A^T y) &= b^T y \\ 0 &= -1 \end{aligned}$$

This is clearly a contradiction.

Theorem 2.2.6
Let $A \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^m$. Only one of the following has a solution

(1) x such that $Ax \le b$

(2) u such that $A^T u = 0, u \ge 0, b^T u < 0$

Proof ($\neg$(1) $\vee$ $\neg$(2))
If both were true

$$0 > b^T u \ge x^T \left(A^T u\right) = 0$$

which is a contradiction.

Proof (¬(1) $\implies$ (2))
Fourier-Motzkin Elimination on $Ax \leq b$ derives

$$0^T x \leq \gamma < 0, u \in \mathbb{R}^m_+, u^T A = 0^T, u^T b = \gamma < 0$$

Theorem 2.2.7 (Farkas' Lemma)
Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ one and only one of the following holds
(I) $\exists x \in \mathbb{R}^n, Ax \leq b, x \geq 0$
(II) $\exists y \in \mathbb{R}^m, A^T y \geq 0, y \geq 0, b^T y < 0$

Proof (¬(1) ∨ ¬(2))
Identical.

Proof (¬(1) $\implies$ (2))
Reduce the assumptions to the theorem above.

Consider

$$\begin{pmatrix} A \\ -I \end{pmatrix} x \leq \begin{pmatrix} b \\ 0 \end{pmatrix}$$

The dual is given by

$$(A^T, -I)y = 0, y \geq 0, (b^T, 0)y < 0$$

2.3 LP-Duality

2.3.1 Motivation

How can we prove bounds on the optimal value of a Linear Program? Nonegative combinations of the contraints produce upper bounds as long as the coefficients are greater than those in the function.

Example 2.3.1 (LP)
Consider the following LP: $\max z(x) := 2x_1 + x_2$ such that

$$\begin{pmatrix} 1 & -2 \\ 4 & 1 \\ -1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq \begin{pmatrix} 2 \\ 10 \\ 0 \\ 0 \end{pmatrix}$$

Note that

$$(0, 1, 2, 0)Ax \leq (0, 1, 2, 0)b$$

gives us

$$2x_1 + x_2 \leq 10$$

But any feasible solution of the original LP must satisfy the inequality above as it is a nonnegative linear combination of the original constraints and $x \geq 0$. So the problem is optimal if the objective value is 10!

one solution is

$$\bar{x}^T := (0, 10)$$

Definition 2.3.1 (Unbounded)
A LP is unbounded if for every $M \in \mathbb{R}$ there is a feasible x such that

$$z(x)$$

is a better objective value.

Remark that an unbounded LP is certainly feasible and no unfeasible LP is unbounded.

Definition 2.3.2 (Bounded)
$S \subseteq \mathbb{R}^n$ is bounded if there is some $M \in \mathbb{R}_{++}$ such that

$$S \subseteq [-M, M]^n$$

Definition 2.3.3 (dual)
Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ with the primal LP (P): $\max c^T x$ such that

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

The Dual denoted (D) of (P) is $\min b^T y$ such that

$$A^T y \geq c, y \geq 0$$

Note that

- (P) being a max-LP gives us (D) being a min-LP
- each primal constraint gives rise to a dual variable
- each primal variable gives rise to a dual constraint

2.3.2 Weak Duality

Theorem 2.3.2 (Weak Duality)
Let P be a max LP with objective function $c^T x$ and D the dual of P. D is a min LP with objective function $b^T y$.
Let $\overline{x}$ be a feasible solution to P and $\overline{y}$ a feasible solution to D. Then

$$c^T \overline{x} \leq b^T \overline{y}$$

Proof
We show this for P in SIF: max $c^T x$ with $Ax \leq b, x \geq 0$. Constructing the dual, we have min $b^T y$ with $c^T \leq y^T A, y \geq 0$.

Note that $Ax \leq b \implies y^T Ax \leq y^T b$ so

$$c^T \overline{x} \leq \overline{y}^T A \overline{x} \leq \overline{y}^T b = b^T \overline{y}$$

Corollary 2.3.2.1
If $\bar{x}, \bar{y}$ are feasible for (P), (D), respectively and

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

then $\bar{x}, \bar{y}$ are both optimal.

Proof
By Weak Duality, $c^T \bar{x}, b^T \bar{y}$ are lower and upper bounds on the objective values of (D), (P), respectively. But they each attain the lower/upper bounds so must be optimal.

Corollary 2.3.2.2
(P) is unbounded means (D) is infeasible and vice versa.

Proof
By application of Weak Duality on the contrapositive of the statement.

Proposition 2.3.3

$$\text{dual} \circ \text{dual}(P) = P$$

2.3.3 Strong Duality

Theorem 2.3.4 (Strong Duality)
If (P), (D) have feasible solutions then they have optimal solutions $\bar{x}, \bar{y}$ such that

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

Proof
Let P: max c^Tx such that $Ax \leq b, x \geq 0$, D: min b^Ty such that $A^Ty \geq c, y \geq 0$.

We wish to show that

$$\left.\begin{array}{ll} Ax \leq b, & x \geq 0 \\ -A^Ty \leq -c, & y \geq 0 \\ -c^Tx + b^Ty \leq 0 & \end{array}\right\} \text{is feasible}$$

Note the last inequality coupled with Weak Duality forces the feasible values have equivalence.

In matrix form we have

$$\underbrace{\begin{bmatrix} A & 0 \\ 0 & -A^T \\ -c^T & b^T \end{bmatrix}}_{A'} \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{x'} \leq \underbrace{\begin{bmatrix} b \\ -c \\ 0 \end{bmatrix}}_{b'}$$

Suppose for a contradiction that $A'x' \leq b'$ is infeasible. By Farkas' Lemma, there is some $u, v, \lambda \geq 0$ such that

$$\begin{pmatrix} A^T & 0 & -c \\ 0 & -A & b \end{pmatrix} \begin{pmatrix} u \\ v \\ \lambda \end{pmatrix} \geq 0$$

$$\begin{pmatrix} u \\ v \\ \lambda \end{pmatrix} \geq 0$$

$$\begin{pmatrix} b^T & -c^T & 0 \end{pmatrix} \begin{pmatrix} u \\ v \\ \lambda \end{pmatrix} < 0$$

In particular, notice we have

$$b^Tu < c^Tv$$

Case I: $\lambda > 0$. Then $\frac{u}{\lambda}, \frac{v}{\lambda} \geq 0$. From (1), $A^T(\frac{u}{\lambda}) \geq c$. we get $A(\frac{v}{\lambda}) \leq b$. So $\frac{u}{\lambda}, \frac{v}{\lambda}$ are feasible solutions for D, P respectively but

$$b^T\left(\frac{u}{\lambda}\right) < c^T\left(\frac{v}{\lambda}\right)$$

which contradicts Weak Duality

Case II: $\lambda = 0$. Since P, D are feasible, we can find feasible solutions $\overline{x}, \overline{y}$. Then, we have

$$\left.\begin{aligned} b^T u &\geq (A\overline{x})^T u = \overline{x}^T A^T u = 0 \\ c^T v &\leq (A^T\overline{y})^T v = \overline{y}^T A v = 0 \end{aligned}\right\} b^T u \geq c^T v$$

which contradicts our assumptions.

Either way, we reach a contradiction so we conclude the proof.

Lemma 2.3.5
if P is feasible and D is infeasible, then P is unbounded.

Proof (2 $\implies$ 1)
If P has an optimal solution, then P is feasible and not unbounded. By (the contrapositive to) the lemma, D is feasible.

Then by (2), D has an optimal solution and $\text{opt}(P) = \text{opt}(D)$.

Lemma 2.3.6
If (P) is feasible and (D) is unfeasible, then (P) is unbounded.

Proof
Suppose (P) is feasible, there is some $\bar{x} \in \mathbb{R}^n$ such that

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

Since (D) is unfeasible, we can apply the second case of Farkas' Lemma, there is some d such that

$$-Ad \geq 0, d \geq 0, -c^T d < 0$$

or

$$Ad \leq 0, d \geq 0, c^T d > 0$$

For all $\lambda \in \mathbb{R}_+$ considr

$$x(\lambda) := \bar{x} + \lambda d$$

we have

$$Ax(\lambda) = A(\bar{x} + \lambda d) = A\bar{x} + \lambda A d \leq b$$

by nonnegativity and nonpositivity.

But $c^T(\bar{x} + \lambda d)$ can be made arbitrily large, demonstrating the unboundedness of (P).

Theorem 2.3.7
If (P) has an optimal solution then (D) also has an optimal solution. Moreover, they have equivalent optimal values.

Proof
By the Lemma.

2.4 Foundamental Theorem of Linear Programming

Theorem 2.4.1 (Foundamental Theorem of Linear Programming)
Every LP is only one of the following:

(i) has an optimal solution

(ii) unbounded

(iii) infeasible

Proof
If (P) is feasible with dual infeasible, it is unblunded. If the dual is feasible, then an optimal solution exists.

2.5 Applications & Interpretations of (Strong) Duality

2.5.1 Complementary Slackness

Conditions and structural characterizations of optimal solutions.

Let

$$(P) : \max c^T x \quad Ax \leq b, x \geq 0$$
$$(D) : \min b^T y \quad y^T A \geq c^T, y \geq 0$$

Let $\bar{x}, \bar{y}$ be feasible solutions to (P), (D) respectively. Then $\bar{x}$ is optimal for (P), $\bar{y}$ optimal for (D) if and only if

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

by Weak and Strong Duality.

In particular, we have

$$c^T\bar{x} \underbrace{=}_{(i)} \bar{y}^T A\bar{x} \underbrace{=}_{(ii)} \bar{y}^T b$$

from the proof of weak duality.

(i) $0 = \sum_{j=1}^n \underbrace{\left(c_j - (\bar{y}^T A)_j\right)}_{\leq 0} \underbrace{\bar{x}_j}_{\geq 0}$ so

$$\forall j \left(\bar{x}_j = 0 \vee (\bar{y}^T A)_j = c_j\right)$$

which gives us a tight contraint for $\bar{y}$

(ii) $0 = \sum_{i=1}^m \underbrace{\bar{y}_i}_{\geq 0} \underbrace{\left((A\bar{x})_i - b_i\right)}_{\leq 0}$ so

$$\forall i \left(\bar{y}_i = 0 \vee (A\bar{x})_i = b_i\right)$$

which gives us a tight contraint for $\bar{x}$

Theorem 2.5.1 (Complementary Slackness Theorem)
$\bar{x}, \bar{y}$ feasible solutions to $(P), (D)$ respectively are optimal if and only if

(a) $\forall j, \bar{x}_j = 0$ or the dual is tight for $\bar{y}$

(b) $\forall i, \bar{y}_i = 0$ or the primal is tight for $\bar{x}$

Example 2.5.2 (applying CS conditions)
Let (P) : $\max 5x_1 + 3x_2 + 5x_3$ such that

$$x_1 + 2x_2 - x_3 \leq 2 \quad (1)$$
$$3x_1 + x_2 + 2x_3 \leq 4 \quad (2)$$
$$-x_1 + x_2 + x_3 \leq -1 \quad (3)$$
$$x_2 \leq 0 \quad (4)$$
$$x_3 \geq 0 \quad (5)$$

With dual (D) : $\min 2y_1 + 4y_2 - y_3$ such that

$$y_1 + 3y_2 - y_3 = 5 \quad (6)$$
$$2y_1 + y_2 + y_3 \leq 3 \quad (7)$$
$$-y_1 + 2y_2 + y_3 \geq 5 \quad (8)$$
$$y_1, y_2, y_3 \geq 0 \quad (9)$$

Is $\bar{x} = (1, -1, 1)^T$ optimal to (P)? If so, there must exist a dual feasible solution $\bar{y}$

satisfying the CS requirements. Since $\nexists j, \bar{x}_j = 0$, we must have

$$
\begin{aligned}
& (6) \\
2\bar{y}_1 + \bar{y}_2 + \bar{y}_3 &= 3 \\
-\bar{y}_1 + 2\bar{y}_2 + \bar{y}_3 &= 5
\end{aligned}
$$

Plugging in values for the CS requirements for $\bar{x}$ we must have

$$
\begin{aligned}
\boxed{\bar{y}_1 = 0} &\vee \bar{x}_1 + 2\bar{x}_2 - \bar{x}_3 = 2 && \text{RHS is false} \\
\bar{y}_2 = 0 &\vee \boxed{3\bar{x}_1 + \bar{x}_2 + 2\bar{x}_3 = 4} && \text{to fulfill CS for } \bar{y} \\
\bar{y}_3 = 0 &\vee \boxed{-\bar{x}_1 + \bar{x}_2 + \bar{x}_3 = -1} && \text{to fulfill CS for } \bar{y}
\end{aligned}
$$

So

$$
\begin{aligned}
\bar{y}_2 + \bar{y}_3 &= 3 \\
2\bar{y}_2 + \bar{y}_3 &= 5
\end{aligned}
$$

and $\bar{y} = (0, 2, 1)^T$ necessarily. After verifying for feasibility of $\bar{y}$ we conclude it is the unique optimal solution as it is the only feasible one satisfying CS.

2.5.2 Valid Inequalities

Definition 2.5.1 (Valid Inequality)
We say that an inequality

$$\alpha^T x \leq \beta, \alpha, x \in \mathbb{R}^n, \beta \in \mathbb{R}$$

is **valid** for a set $S \subseteq \mathbb{R}^n$ if

$$\forall \bar{x} \in S, \alpha^T \bar{x} \leq \beta$$

Example 2.5.3
Consider a polyhedron $P := \{x \in \mathbb{R}^N : Ax \leq b\}$. Every valid inequality for P is implied by an inequality derived via a nonnegative linear combination of the constraints of P.

More rigorously, $\alpha^T x \leq \beta$ is valid for P if and only if

$$\max_{x \in P} \alpha^T x \leq \beta$$

by Strong Duality, this happens if and only if there is a dual feasible y, ie

$$\exists y \geq 0, y^T A \geq \alpha^T$$

such that $b^T y \leq \beta$, so the optimal values are equivalent and

$$y^T A x = \alpha^T x \leq \beta^T y \implies \alpha^T x \leq \beta$$

2.5.3 Geometric Interpretations of Farkas' Lemma

Definition 2.5.2 (cone)
$K \subseteq \mathbb{R}^n$ is a **cone** if

- $0 \in K$
- $\forall x \in K, \forall \lambda \geq 0, \lambda x \in K$
- $\forall x, y \in K, x + y \in K$

Lemma 2.5.4
The intersection of arbitrary family of cones is a cone.

for $S \subseteq \mathbb{R}^n$, $\text{cone}(S)$ denotes the smallest cone containing S.

$$\text{cone}(S) = \bigcap_{i \in I} C_i$$

such that $i \in I \implies S \subseteq C_i$ is the index set of all cones generated by S.

Lemma 2.5.5
If $S \subseteq \mathbb{R}^n$ is finite, ie $S = \{a^{(1)}, \ldots, a^{(k)}\}$, then

$$\text{cone}(S) = \left\{ x : x = \sum_i \lambda_i a^{(i)}, \lambda_i \geq 0 \right\}$$

Let $A = [A_1, \ldots, A_n] \in \mathbb{R}^{m \times n}$. The following are equivalent:

(1) $b \in \mathbb{R}^m, b \notin \text{cone}(\{A_1, \ldots, A_n\})$

(2) $Ax = b, x \geq 0$ has no solutions

(3) $\exists y \in \mathbb{R}^m, y^T A \geq 0, y^T b < 0$. In other words, the hyperplane $\{\delta \in \mathbb{R}^m : y^T \delta = 0\}$ seperates b from the cone.

2.5.4 Geometric Interpretation of Duality

Definition 2.5.3 (tight/active)
We say $\alpha^T x \leq \beta$ is **tight/active** at $\bar{x}$ if $\alpha^T x = \beta$

Proposition 2.5.6
Let $(P) : \max c^T x$ such that $Ax \leq b$, and $\bar{x}$ be a feasible solution to (P).
Let $A^= x \leq b^=$ be the contraints of $Ax \leq b$ which are tight at $\bar{x}$.
Then $\bar{x}$ is an optimal solution to (P) if and only if $c \in \text{cone}(\text{rows of } A^=)$.

Proof
$\bar{x}$ is an optimal solution if and only if $\exists \bar{y}$ such that $\bar{y}$ is a feasible solution to the dual

$$(D) : \min b^T y, y \geq 0, y^T A = c^T$$

and $\bar{y}$ satisfies the CS conditions with $\bar{x}$.

The above is true if and only if $\bar{y} \geq 0, A^T \bar{y} = c$ and $\bar{y}_i > 0 \implies (A\bar{x})_j = b_j$, meaning $(A\bar{x})_j = b_j$ is a constraint from $A^= x \leq b^=$.

Which in turn is true if and only if $c = \sum a_i^T y_i$, i corresponds to rows in A.

We can see cones and the above as some sort of a *one-sided row space*.

Example 2.5.7
$\max x_1 + 3x_2$ such that

$$\begin{aligned} 2x_1 + x_2 &\leq 10 && (1) \\ x_1 + x_2 &\leq 6 && (2) \\ -x_1 + x_2 &\leq 4 && (3) \\ x_1, x_2 &\geq 0 && (4) \end{aligned}$$

We get $x = (1, 5)^T$ and it is optimal since

$$c = (1, 3)^T \in \text{cone}\left(\{(1, 1)^T, (-1, 1)^T\}\right)$$

2.5.5 Physical Interpretation / Intuition

Think of each hyperplane $a_i^T x = b_i$ as a wall, with a free particle inside the feasible region subjected to a force of c.

To reach equilibrium, c should be balanced by other forces (normal force from walls that the particle touches $[A^=]$).

If $\bar{x}$ is resting position, then c is balanced by a nonnegative combination of constraints in $A^=$.

$$-c = \sum_{i\text{: particle touches wall } i} -a_i^T y_i, y_i \geq 0$$

2.5.6 Strong Duality & Farkas' Lemma

Proposition 2.5.8 (Strong Duality Implies Farkas' Lemma)
(P) : $\max 0^T x$ such that $Ax \leq b, x \geq 0$ is infeasible or has optimal value of 0.
(D) : $\min b^T y$ such that $y^T A \geq 0, y \geq 0$ is feasible with $\bar{y} = 0$ as feasible solution.
(P) is infeasible $\iff$ (D) unbounded $\iff$ $b^T y$ has a negative solution.

2.5.7 Economic Interpretation of Duality & Sensitivity Analysis

(P) : $\max c^T x$ such that $Ax \leq b, x \geq 0$ and

$$\sum a_{ij} x_j \leq b_i$$

where a_{ij} is the number of units of resource i to produce 1 units of product j and b_i is the supply of resource i.

(D) : $\min b^T y$ such that $y \geq 0, A^T y \geq c$.

By Strong Duality, $c^T \bar{x} = b^T \bar{y}$ if $\bar{x}, \bar{y}$ are optimal values. If $b_i \to b_i + \epsilon$, $b^T \bar{y} \to b^T \bar{y} + \epsilon \bar{y}$.

$\bar{y}_i$ is the rate of change of optimal value with respect to change in b_i called the **shadow price** of resource i.

2.6 Summary of Duality Theorems for LPs

We begin this section by remarking that complementary slackness conditions are not always helpful!

Example 2.6.1
Consider $\max c^T x$ such that

$$Ax \leq 0, x \geq 0$$

is $x = 0$ optimal?

The dual is $\min 0^T y$ such that

$$A^T y \geq c, y \geq 0$$

Unfortunately, the CS conditions say nothing about y as every x_i is!

In fact, any feasible y would show that $x = 0$ is an optimal solution.

Definition 2.6.1 (LP Equivalence)
(P), (P') are equivalent if

(1) (P) has optimal solutions if and only if (P') does

(2) (P) is infeasible if and only if (P') is

(3) (P) is unbounded if and only if (P') is

(4) certificates of optimality, infeasibility, unboundedness can be "easily" converted between the problems

Example 2.6.2

- $\min c^T x \iff -\max -c^T x$
- $a^T x = \alpha \iff a^T \leq \alpha$ and $-a^T x \leq -\alpha$
- x_j is free $\iff$ we introduce 2 new nonegative variables u_j, v_j with

$$x_j = u_j - v_j, u_j, v_j \geq 0$$

- $a^T x \leq \alpha \iff a^T x + x_{n+1} = \alpha, x_{n+1} \geq 0$

Example 2.6.3
(P): $\max c^T x$ such that

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

(P'): $\max c^T x$ such that

$$\begin{bmatrix} A \\ -A \end{bmatrix} x \leq \begin{bmatrix} b \\ -b \end{bmatrix}, x \geq 0$$

(D): $\min [b^T, -b^T] \begin{bmatrix} u \\ v \end{bmatrix}$ such that

$$\begin{bmatrix} A^T & -A^T \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} \geq c, u, v \geq 0$$

(D'): $\min b^T(u - v)$ such that

$$A^T(u - v) \geq c, u, v \geq 0$$

note that this is the same as $y := u - v$ free as a single variable.

2.6.1 General Formula for LPs

(P) max LP	(D) min LP
i-th constraint is of the $\leq$ type	i-th dual variable of is ≥ 0
i-th constraint is of the $\geq$ type	i-th dual variable of is ≤ 0
i-th constraint is of the $=$ type	i-th dual variable of is free
j-th primal variable is ≥ 0	j-th dual constraint if of the $\geq$ type
j-th primal variable is ≤ 0	j-th dual constraint if of the $\leq$ type
j-th primal variable is free	j-th dual constraint if of the $=$ type

Theorem 2.6.4 (Strong Duality for General LPs)
Let (P), (D) be primal, dual LPs.

1. If (P), (D) have feasible solutions then they both have optimal solutions and optimal objective values are the same.
2. If one of (P) have an optimal solution, then they both do with equivalent optimal objective values.

Theorem 2.6.5 (CS Conditions)
Let (P), (D) be pairs of primal-dual LPs.
Let $\bar{x}$ be feasible in (P) and $\bar{y}$ feasible in (D).
Then $\bar{x}, \bar{y}$ are optimal in thei respective problems if and only if:

$$\forall j \in [n], \bar{x}_j = 0 \vee j\text{-th dual constraint is tight}$$
$$\forall i \in [m], \bar{y}_i = 0 \vee i\text{-th primal constraint is tight}$$

2.7 Geometry of Polyhedra

2.7.1 Notation and Definitions

Definition 2.7.1 (Convex)
$S \subseteq \mathbb{R}$ is convex if forall $x, y \in S$

$$\{\lambda x + (1 - \lambda)y : \lambda \in [0, 1]\} \subseteq S$$

Example 2.7.1
$\varnothing, \{1\}$, ellipsoids, half-spaces, $\mathbb{R}^n$ are convex.

$\{1,2\}$, donut, random V are not convex.

Proposition 2.7.2
$A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, F := \{x \in \mathbb{R}^n : Ax \leq b\}$
F is convex.

Proof
Suppose $x, y \in F$

$$\begin{aligned} A(\lambda x + (1-\lambda)y) &= \underbrace{\lambda}_{\geq 0} \overbrace{Ax}^{\leq b} + \underbrace{(1-\lambda)}_{\geq 0} \overbrace{Ay}^{\leq b} \\ &\leq \lambda b + (1-\lambda) b \\ &= b \end{aligned}$$

Definition 2.7.2 (Closed Half-Space)
$\{x \in \mathbb{R}^n : a^T x \leq \alpha\}$ for some $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$

Note that $\mathbb{R}^n, \varnothing$ are closed half-spaces.

Proposition 2.7.3
Let $\{C_\lambda\}_{\lambda\in\Lambda}$ be a collection of convex sets.

$$\bigcap_{\lambda\in\Lambda} C_\lambda$$

is convex.

Definition 2.7.3 (Polyhedron)
A **polyhedron** is an intersection of finitely many closed half-spaces.

Remark that Polyhedra are convex sets.

In addition, the feasible region of LP problems are polyhedra.

Moreover, the optimal solution sets of LP problems are polyhedra as well.

Definition 2.7.4 (Convex Hull)
$\text{conv}(S), S \subseteq \mathbb{R}^n$ is the smallest convex set containin S.

$$\text{conv}(S) := \bigcap_{S\subseteq H, H \text{ is convex}} H$$

Definition 2.7.5 (Convex Combination)
For $x^{(1)}, \ldots, x^{(k)} \in \mathbb{R}^n$ is

$$\sum_{i=1}^{k} \lambda_i x^{(i)}, \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0$$

Proposition 2.7.4
$S \subseteq \mathbb{R}^n$ is convex if and only if it contains all convex combinations of its elements.

Proof
The forwards direction is trivial. We show the reverse.

Let us argue by induction on k.

For $k = 1, 2$, it is true by definition. Now suppose it holds for all $n \leq k$.

We have

$$\bar{x} := \sum i = 1^{k+1} \lambda_i x^{(i)}$$

where without loss of generality $\lambda_i > 0$.

Write

$$\begin{aligned}
\bar{x} &= (1-\lambda_{k+1}) \underbrace{\left(\sum_{i=1}^{k} \frac{\lambda_i}{1-\lambda_{k+1}} x^{(i)} \right)}_{\in S, \text{induction hypothesis}} + \lambda_{k+1} x^{(k+1)} \\
&= (1-\lambda_{k+1})\hat{x} + \lambda_{k+1} x^{(k+1)} && \text{reduced to base case } k = 2 \\
&\in S
\end{aligned}$$

Corollary 2.7.4.1
A convex hull of S is the set of all convex combinations of elements of S.

Theorem 2.7.5 (Carathéodory, 1907)
Let $S \subseteq \mathbb{R}^n$, every point in $\operatorname{conv}(S)$ can be expressed as a convex combination of at most $n+1$ points in S.

Proof
Let $\bar{x} \in \operatorname{conv}(S)$

$$\bar{x} = \sum_{i=1}^{k} \lambda_i x^{(i)}, k \geq n+2$$

Consider

$$\begin{pmatrix} x^{(1)} \\ 1 \end{pmatrix}, \ldots, \begin{pmatrix} x^{(k)} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$$

They live in $\mathbb{R}^{n=1}$ and must be linearly dependent by the dimension.

There are μ_i

$$\sum_{i=1}^{k} \mu_i x^{(i)} = 0, \sum_{i=1}^{k} \mu_i = 0$$

Let

$$\bar{\alpha} := \max\{\alpha : \lambda + \alpha\mu \geq 0\}$$

this is valid as there must be at least one negative μ_i.

Let

$$\bar{\lambda} := \lambda + \bar{\alpha}\mu$$

and note that at least one entry is zero by construction.

We have

$$\sum \bar{\lambda}_i x^{(i)} = \sum \underbrace{\lambda_i x^{(i)}}_{=\bar{x}} + \alpha \underbrace{\sum \vec{u}_i x^{(i)}}_{=0} = \bar{x}$$

with

$$\sum \bar{\lambda}_i x^{(i)} = \sum \underbrace{\lambda_i}_{=1} + \alpha \underbrace{\sum \vec{u}_i}_{=0} = 1$$

So we reduced $\bar{x}$ to a convex combination of at most $k-1$ points. We can repeat this argument until we express $\bar{x}$ with $n+1$ points.

Definition 2.7.6 (Affine Combination)
If

$$\begin{pmatrix} x^{(1)} \\ 1 \end{pmatrix}, \dots, \begin{pmatrix} x^{(k)} \\ 1 \end{pmatrix} \in \mathbb{R}^{n+1}$$

are linearly dependent in $\mathbb{R}^{n+1}$ then

$$x^{(1)}, \dots, x^{(k)}$$

are affinely dependent in $\mathbb{R}^n$.

2.7.2 Extreme Points

Definition 2.7.7 (Extreme Point)
Let $S \subseteq \mathbb{R}^n$ be convex, we say $\bar{x}$ is an **extreme point** of S if

(1) $\bar{x} \in S$

(2) there do not exist distinct $u, v \in S \setminus \{\bar{x}\}$ such that

$$\bar{x} = \frac{1}{2}(u + v)$$

Equivalently

$$\forall \alpha \in \mathbb{R}^n, \alpha \neq 0, \bar{x} + \alpha \notin S \vee \bar{x} - \alpha \notin S$$

In a polyhedron, extreme points are an intersection of some hyperplanes (defined by some linearly independent tight restrictions).

Theorem 2.7.6
Let $\bar{x} \in S \subseteq \mathbb{R}^n$ be convex. $\bar{x}$ is an extreme point of S if and only if $S \setminus \{\bar{x}\}$ is convex.

Proof (sketch)
If $\bar{x}$ is an extreme point, then

$$S \setminus \{\bar{x}\}$$

is still convex.

Theorem 2.7.7
Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}, \bar{x} \in P$ and $A^= x \leq b^=$ be the constraints that are tight at $\bar{x}$.
$\bar{x}$ is an extreme point $\iff$ $\text{rank}(A^=) = n$

Proof ($\implies$)
Suppose $\text{rank}(A^=) < n$, then $\exists \alpha \in \mathbb{R}^n, \alpha \neq 0$ such that $A^= \alpha = 0$.

We claim that for sufficiently small $\epsilon > 0$, $x^{(1)} = x + \epsilon\alpha, x^{(2)} = x - \epsilon\alpha \in P$, so it cannot be an extreme point.

We have

$$\left.\begin{aligned} A^= x^{(1)} &= A^= x + \epsilon(A^= \alpha) = A^= \bar{x} \\ A^= x^{(2)} &= A^= x + \epsilon(A^= \alpha) = A^= \bar{x} \end{aligned}\right\} = b^=, \text{ So every constraint of } A^= x \leq b^= \text{ is tight at } x^{(1)}, x^{(2)}$$

Now, consider a constraint $a_i^T x \leq b_i$ which is NOT tight at $\bar{x}$. Then

$$\left.\begin{aligned} a_i^T x^{(1)} = \quad & a_i^T \bar{x} + \epsilon(a_i^T \alpha) < b_i + \epsilon(a_i^T \alpha) \\ a_i^T x^{(2)} = \quad & a_i^T \bar{x} - \epsilon(a_i^T \alpha) < b_i - \epsilon(a_i^T \alpha) \end{aligned}\right\} \text{a small } \epsilon \text{ satisfies the inequalities} \leq b_i$$

Since $x^{(1)}, x^{(2)} \in P$, the line segment in between them lie in P. In particular, $\bar{x}$ is on that line so $\bar{x}$ cannot be an extreme point.

Proof ($\impliedby$)
Suppose $\text{rank}(A^=) = n$ but $\bar{x}$ is not an extreme point.

Then $\exists u \neq v \in P, \exists \lambda \in (0,1), \bar{x} = \lambda u + (1-\lambda)v$. It follows that

$$\begin{aligned} b^= &= A^= \bar{x} = A^=(\lambda u + (1-\lambda)v) = \lambda A^= u + (1-\lambda) A^= v \\ 0 &= \lambda A^= u - \lambda b^= - b^= + \lambda b^= + (1-\lambda) A^= v \\ 0 &= \lambda \underbrace{(A^= u - b^=)}_{\leq 0} + (1-\lambda) \underbrace{(A^= v - b^=)}_{\leq 0} \end{aligned}$$

So $A^= u = b = A^= v \implies A^=(v-u) = 0$ but that is a contradiction as $v \neq u$ and A is of full rank.

We present an alternative proof

Lemma 2.7.8
$a \in \mathbb{R}^n, \alpha \in \mathbb{R}, \bar{x} \in \mathbb{R}^n, H := \{s \in \mathbb{R}^n : a^T x \leq \alpha\}$
If $a^T x = \alpha$ and $u, v \in H$ such

$$\frac{1}{2}u + \frac{1}{2}v = \bar{x}$$

then

$$a^T u = a^T v = \alpha$$

Proof
$\alpha \geq \frac{1}{2}a^T u + \frac{1}{2}a^T v = a^T \bar{x} = \alpha$ but

$$a^T u, a^T v \leq \alpha \implies a^T u = a^T v = \alpha$$

Proof ($\Longleftarrow$)
Suppose for a contradiction that $\text{rank}(A^=) = n$ and $\bar{x}$ is NOT an extreme point of the polyhedra.

There must be $u, v \in F \setminus \{\bar{x}\}$ such that

$$\frac{1}{2}u + \frac{1}{2}v = \bar{x}$$

Note that $u, v, \bar{x}$ satisfy every inequality in $A^= x \leq b$ by the lemma which is a contradiction since a matrix with full column rank has unique solutions to every equation.

Corollary 2.7.8.1
(i) if $\text{rank}(A) < n$ there are no extreme points.

(ii) Every Polyhedron has a finite number of extreme points upper bounded by $\binom{m}{n}$

Proof
If $\bar{x}$ is an extreme point and $A^= x \leq b^=$ are the tight constraints of $\bar{x}$, then $\text{rank}(A^=) = n$ and $\bar{x}$ is the unique solution to $A^= x = b^=$.

It follows that the number of extreme points is at most the number of subsystems $A'x \leq b'$ with $\text{rank}(A') = n$. This in turn is bounded above by $\binom{m}{n}$.

In particular, this means that the number of constrainsts must be at least the number of variables in order to have ANY extreme points.

We say $P \subseteq \mathbb{R}^n$ **has a line** if $\exists \bar{x} \in P, d \in \mathbb{R}^n, d \neq 0$ such that

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

Definition 2.7.8 (Pointed Polyhedron)
A polyhedron is **pointed** if it does not contain a line.

Proposition 2.7.9
$P \subseteq \mathbb{R}^n$ is a nonempty, pointed polyhedron if and only if P has an extreme point.

Theorem 2.7.10
Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polyhedron with no lines. If the LP $\max c^T x$ such that $x \in P$ has optimal solution, then it always has an optimal solution that is an extreme point of P.

Proof
Let $\bar{x}$ be the optimal solution to $\max c^T x$ such that $x \in P$ for which the number of tight contraints at $\bar{x}$ is maximized. Let $A^= x \leq b^=$ be the tight constraints.

Suppose $\text{rank}(A^=) < n$, else $\bar{x}$ is an extreme point by our work before. So $\exists \alpha \neq 0, A^= \alpha = 0$.

Now, as $\bar{x}$ is not an extreme point, by the characterization of extreme points, $\exists \epsilon > 0$ such that

$$\bar{x} \pm \epsilon\alpha \in P$$

Note we must have $c^T\alpha = 0$ or else one of the two values has greater objective value than $c^T\bar{x}$, contradicting the optimality of $\bar{x}$.

So all points on the line

$$L = \{\bar{x} + \lambda\alpha : \lambda \in \mathbb{R}\}$$

have $c^T x = c^T \bar{x}$

But since $L \notin P$ WLOG there is a maximum λ^* such that

$$x' = \bar{x} + \lambda^*\alpha \in P$$

x' "activates" one more constraint than $\bar{x}$ while achieving the same objective value, we can now apply the same method and pick up inequalities until reaching n, thus guaranteeing we have found an extreme point.

Theorem 2.7.11
Let $P = \{x \in \mathbb{R}^n : Ax \leq n\}$ be a polyhedron and $\bar{x} \in P$. Then $\bar{x}$ is an extreme point of $P \iff \exists c \in \mathbb{R}^n$ such that $\bar{x}$ is the unique optimal solution to

$$\max c^T x, x \in P$$

2.7.3 Polytopes, Polyhedral Cones, & more Geometric Objects

Definition 2.7.9 (Polytope)
A bounded polyhedron

Theorem 2.7.12
Let $F \subseteq \mathbb{R}^n$ be a polytope.
F is the convex hull of its extreme points.

Proof
If $F = \varnothing$, we are done.

Elsewise, let

$$v^{(1)}, \dots, v^{(k)}$$

be all the extreme points of F and note that there are finitely many by our previous work.

Define

$$\bar{F} := \operatorname{conv}(v^{(1)}, \dots, v^{(k)})$$

and note $\bar{F} \subseteq F$ since F is convex and contains all the points above.

Suppose there is some $\bar{x} \in F \setminus \bar{F}$, so the system

$$\sum_{i=1}^k \lambda_i v^{(i)} = \bar{x}, \sum_{i=1}^k \lambda_i = 1, \lambda \geq 0$$

is infeasible.

By Farkas' Lemma, there is some $\bar{y} \in \mathbb{R}^n, \bar{\alpha} \in \mathbb{R}$ such that

$$\bar{y}^T \bar{x} + \bar{\alpha} < 0, \bar{y}^T v^{(i)} + \bar{\alpha} \geq 0, \forall i \in [k]$$

Consider the LP: $\min \bar{y}^T x$ such that

$$x \in F$$

By asumption, F is bounded and nonempty so there are optimal solutions. In particular, by our previous work, there is an optimal solution being an extreme point.

But none of $v^{(i)}$ are optimal solutions by construction so we have the desired contradiction.

Theorem 2.7.13
The convex hull of any finite subset of $\mathbb{R}^n$ is a polytope.

Definition 2.7.10 (Minkowsky Sum)

$$S + T := \{s + t \in \mathbb{R}^n : s \in S, t \in T\}$$

Definition 2.7.11 (Polyhedral Cone)
A set that is simutaneously a cone and a polyhedron.

Theorem 2.7.14
Let $F \subseteq \mathbb{R}^n$ be a nonempty pointed polyhedon. There is a polytope $P \subseteq \mathbb{R}^n$ and pointed polyhedral cone $K \subseteq \mathbb{R}^n$ such that

$$F = P + K$$

The decomposition above is not unique in general but if we can get one using the extreme points of F.

Theorem 2.7.15
P is a polyhedron $\iff$ $P = Q + C$ where Q is a polytope, C is a polyhedral cone.

Lemma 2.7.16
$\max c^T x$ such that $Ax \leq b$ is unbounded $\iff$ $\max c^T x$ such that $Ax \leq 0$ is unbounded.

Proof
Consider the duals of the LPs in question.

By Farkas' Lemma, $\min b^T y$ such that $c^T \leq y^T A$ is infeasible if there is some $0 \leq \alpha \in \mathbb{R}^n$ such that

$$A\alpha = 0, c^T \alpha > 0$$

But note that the conditions does not depend on b at all! This means that α is also a certificate of infeasibility for $\min 0^T y$ such that $c^T \leq y^T A$.

Proof ($\implies$, Case I)
Let us consider $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ if it is pointed.

Take

$$Q = \text{conv}(S)$$

where S is the set of extreme points of P, as well as

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

If $P = \emptyset$, then $\emptyset = Q$, so we certainly have $P = Q + C$. Otherwise, let $u \in Q, v \in C$, we have

$$A(u + v) = Au + Av \leq Au \leq b$$

so $Q + C \subseteq P$.

Now, suppose $\exists \bar{x} \in P \setminus (Q + C)$. So there is some $c \in \mathbb{R}^n$ such that

$$\max_{x \in P} c^T x > \max_{x \in Q+C} c^T x \quad (*)$$

(check)

The LP $\max_{x \in Q+C} c^T x$ is feasible as the feasible region is by assumption nonempty. But LP $\max c^T x$ such that $x \in P$ is unbounded $\iff$ the LP $\max c^T x$ such that $x \in C$ is unbounded by the lemma above. This means that the LP $\max_{x \in P} c^T x$ is also bounded!

However, this means there is an optimal solution to $\max_{x \in P} c^T x$ which means there is an optimal solution in Q as it contain all the extreme points of P. This contradicts (*).

Proof ($\implies$, Case II)
Now suppose P contains a line.

We have $\text{rank}\, A < n$. Let

$$L := \{\alpha \in \mathbb{R}^n : A\alpha = 0\}$$

and note that this has dimension at least 1 since $\text{rank}\, A < n$.

Define

$$P' := P \cap L^\perp$$

where

$$L^\perp := \left\{x \in \mathbb{R}^n : x^T = z^T A, z \in \mathbb{R}^n\right\}$$

which is the row space of of A. Note that $L^\perp$ is a polyhedron as it is the projection of A. Also note that L, L^T are orthogonal as the null space is the space of vectors which are orthogonal to the rows of A, which the rows space is precisely the span of the rows of A.

First we claim that P' is a pointed polyhedron. Suppose not, then P' contains a line, meaning there is some $\bar{x} \in P', d \in \mathbb{R}^n, d \neq 0$ such that

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

Note $\bar{x} + d \in P' \implies \bar{x} + d \in L^\perp$.

We must have $Ad = 0$ ie $d \in L$, else by choosing a sufficiently large λ, $A(\bar{x} + \lambda d) > b$. Then, by orthogonality:

$$0 = (\bar{x} + d)^T d = \bar{x}^T d + d^T d = z^T A d + d^T d = d^T d \implies d = 0$$

But then, the supposed line does not exist!

We then claim that $P = P' + L$

To see $\subseteq$, let $x = p + q \in P$ with $p \in L, q \in L^T$. Then

$$q \in P \cap L^T = P'$$

To see $\supseteq$, let $u \in P', v \in L$ so

$$A(v + u) = Au \leq b \implies u + v \in P$$

Proof ($\impliedby$)
Since Q, C are polyhedra, their Minkowski Sum : P is also a polyhedron.

Proposition 2.7.17
If $K \subseteq \mathbb{R}^n$ is a polyhedral cone, then there is an A such that

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

Proof
Consider any tight restriction $a_i^T x \leq b_i$, we can scale x by any positive factor and it will still be in P.

The only way that is possible is if $b_i = 0$.

Definition 2.7.12 (Affine Subspace)
The solution set of $Ax = b$ for some A, b.

Note that every linear subspace of $\mathbb{R}^n$ can be expressed as

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

for some $A \in \mathbb{R}^{m \times n}$.

Suppose

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

Then there is some l such that $Al = b$.

Then

$$S = l + \{x : Ax = 0\}$$

Definition 2.7.13 (Lineality Space)
The largest affine subspace contained in a polyhedron.

Theorem 2.7.18
Let $F \subseteq \mathbb{R}^n$ be a nonempty polyhedron.
There exist a pointed polyhedral cone $K \subseteq \mathbb{R}^n$, and a polytope $P \subseteq \mathbb{R}^n$ such that

$$F = P + K + L$$

where L is the lineality space of F.

2.7.4 Additional Content

Theorem 2.7.19
Let $C \subseteq \mathbb{R}^n$ be a compact convex set and $S \subseteq C$.
The following are equivalent:

(i) $\overline{\operatorname{conv}(S)} = C$

(ii) $\inf\{h^T x : x \in S\} = \min\{h^T x : x \in C\}$

(iii) $\operatorname{ext} C = \bar{S}$

Theorem 2.7.20 (Characterization of Polytopes)
P is a polytope $\iff$ $P = \operatorname{conv}(S)$ for a finite set $S \subseteq \mathbb{R}^n$.
Moreover, we take S to be the set of extreme points of P.

Proof ($\implies$)
Let P be a polytope, and S the set of extreme points of P. We have $\operatorname{conv}(S) \subseteq P$ as P is convex.

Suppose there is an $\bar{x} \in P \setminus \operatorname{conv}(S)$. Then there are α, β such that $\alpha^T \bar{x} > \beta$ but $\alpha^T x < \beta$ for all $x \in \operatorname{conv}(S)$.

So $\max \alpha^T x, x \in P > \max \alpha^T x, x \in \operatorname{conv}(S)$.

Since the LP has an optimal solution, it must have an optimal solution in S, the set of extreme points.

We have the desired contradiction.

Proof ($\Longleftarrow$)
It suffices to show that any convex hull is a polyhedron. Then, the finiteness forces us to have a bounded polyhedron.

We claim for $\mathbb{R}^n \supseteq P := \{x_1, \dots, x_{k_0}\}$,

$$\operatorname{conv}(P) = \left\{ \sum_{i=1}^{k} \lambda_i x_i : k \leq k_0, \sum_{i=1}^{k} \lambda_i = 1, \lambda_i \geq 0 \right\}$$

To see $\supseteq$, we argue by induction that any convex set C containing P also contains the RHS.

For the case $k = 1$ this is trivial.

Elsewise, we can consider $0 \leq \lambda' := \sum_{i=1}^{k-1} \lambda_i \leq 1$. Scale this down and define $\mu_i := \frac{\lambda_i}{\lambda'}$.

By the induction hypothesis, $q := \sum_{i=1}^{k-1} \mu_i x_i \in C$, since $\sum_{i=1}^{k-1} \mu_i = 1$ by construction.

But then

$$\sum_{i=1}^{k} \lambda_i x_i = \lambda' \left(\sum_{i=1}^{k-1} \mu_i x_i \right) + \lambda_k x_k \subseteq C$$

since $1 = \lambda' + \lambda_k = \sum_{i=1}^{k-1} + \lambda_k$.

Now to see $\subseteq$, it suffices to show that the RHS is a convex set.

Let $x = \sum_{i=1}^{k} \lambda_i x_i, y = \sum_{i=1}^{k} \mu_i x_i$.

For any $\xi \in [0, 1]$, we have

$$\sum_{i=1}^{k} (\xi \lambda_i + (1 - \xi)\mu_i) = 1 \implies \xi x + (1 - \xi) y \in \text{RHS}$$

Theorem 2.7.21
P is a polyhedral cone $\iff P = \operatorname{cone}(S)$ for a finite set $S \subseteq \mathbb{R}$

Proof ($\Longrightarrow$)
Let P be a polyhedral cone, and $Q = \{x \in P : -1 \leq x_j \leq 1, \forall j \in [n]\}$. We are able to do this as we can freely "scale" values in a cone and stay within a cone.

Q is a polytope, so $Q = \text{conv}(S)$ where $S = \left\{q^{(1)}, \ldots, q^{(k)}\right\} \subseteq \mathbb{R}^n$. Since $P \supseteq Q \supseteq S$ and P is a cone, we must have $\text{cone}(S) \subseteq P$.

Let $x \in P$, there is some $\delta \in \mathbb{R}$ such that

$$\delta > 0, \frac{x}{\delta} = \sum \lambda_i q^{(i)}, \lambda_i \geq 0, \sum \lambda_i = 1$$

So $x = \delta \sum \lambda_i q^{(i)} \implies x \in \text{cone}(S)$. In other words, $P \subseteq \text{cone}(S)$.

Proof ($\Longleftarrow$)
If $S = \left\{q^{(1)}, \ldots, q^{(k)}\right\}$, then

$$\text{cone}(S) = \left\{\sum \lambda_i q^{(i)} : \lambda_i \geq 0, \forall i\right\}$$

We show that this is a polyhedron, and so is by definition a polyhedral cone.

$$\begin{aligned}\text{cone}(S) &= \left\{x \in \mathbb{R}^n : \exists 0 \leq \lambda \in \mathbb{R}^n, x = \left[q^{(1)}, \ldots, q^{(k)}\right] \lambda\right\} \\ &= \left\{x \in \mathbb{R}^n : \exists 0 \leq \lambda \in \mathbb{R}^n, 0 \geq \left[x, q^{(1)}, \ldots, q^{(k)}\right] \begin{bmatrix} 1 \\ -\lambda \end{bmatrix}, 0 \geq \left[x, q^{(1)}, \ldots, q^{(k)}\right] \begin{bmatrix} 1 \\ \lambda \end{bmatrix}\right\}\end{aligned}$$

Through Fourier-Motzkin Elimination, we can completely eliminate λ_i's, arriving at a system of inequalities independent of λ, thus demonstrating that $\text{cone}(S)$ is in fact a polyhedron.

Proposition 2.7.22
(i) If S, T are polyhedra, $S + T$ is also a polyhedron
(ii) If S, T are polytope, $S + T$ is also a polytope
(iii) If S, T are polyhedral cones, $S + T$ is also a polyhedral cone

Proof (i)
Suppose

$$\begin{aligned} S &= \{x \in \mathbb{R}^n : Ax \leq b\} \\ T &= \{x \in \mathbb{R}^n : \alpha x \leq \beta\}\end{aligned}$$

We then have

$$\begin{aligned} S+T &= \{x \in \mathbb{R}^n : x = s+t, As \le b, \alpha t \le \beta\} \\ &= \{x \in \mathbb{R}^n : As \le b, \alpha(x-s) \le \beta\} \\ &= \left\{ x \in \mathbb{R}^n : \begin{bmatrix} 0 & A \\ \alpha & -\alpha \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} \le \begin{bmatrix} b \\ \beta \end{bmatrix} \right\} \end{aligned}$$

We can then apply Fourier-Motzkin Elimination to take s out of the picture and complete the proof.

Proof (ii)
By our work above we know $S = \text{conv}(s_1, \dots, s_k), T = \text{conv}(t_1, \dots, t_l)$. We now claim

$$S+T = \text{conv}_{i \le k, j \le l}(s_i + t_j)$$

Let $s \in S, t \in T$.

To see $\subseteq$, consider $s+t \in S+T$, we have

$$\begin{aligned} s &= \sum_{i=1}^{k} \lambda_i s_i, \lambda \ge 0, \sum \lambda_i = 1 \\ t &= \sum_{j=1}^{l} \mu_j t_j, \mu \ge 0, \sum \mu_i = 1 \\ s+t &= \sum_{i=1}^{k} \left(\sum_{j=1}^{l} \mu_i \right) \lambda_i s_i + \sum_{j=1}^{l} \left(\sum_{i=1}^{k} \lambda_i \right) \mu_j t_j \\ &= \sum_{i=1}^{k} \sum_{j=1}^{l} \lambda_i \mu_j \left(s_i + t_j\right) \end{aligned}$$

with $\sum_{i=1}^{k} \sum_{j=1}^{l} \lambda_i \mu_j = \left(\sum_{i=1}^{k} \lambda_i \right) \left(\sum_{j=1}^{l} \mu_j \right) = 1$

To see $\supseteq$, it suffices to show that $S+T$ is convex. Consider

$$\xi \left(s^{(1)} + t^{(1)}\right) + (1-\xi)\left(s^{(2)} + t^{(2)}\right) = \xi s^{(1)} + (1-\xi) t^{(1)} + \xi t^{(2)} + (1-\xi) t^{(2)}$$

which is cleary in $S+T$ by definition, so we are done.

Proof (iii)
We claim $S+T = \text{cone}_{i\le k, j\le l}(s_i, t_j)$ where

$$\begin{aligned} S &= \text{cone}(s_i) \\ T &= \text{cone}(t_j) \end{aligned}$$

To see $\subseteq$, we first decompose an element from the Minkowski Sum as two elements from S, T, we then write it as a nonegative sum of the generating sets of S, T respectively.

To see $\supseteq$ we note that $S+T$ is a cone since we can add nonnegative linear combinations of items from S, T freely decompose the sum as a binary addition between an element from S, T.

2.8 Simplex Method

2.8.1 Introduction

Definition 2.8.1 (Simplex)
$S \subseteq \mathbb{R}^n$ such that

$$S = \text{conv}(v^{(1)}, \dots, v^{(n+1)})$$

where the $v^{(i)}$'s are affinely independent.

Given $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m, c \in \mathbb{R}^n$ and consider: $\max c^T x$ subject to

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

Assume $\text{rank}(A) = m$. If not, we can apply Gaussian Elimination on $[A|b]$ to check either

(i) $Ax = b$ is infeasible

(ii) $\text{rank}\, A = m$

(iii) $Ax = b$ has redundant equations

Definition 2.8.2 (Basis)
We say $B \subseteq [n]$ is a **basis of** A if A_B is square and nonsingular.
Equivalently, $|B| = m$, $\text{rank}(A_B) = m$.

Let $B \subseteq [n]$ be a basis and $N = [n] \setminus B$. The system $Ax = b, x_N = 0$ has a unique solution

$$x_B = A_B^{-1} b, x_N = 0$$

we say that the above is the **basic solution** corresponding to B.

Definition 2.8.3 (Basic Solution)
$\bar{x}$ is called a basic solution to $Ax = b$ if there is a basis B such that $\bar{x}$ is the basic solution corresponding to B.

Definition 2.8.4 (Basic Feasible Solution)
We say that a basic solution $\bar{x} \geq 0$ is a **basic feasible solution** (BFS).

For a basis B, $x_j, j \in B$ are basic variables with others being non-basic variables.

Theorem 2.8.1
Let $A \in \mathbb{R}^{m \times n}, \operatorname{rank} A = m, b \in \mathbb{R}^m, F = \{x \in \mathbb{R}^m : Ax = b, x \geq 0\}$.
Suppose $\bar{x} \in F$. TFAE

(1) $\bar{x}$ is a basic feasible solution to $Ax = b, x \geq 0$

(2) $\{A_j : \bar{x}_j > 0\}$ is linearly independent

(3) $\bar{x}$ is an extreme point of F.

Let B be a basis of A and $\bar{x}$ be the basic feasible solution determined by B. Consider

$$\begin{aligned} \max c^T x \\ A_B x_B + A_N x_N &= b \\ x &\geq 0 \end{aligned}$$

There are unique solutions

$$\begin{aligned} x_B &= A_B^{-1} b - A_B^{-1} A_N x_N \\ \bar{y} &= A_B^{-T} c_B \end{aligned}$$

Definition 2.8.5 (Feasible Basis)
A subset

$$B \subseteq [n]$$

is a feasible basis for (P) if B is a basis of A and the BFS determined by B is feasible in (P).

Note that a feasible basis determines a unique BFS of (P).

Given a reasible basis B of (P), how can we improve the objective value?

Since x_B is "locked" by feasibility, we must take some $j \in N$ and set it to nonzero value. Assume we pick $k \in N$. To maintain feasibility of the new solution, we must set

$$x_k := \alpha \geq 0$$

Moreover, to maintain $Ax = b$ by the relation

$$A_B x_B + A_k x_k = b$$

we actually need

$$\begin{aligned} x_B &= A_B^{-1} b - (A_B^{-1} A_k)\alpha \geq 0 \\ \alpha &\geq 0 \end{aligned}$$

Define $d := A_B^{-1} A_k$. We can determine how large α can be set. We must choose it so that for all $j \in B, d_j > 0$, then

$$\bar{x}_j - \alpha d_j \geq 0 \implies \alpha \leq \frac{x_j}{d_j}$$

Therefore, we can let

$$\alpha := \min\left\{\frac{\bar{x}_j}{d_j} : j \in B, d_j > 0\right\}$$

and our new feasible solution is of the form

$$x' := x + \alpha \bar{d}$$

where

$$\bar{d}_j := \begin{cases} -d_j, & j \in B \\ 1, & j = k \\ 0, j \in N \setminus \{k\} \end{cases}$$

However, if $d_j \leq 0$ for all $j \in B$, then the feasible region F of (P) is unbounded since F contains the ray

$$x(\lambda) := x + \lambda \bar{d}$$

We will ignore this case for now.

Proposition 2.8.2
Suppose that the above does not occur, then x' is a BFS.

Proof
The index l which achieved the minimal value of α can be removed from B so

$$B' := B \cup \{k\} \setminus \{l\}$$

is the desired basis.

How can we be sure that the $k \in N$ that we picked actually improves the objective value z?

Write

$$c^T x = c_B^T x_B + c_N^T x_N$$

For all basic feasible solutions x_B

$$\begin{aligned} c^T x &= c_B^T A_B^{-1} b - c_B^T (A_B^{-1} A_N x_N) + c_N^T x_N \\ &= c_B^T A_B^{-1} b + (c_N^T - c_B^T A_B^{-1} A_N) x_N \end{aligned}$$

Let $\bar{y}$ be the unique solution to the system $A_B^T y = c_B$ so that

$$\begin{aligned} c^T x &= \underbrace{c_B^T A_B^{-1} b}_{\text{current objective value}} + (c_N^T - \bar{y}^T A_N) x_N \\ &= \text{constant} + \sum_{j \in N} \bar{c}_j x_j \end{aligned}$$

where $\bar{c} = c_N^T - \bar{y}^T A_N$.

If $\bar{c} \le 0$, then $\bar{y}$ satisfies $A^T \bar{y} \ge c$.

If both $\bar{x}, \bar{y}$ are feasible, by CS Conditions, they would be optimal in their respectively problems.

Since $(A^T)^{-1} = (A^{-1})^T$, we will then write this as

$$A^{-T}$$

Proposition 2.8.3
If $\bar{c}_N \le 0$, then the current BFS is optimal in (P).

Proof
For every feasible solution x of (P),

$$z := \bar{c}^T x + \bar{v} = 0^T x_B + \overbrace{\bar{c}_N^T \underbrace{x_N}_{\ge 0}}^{\le 0} \le c_B^T A_B^{-1} b = c^T x$$

If $\bar{c}_N < 0$, then $\bar{x}$ is the unique optimal solution.

Proof (alternative)
The dual of (P) is (D):

$$\begin{aligned} \min\, & b^T y \\ & A^T y \ge c \end{aligned}$$

Consider $\bar{y} = A_B^{-T} c_B$. Then

$$A^T \bar{y} \geq c \iff 0 \geq c - A^T \bar{y} = \bar{c}$$

Therefore, $\bar{y}$ is a feasible solution of (D) if and only if $c_N \leq 0$. We have feasible solutions $\bar{x}, \bar{y}$ for (P), (D), with equivalent objective values

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

which are optimal by duality.

Note that the converse to the lemma is NOT necessarily true.

2.8.2 Simplex Details

The input is

$$(A, b, c, \bar{x}, B)$$

such that

(i) A, b, c defines an LP (P), in SEF

(ii) B is a feasible basis for (P), determining a BFS $\bar{x}$

1) Solve $A_B^T y = c_B$ with solution $\bar{y}$
2) let $\bar{c}_N := c_N - A_N^T \bar{y}$. If $\bar{c}_N \leq 0$, then we have optimal solutions $\bar{x}, \bar{y}$
3) Pick $k \in N$ such that $\bar{c}_k > 0$
4) Solve $A_B d = A_k$ with solution d
5) Define

$$\bar{d}_j := \begin{cases} 1, & j = k \\ 0, & j \in N \setminus \{k\} \\ -d_j, & j \in B \end{cases}$$

If $\bar{d} \geq 0$, then stop, as (P) is unbounded with certificate $\bar{x}, \bar{d}$. This is due to the fact that $c^T x(\lambda) = c^T x + \lambda \bar{c}_k \to \infty$

6) Compute

$$\bar{\alpha} := \frac{\bar{x}_l}{-\bar{d}_l}$$

where l is the index achieving minimal ratio for $\bar{d}_l < 0$

7) Take the new basis

$$B' := B \cup \{k\} \setminus \{l\}$$

and new feasible solution

$$\bar{x}' := \bar{x} + \alpha\bar{d}$$

8) goto 1) with $B = B', \bar{x} = \bar{x}'$

2.8.3 Termination

Theorem 2.8.4
The Simplex Method applied to LP problems in SEF with a basic feasible solution termintes in at most

$$\binom{m}{n}$$

iterations provided that $\bar{\alpha} > 0$ (when all $\bar{x}_i \neq 0$) in each iteration.
When the algorithm stops, it gives either a certificate of optimality or unboundedness.

Definition 2.8.6 (Degenerate)
A BFS $\bar{x}$ determined by basis B of A is degenerate if $\bar{x}_i = 0$ for some $i \in B$.

Notions of degenerate basic feasible soutions and degenerate basis are defined similarly.

When degeneracy happens, $\bar{\alpha} = 0$ and the algorithm may not make progress. the improvement is in the choice of k or l. When there is a tie, choose the possible k, l with the lowest index. This is the smallest index rule and ensures termination.

Definition 2.8.7 (Nondegenerate)
An LP problem (P) with constraints

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

is nondegenerate if every basis of A is nondegenerate.

Theorem 2.8.5 (Bland's Rule)
The Simplex Method applied to LP problems in SEF with a basic feasible solution and utilizing the smallest subscript rule (Bland's Rule) termintes in at most

$$\binom{m}{n}$$

When the algorithm stops, it gives either a certificate of optimality or unboundedness.

Proof
Assignment 3.

2.9 Lexicographic Simplex Method

Our goal is to remove degeneracy.

Consider

$$\begin{aligned} \max z &:= c^T x \\ Ax &= b \\ x &\geq 0 \end{aligned}$$

The idea is to perturb the RHS in a way that we never get

$$(A_B^{-1} b)_j = 0$$

for any $i \in B$ for any basis B.

Consider (P') given by

$$\begin{aligned} \max z &:= c^T x \\ Ax &= \begin{array}{c} \begin{array}{ccccc} & \epsilon_1 & \epsilon_2 & \dots & \epsilon_m \end{array} \\ \begin{pmatrix} b_1 & 1 & 0 & \dots & 0 \\ b_2 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ b_m & 0 & 0 & \dots & 1 \end{pmatrix} \end{array} \\ x &\geq 0 \end{aligned}$$

where $1 >> \epsilon_1 >> \dots >> \epsilon_m > 0$.

We apply the Simplex Method to (P'), in computing

$$\min \left\{ \frac{\bar{x}_i}{d_i} \right\}$$

we need to compare expressions like

$$\begin{aligned} \beta_0 + \beta_1 \epsilon_1 + \beta_2 \epsilon_2 + \dots + \beta_m \epsilon_m \\ \gamma_0 + \gamma_1 \epsilon_1 + \gamma_2 \epsilon_2 + \dots \gamma_m \epsilon_m \end{aligned}$$

We can determine that the first expression is lexicographically larger than the second if for the smallest i such that

$$\beta_i \neq \gamma_i$$

we have

$$\beta_i > \gamma_i$$

For any basis B of A, the corresponding x_B for (P') is

$$x_B = \begin{array}{c} \begin{array}{cc} \phantom{A_B^{-1}b} & \epsilon \end{array} \\ \begin{pmatrix} A_B^{-1}b & A_B^{-1} \end{pmatrix} \end{array}$$

and A_B^{-1} cannot have zero rows.

So the Lexicographic Simplex Method induces a total order on the bases.

Proposition 2.9.1
The LP problem (P') is nondegenerate.

Apply the Simplex Method to (P') is applying the Lexicographic Simplex Method to (P).

Theorem 2.9.2
Lexoicographic Simplex Method applied to (P) with a starting BFS terminates in at most

$$\binom{m}{n}$$

iterations.
The resulting basis from the Lexographic Simplex Method proves the same claim for (P).

2.9.1 Summary

Degeneracy can lead to both cycling and stalling where cycling is problematic in theory and stalling is problematic in practice.

We can avoid this by perturbing b be a small amount, but different "independent" amounts for each b_i.

Remark that Non-degeneracy is generic in the sense that randonly generated LP's are non-degenerate. However, LPs in practice are almost always fomulated, which leads to degeneracy.

2.10 Two Phase Method

Given (P)

$$
\begin{aligned}
\max z := c^T x \\
Ax = b \\
x \geq 0
\end{aligned}
$$

we can introduce artificual variables $x_{n+1}, \ldots, x_{n+m}$ and solve the auxiliary LP (P-AUX)

$$
\begin{aligned}
\max -s \\
\begin{bmatrix} A & I \end{bmatrix} \begin{bmatrix} x \\ s \end{bmatrix} = b \\
x, s \geq 0
\end{aligned}
$$

The basis corresponding to s yields a trivial BFS of (P-AUX). But it is certainly bounded, and so has an optimal solution.

Proposition 2.10.1
(P-AUX) has an optimal objective value equal to zero if and only if (P) has a feasible solution.

We can then solve (P-AUX) with the Simplex Method. If the optimal value is not zero, then (P) is infeasible with the last $\bar{y}$ being a certificate of infeasibility.

2.10.1 Dual of the Auxiliary Program

(D-AUX) given by

$$
\begin{aligned}
\min b^T y \\
A^T y \geq 0 \\
I y \geq -1
\end{aligned}
$$

In the case that the optimal value of (P-AUX) is not zero, the last $\bar{y}$ computed by the Simplex Method is an optimal solution of (D-AUX) with

$$
b^T y < 0
$$

So $\bar{y}$ is a certificate of infeasibility for (P).

Notice that we then have an algorithm proof of the Fundamental Theorem of Linear Programming (with some tie-breaking rule).

2.10.2 Alternative Approach

(P')

$$\begin{aligned} \max 0^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

with dual (D')

$$\begin{aligned} \min b^T y \\ A^T y \geq 0 \end{aligned}$$

Notice that 0_m is a BFS of of (D').

We can then run Simplex on (D') and if it is unbounded, (P') and therefore (P) is infeasible. Else if there is an optimal solution, (P') and therefore (P) is feasible.

2.11 Abridged Complementary Slackness

Theorem 2.11.1
(P)

$$\begin{aligned} \max c^T x \\ Ax = b \\ x \geq 0 \end{aligned}$$

(D)

$$\begin{aligned} \min b^T y \\ A^T y \geq c \end{aligned}$$

Suppose (P) has a optimal solution. The (P), (D) have optimal solutions $\bar{x}, \bar{y}$ such that for all $j = 1, 2, \ldots, n$

$$\bar{x}_j (A^T \bar{y} - c)_j = 0 \wedge \bar{x}_j + (A^T \bar{y} - c)_j > 0$$

so it is complementary slackness, with with non-tight part being "positive"

Proof
We will prove that for every $j \in [n]$ either there is some $x^{(j)}$ optimal in (P) such that $x_j^{(j)} > 0$ or there is some $y^{(j)}$ optimal in (D) such that

$$(A^T y^{(j)} - c) > 0$$

Since (P) has an optimal solution by Strong Duality Theorem, so does (D). Let $\bar{z}$ be the optimal value.

Consider (Pj)

$$\begin{aligned} \max e_j^T x& \\ Ax &= b \\ c^T x &\geq \bar{z} \\ x &\geq 0 \end{aligned}$$

As well as (Dj)

$$\begin{aligned} \min b^T y + \bar{z}\eta& \\ A^T y + \eta c &\geq e_j \\ \eta &\leq 0 \end{aligned}$$

If (Pj) has a feasible solution with positive objective value, we have our $x^{(j)}$ as desired.

Else, we may assume the optimal value of (Pj) is zero. Then, by the Strong Duality applied to (Pj) and (Dj), there is some $\hat{y} \in \mathbb{R}^m, \hat{\eta} \in \mathbb{R}$ such that

$$\begin{aligned} \hat{\eta} &\leq 0 \\ A^T \hat{y} + \hat{\eta} &\geq e_j \\ b^T \hat{y} &= -\bar{z}\hat{\eta} \end{aligned}$$

<u>Case I</u>: If $\hat{\eta} = 0$, then $A^T \hat{y} \geq e_j, b^T \hat{y} = 0$. Since **(D)** has an optimal solution, say y^*, we can take

$$y^{(j)} := y^* + \hat{y}$$

and $y^{(j)}$ is optimal in (D) with $(A^T y^{(j)} - c)_j \geq 1 > 0$.

<u>Case II</u>: If $\hat{\eta} < 0$, we can take

$$y^{(j)} := \frac{\hat{y}}{-\hat{\eta}}$$

Note that

$$\begin{aligned} A^T y^{(j)} &\geq c - \frac{1}{\hat{\eta}} e_j \\ b^T y^{(j)} &= \bar{z} \end{aligned}$$

That is, $y^{(j)}$ is an optimal solution of (D) with $(A^T \hat{y}^j - c)_j \geq -\frac{1}{\hat{\eta}} > 0$.

Now, let

$$B := \{j : \exists x^{(j)}\}, N := [n] \setminus B$$

Define

$$\bar{x} := \frac{1}{|B|} \sum_{j \in B} x^{(j)}, \bar{y} := \frac{1}{|N|} \sum_{j \in N} y^{(j)}$$

If $B = \varnothing$, then $\bar{x} := 0$ is the unique optimal solution of (P). But $\bar{y}$ still satisfies the statement.

If $N = \varnothing$, then there is a unique optimal solution $\bar{y}$ with $\operatorname{rank} A = m$ (which we may assume WLOG), satisfying

$$A^T \bar{y} = c$$

Again $\bar{x}$ still satisfies the statement.

3 Combinatorial Optimization

3.1 Motivational Examples

Example 3.1.1 (Assignment Problem)
Suppose we have a set of jobs J, set of workers W.

We are given $c_{ij} \in \mathbb{R}$ for each $i \in W, j \in J$ describing the "compatibility" of worker i to job j.

We want to assign workers to jobs bijectively such that the summation of c_{ij} for the assignment pairs is maximized.

Let
$$x_{ij} := \begin{cases} 1, & \text{worker } i \text{ assigned to job } j \\ 0, & \text{otherwise} \end{cases}$$

We wish to maximize $\max \sum_{i \in W} \sum_{j \in J} c_{ij} x_{ij}$ subject to

$$\begin{aligned} \sum_{j \in J} x_{ij} &= 1 && \forall i \in W \\ \sum_{i \in W} x_{ij} &= 1 && \forall j \in J \\ x_{ij} &\geq 0 \\ x_{ij} &\in \mathbb{Z} \end{aligned}$$

(AP) has feasible solutions if and only if $|W| = |J|$.

The linear equations contain a redundant one.

Definition 3.1.1 (Graph)
$G = (V, E)$ in a graph.
E is a subset of pairs $\{u, v\}$ where $u \neq v \in V$.

We will define graphs to be by default simple graphs in this course.

Definition 3.1.2 (Matching)
A matching in a graph G is a subset $M \subseteq E$ such that every vertex $v \in V(G)$ is incident with at most one edge in M.

Definition 3.1.3 (Perfect Matching)
A matching in G is perfect if it satisfies every vertex (ie cardinality is exactly $\frac{|V|}{2}$.

Definition 3.1.4 (Maximum Weight Matching Problem)
Let $w_e \in \mathbb{R}$ for every $e \in E$.
The said problem in G is to find a matching M in such that

$$\sum_{e \in M} w_e$$

is maximized

Definition 3.1.5 (Maximum Weight Perfect Matching Problem)
The said problem is a maximum weight matching problem with the added constraint that the matching must be perfect.

Definition 3.1.6 (Bipartite)
A graph $G = (V, E)$ is bipartite if there is a bipartition A, B of V such that

$$\{u, v\} \in E \implies u \in A, v \in B \vee \text{vice versa}$$

Definition 3.1.7 (Complete)
A graph is complete if for every $u, v \in V$

$$\{u, v\} \in E$$

We can formalize the Assignment Problem as a Maximum Weight Perfect Matching Problem in a (complete) bipartite graphs.

3.2 Pure Integer Programming Problems

$$\begin{aligned} \max c^T x& \\ Ax &\leq b \\ x &\in \mathbb{Z}^n \end{aligned}$$

The LP relaxation of an Integer Program is (LP)

$$\begin{aligned} \max c^T x \\ Ax \le b \end{aligned}$$

If we solved the LP and obtained an optimal solution $\bar{x}$ and $\bar{x} \in \mathbb{Z}^n$, then we are done!

Regardless if $\bar{x} \in \mathbb{Z}^n$, $c^T \bar{x}$ is an upper bound on the optimal objective value of the IP.

Definition 3.2.1 (Integer Hull)
If

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

the integer hull of P is

$$P_I := \text{conv}(P \cap \mathbb{Z}^n)$$

Note that P_I is the set of feasible solutions of our (IP). In some sense, P is the "closest" approximation to P_I.

Theorem 3.2.1
Let $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ and

$$S := \{x \in \mathbb{Z}^n : Ax \le b\}$$

where S is bounded.
Then conv(S) is a polytope.

By boundedness, S is finite and thus the convex hull is just the convex combination of all the finite points. It is by definition a Simplex, which is a polytope.

Theorem 3.2.2
Let $A \in \mathbb{Q}^{m \times n}, b \in \mathbb{Q}^m$ and

$$P := \{x \in \mathbb{Z}^n : Ax \le b\}$$

Then P is a polyhedron.

Corollary 3.2.2.1
Let $A \in \mathbb{Q}^{m \times n}, b \in \mathbb{R}^m$

$$S := \{x \in \mathbb{Z}^n : Ax \le b\}$$

Then conv(S) is a polyhedron.

Proof
Let A, b, S be as above and $\xi \in \mathbb{Z}_{++}$ be the LCM of all denominator of rational A_{ij}.

$\xi A \in \mathbb{Z}^{m\times n}$.

Then $S = \{x \in \mathbb{Z}^n : \underbrace{\xi A x}_{\in \mathbb{Z}^m} \leq \xi b\}$. But this is the same as

$$\{x \in Z^n : \xi A x \leq \lfloor \xi b \rfloor\}$$

Let

$$\tilde{A} = \xi A, \tilde{b} = \lfloor \tilde{b} \rfloor$$

we have

$$S = \{x \in \mathbb{Z}^n : \tilde{A}x \leq \tilde{b}\}$$

By the previous theorem, $\text{conv}(S)$ is a polyhedron.

We now as when is $P = P_I$

If $P = \varnothing$, then the statement holds. Elsewise, if P is bounded then

$$P = P_I \iff \text{ext}(P) \subseteq \mathbb{Z}^n$$

Theorem 3.2.3
Let $A \in Q^{m\times n}, b \in \mathbb{Q}^m$ such that

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

is nonempty and bounded.
Then $P = P_I$ if and only if for all $c \in \mathbb{Z}^n$, the LP

$$\begin{aligned} \max c^T x& \\ x \in P& \end{aligned}$$

has an integer optimal value.

Proof
($\Rightarrow$) Suppose $P = P_I$. Then $\text{ext}(P) \subseteq \mathbb{Z}^n$ and by our work in class, since P is pointed, for all $c \in \mathbb{Z}^n$, the LP has an extreme point $\bar{x}$ of P that is optimal. Since $\bar{x} \in \mathbb{Z}^n, c \in \mathbb{Z}^n$, we get that

$$c^T \bar{x} \in \mathbb{Z}^n$$

and we are done.

($\Leftarrow$) Suppose that forall $c \in \mathbb{Z}^n$, the optimal value of LP is an integer. We show that every extreme point $\bar{x}$ of P is in $\mathbb{Z}^n$.

Let $\bar{x} \in P$ be an arbitrary extreme point of P. Let $A^= x \leq b^=$ be the tight constraints at $\bar{x}$.

Since $\bar{x}$ is an extreme point of P, we have $\operatorname{rank} A^= = n$. Define

$$\bar{y}_i := \begin{cases} \xi, & (A\bar{x})_i = b_i \\ 0, & \text{else} \end{cases}$$

where $\xi \in \mathbb{Z}_{++}$ is such that $\xi A \in \mathbb{Z}^{m \times n}$.

Choose $\bar{c} := A^T \bar{y} \in \mathbb{Z}^n$. Observe that by CS Conditions, $\bar{x}$ is an optimal solution of the LP

$$\begin{aligned} \max \bar{c}^T x & \\ x \in P & \end{aligned}$$

In fact, $\bar{x}$ is the unique optimal solution for this LP. Indeed, by CS Conditions again, we see that all optimal solutions of the LP must satisfy

$$A^= x = b^=$$

But $\operatorname{rank} A^= = n$, so there is a unique solution (ie $\bar{x}$).

Since $\bar{x}$ is the unique optimal solution, there exists $r > 0$ such that

$$\bar{c}^T \bar{x} - \bar{c}^T \tilde{x} > r$$

for all other extreme points $\tilde{x} \in P$.

By finiteness, there is also $R > 0$ such that

$$|\bar{x}_j - \tilde{x}_j| < R$$

for all j and for all other extreme points $\tilde{x}$ of P.

Let $M \in \mathbb{Z}_{++}$ be some positive integer satisfying

$$M > \frac{R}{r}$$

Now, take $k \in [n]$. Define $\hat{c} \in \mathbb{Z}^n$ be

$$\hat{c}_j := \begin{cases} M\bar{c}_j + 1, & j = k \\ M\bar{c}_j, & \text{else} \end{cases}$$

Let us consider

$$\begin{aligned}
\hat{c}^T\bar{x} - \hat{c}^T\tilde{x} &= \hat{c}_j^T(\bar{x} - \tilde{x}) \\
&= \overbrace{M\underbrace{\bar{c}^T(\bar{x} - \tilde{x})}_{>r}}^{>R} + \overbrace{\bar{x}_k - \tilde{x}_k}^{>-R} \\
&> R - R \\
&= 0
\end{aligned}$$

Therefore, $\bar{x}$ is the unique optimal solution of

$$\max\{\hat{c}^T x : x \in P\}$$

Since $M\bar{c}$ and $\hat{c}$ are in $\mathbb{Z}^n$ with $\bar{x}$ the corresponding optimal solution,

$$\hat{c}^T\bar{x} = M\bar{c}^T\bar{x} + \bar{x}_k \in \mathbb{Z}$$

But $\hat{c}^T\bar{x}, M\bar{c}^T\bar{x} \in \mathbb{Z}$ and so

$$x_k \in \mathbb{Z}$$

It follows that $P = P_I$ by the arbitrary choice of k.

Theorem 3.2.4
Let $A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^m$.
Suppose $P := \{x \in \mathbb{R}^n : Ax \leq b\}$ is nonempty and bounded. Then, TFAE

(i) $P = P_I$

(ii) Every extreme point of P is in $\mathbb{Z}^n$

(iii) $\forall c \in \mathbb{R}^n$, the LP problem $\max\{c^T x : Ax \leq b\}$ has an optimal solution $\bar{x} \in \mathbb{Z}^n$.

(iv) $\forall c \in \mathbb{Z}^n$, $\max\{c^T x : Ax \leq b\} \in \mathbb{Z}$

(v) $\forall c \in \mathbb{Z}^n$, $\min\{b^T : A^T y = c, y \geq 0\} \in \mathbb{Z}$

3.3 Totally Unimodular Matrices

Definition 3.3.1 (Submatrix)
Let $I \subseteq [m], J \subseteq [n]$ then the submatrix A_{IJ} is given by

$$[a_{ij}]$$

for $i \in I, j \in J$.

Definition 3.3.2
$A \in \mathbb{Z}^{m \times n}$ is totally unimodular (TU) if for every $k \in [m]$ the determinant of every $k \times k$ submatrix of A is in $\{-1, 0, 1\}$.

Clearly, any such A has entries only in $\{-1, 0, 1\}$ ($k = 1$).

Example 3.3.1

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Example 3.3.2

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

is not TU as $\det A = -2$.

Theorem 3.3.3
Let $A \in \mathbb{Z}^{m \times n}$, $\operatorname{rank} A = m \leq n$. Then TFAE

(i) $|\det A_B| = 1$, for every basis B of A

(ii) Every extreme point of $\{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ is in $\mathbb{Z}^n$ for every $b \in \mathbb{Z}^m$

(iii) $A_B^{-1} \in \mathbb{Z}^{m \times m}$ for all bases B of A

Proof (i $\implies$ ii)
Suppose for all bases B of A, $\det A_B \in \{-1, 1\}$.

Let $b \in \mathbb{Z}^m$ be arbitrary and $\bar{x}$ is an arbitrary extreme point of

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

Then there is a basis B of A which determines $\bar{x}$. Thus

$$\bar{x}_N = 0, \bar{x}_B = A_B^{-1} b = \frac{\operatorname{adj} A_B}{\det A_B} b \in \mathbb{Z}^m$$

where adj A is the transpose of the cofactor matrix and the cofactor matrix C is such that c_{ij} is the determinant of the submatrix of A_B with i, j-th row and column removed.

Proof (ii $\implies$ iii)
Assume ii.

Let B be an arbitrary basis. We wish to choose $b := e_i + \alpha A_B 1_m \in \mathbb{Z}^m$ where

$$\alpha := \left\lceil \max_{ij} \left|(A_B^{-1})_{ij}\right| \right\rceil \in \mathbb{Z}_{++}$$

for each $i \in B$.

Consider the basic solution of $Ax = b$ determined by B

$$\begin{aligned} \bar{x}_N &= 0 \\ \bar{x}_B &= A_B^{-1} e_i + \alpha A_B^{-1} A_B 1_m \\ &= A_B^{-1} e_i + \alpha 1_m \\ &\geq 0 \end{aligned}$$

by the choice of α.

So $\bar{x}$ is feasible and so is a BFS (extreme point). By ii, $\bar{x}_B \in \mathbb{Z}^m$. But $\alpha 1_m \in \mathbb{Z}^m$, and so $A_B^{-1} e_i$ is in $\mathbb{Z}^m$ (ie the i-th column is in $\mathbb{Z}^m$ for every i).

Proof (iii $\implies$ i)
Suppose $A_B^{-1} \in \mathbb{Z}^{m \times m}$ for all bases B of A.

Note that

$$\det A_B \det A_B^{-1} = 1$$

Therefore, $\det A_B = \det A_B^{-1} \in \{-1, 1\}$ as desired.

Proposition 3.3.4
Let $A \in \{-1, 1, 0\}^{m \times n}$. TFAE

(i) A is TU

(ii) A^T is TU

(iii) $[A|I]$ is TU

(iv) $\begin{bmatrix} A \\ I \end{bmatrix}$ is TU

(v) $[A|A]$ is TU

(vi) Let D be an $m \times m$ diagonal matrix with only ± 1 on the diagonal. Then DA is TU

Remark that A is TU if and only if every matrix obtained from A via elementary row-ops is TU.

Theorem 3.3.5
Let $A \in \{-1, 0, 1\}^{m \times n}$ be TU and $b \in \mathbb{Z}^m$. Then every extreme point of

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

is in $\mathbb{Z}^n$.

Theorem 3.3.6
Let $A \in \{-1, 0, 1\}^{m \times n}$ be TU, $b \in \mathbb{Z}^m$. Then every extreme point of

$$\{x \in \mathbb{R}^n : Ax \leq b, x \geq 0\}$$

is in $\mathbb{Z}^n$.

Theorem 3.3.7
The node-arc incidence matrix of every di-graph is TU.

Proof
We argue with induction on the size of the square submatrices that every square submatrix has the desired value of the determinant. This holds for $k = 1$ trivially.

Suppose for $k \leq l$ this holds. Let us consider some $(l+1) \times (l+1)$ submatrix C.

Case I: If C has a zero column, then $\det C = 0$

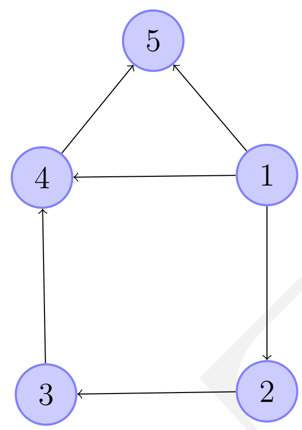
Case II: We have a column with exactly one nonzero element. We can compute $\det C$

with cofactor expansion to get

$$|\det C| = |\det ??|$$

where ?? is some $l \times l$ submatrix and thus the statement holds.

Case III: Every column has two non-zero elements. Notice that every column has a $+1, -1$ entry. So the sum of the rows is 0 and $\det C = 0$.



Example 3.3.8
The indicence matrix of the above is given by

$$\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{array} \begin{array}{c} \begin{array}{cccccc} (1,2) & (1,4) & (1,5) & (2,3) & (3,4) & (4,5) \end{array} \\ \begin{pmatrix} -1 & -1 & -1 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix} \end{array}$$

Corollary 3.3.8.1
The incidence matrix of every (undirected) bipartite graph is TU.

Proof
Let W, J be a bipartition of $G = (V, E)$.

Orient all edges of G from W to J to get a directed graph $\vec{G}$.

By our work above $\vec{G}$ is TU. Let $\vec{A}$ be the incidence matrix of $\vec{G}$.

But we can assume WLOG that $\vec{A}$ has only -1 up to index $|W|$ and 1 after. So

$$\vec{A} = DA$$

where D is a diagonal matrix with ± 1 entries on its diagonal and A the incidence matrix of G.

This completes the proof.

Notice that the above corollary is "tight" in the sense that if G is NOT bipartite, its incidence matrix is not necessarily TU.

This is because G is bipartite if and only if G does not contain an odd cycle.

3.4 König's Theorem

Definition 3.4.1 (node/vertex cover)
$C \subseteq V(G)$ such that every edge is incident with some vertex in C.

Theorem 3.4.1 (König's Theorem, 1916)
In every bipartite graph, the cardinality of a maximal matching is equal to the cardinality of a minimal vertex cover.

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

We may assume G has no isolated vertices.

Let M be a maximal matching and

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

Consider (P)

$$\begin{aligned} \max 1^T x& \\ Ax &\leq 1 \\ x &\geq 0 \\ x &\in \mathbb{Z}^E \end{aligned}$$

and consider (P')

$$\begin{aligned} \max 1^T x& \\ Ax &\leq 1 \\ x &\geq 0 \end{aligned}$$

Note opt(P) is at most opt(P').

Then the dual (D') is

$$\begin{aligned} \min 1^T y \\ A^T y &\geq 1 \\ y &\geq 0 \end{aligned}$$

Notice that (P') is always feasible and there is a trivial bound $|V(G)|/2$ on (P') so it is bounded. Thus it has an optimal solution equal to the optimal value of (D').

Then consider (D)

$$\begin{aligned} \min 1^T y \\ A^T y &\geq 1 \\ y &\geq 0 \\ y &\in \mathbb{Z}^V \end{aligned}$$

Similarly, opt(D') is at most opt(D).

The first IP formulates the maximum cardinality matching problem. Its LP relaxation has a feasible solution $\bar{x} := 0$. Moreoever, every feasible solution x of the LP has that $Ax \leq 1$ and

$$1^T Ax = 2 \cdot 1^T x \leq |V|$$

By the Fundamental Theorem of LP, it has an optimal solution with objective value equal to (D').

We claim that every optimal solution of (D') is such that $y_i \leq 1$. If not, set $y_i = 1$. This maintains feasibility but strictly improves the objective value.

The last problem is equivalent to

$$\begin{aligned} \min 1^T y \\ A^T &\geq 1 \\ y &\in \{0, 1\}^V \end{aligned}$$

which is an IP formulation of the minimum cardinality node cover problem in G.

Since G is bipartite, by corollary previous, A is TU. It follows that A^T is also TU.

By our work with TU matrices, we conclude that all extreme points of both LPs are integral. Therefore, we have equality throughout the above chain and so

$$\max_{M \text{ matching}} |M| = \min_{C \text{ vertex cover}} |C|$$

Remark that this type of proof is extremely robust and can be used to prove things such as the Maximum-Flow Minimum-Cut Theorem. Even if we do not have equality throughout, there maybe perhaps some very good approximation.

Consider (IP)

$$\begin{aligned} \max c^T x \\ Ax \leq b \\ x \geq 0 \\ x \in \mathbb{Z}^n \end{aligned}$$

$\leq$ (P)

$$\begin{aligned} \max c^T x \\ Ax \leq b \\ x \geq 0 \end{aligned}$$

$=$ (D)

$$\begin{aligned} \min b^T y \\ A^T y \geq c \\ y \geq 0 \end{aligned}$$

$\leq$ (ID)

$$\begin{aligned} \min b^T y \\ A^T y \geq c \\ y \geq 0 \\ y \in \mathbb{Z}^m \end{aligned}$$

where A is TU and $b \in \mathbb{Z}^m, c \in \mathbb{Z}^n$.

If we show that all are feasible, then we have equality throughout.

3.5 Maximum Flow

Given $\vec{G} = (V, \vec{E})$ digraph and two distinguished nodes $s, t \in E$, we are given capacities $u_{ij} \in \mathbb{Z}_+$ for all $ij \in E$.

We wish to push as much flow from s to t as possible while respecting capacities.

Let x_{ij} denote the amount of flow from node i to node j. For every $ij \in \vec{E}$.

We wish to

$$
\begin{aligned}
\max \sum_{sj \in E} s_{sj} & \\
\sum_{ij \in E} x_{ij} - \sum_{jk \in E} x_{jk} = 0 & \qquad \forall j \in V \setminus \{s, t\} \\
x_{ij} \leq u_{ij} & \qquad \forall ij \in E \\
x_{ij} \geq 0 & \qquad \forall ij \in E
\end{aligned}
$$

We should restrict to integral solutions but notice that the constraints are given by the incidence matrix and so we can apply the technique we used to prove König's Theorem.

Definition 3.5.1 (Cut)
Let $U \subseteq V$

$$\delta(U) := \{ij \in \vec{E} : i \in U, j \notin U\}$$

This is normally the directed cut but can also be defined for undirected graphs.

Definition 3.5.2 (*st*-Cut)
Let $U \subseteq V$ be such that $x \in U, t \notin U$, then $\delta(U)$ is an *st*-cut.

The integer dual gives the minimum *st*-cut problem.

Theorem 3.5.1
The maximum *st*-flow is equivalent to the minimum *st*-cut.

Proof
Exactly the same as König's.

3.5.1 Integer Programming Formulation

Let

$$c_{uv} := \begin{cases} 1, & u = s \\ 0, & \text{else} \end{cases}$$

We wish to

$$\begin{aligned} \max c^T x & \\ Mx &= 0 \\ x &\le u \\ x &\ge 0 \end{aligned}$$

where M is the incidence matrix of $\vec{G}$ (with rows corresponding to the source s, sink t removed).

Equivalently

$$\begin{aligned} \max c^T x & \\ \begin{bmatrix} M \\ -M \\ I \end{bmatrix} x &\le \begin{bmatrix} 0 \\ 0 \\ u \end{bmatrix} \\ x &\ge 0 \end{aligned}$$

and note that the constraint matrix is TU.

Similarly in the dual

$$\begin{bmatrix} M^T & -M^T & I \end{bmatrix}$$

which is TU.

The main ideas we used in proving König's Theorem apply here as well, leading to the famous Maximum-Flow Minimum-Cut Theorem.

3.5.2 Maximum-Flow Minimum-Cut

Definition 3.5.3 (*st*-Cut)
Let $W \subseteq V$ such that $s \in W, t \notin W$. Then $\delta(w)$ is an *st*-cut in $\vec{G}$.

Definition 3.5.4 (Cut Capacity)
The capacity of a *st*-cut W is

$$\sum_{ij \in \delta(W)} u_{ij}$$

Theorem 3.5.2 (Maximum-Flow Minimum-Cut)
Let $\vec{G} = (V, \vec{E})$ be a directed graph with two distinct nodes $s, t \in V$.
Also, let $u \in \mathbb{R}_+^{\vec{E}}$ be given.
The value of the maximum flow in $\vec{G}$ is equal to the capacity of the minimum *st*-cut in $\vec{G}$.
Furthermore, if $u \in \mathbb{Z}_+^{\vec{E}}$, then there is a maximal flow in $\vec{G}$ which is integral.

Proof
Same as König's Theorem.

3.6 Comments on Integer Programs via Polyhedral Theory

We saw that if A is TU with b, c integral, then under some mild assumptions, we can establish:

Primal (IP) $\leq$ (LP) relaxation $=$ (Dual LP) $\leq$ (Dual LP with integrality constraint) and equality throughout.

From this, we obtain combinatorial min-max theorems and efficient, robust primal-dual algorithms.

However, it is not always possible to find an IP formulation with coefficient matrix being TU OR to have the LP relaxation

$$\max c^T x, x \in P$$

have the property that

$$P = P_I$$

as we usually have $P_I \subseteq P$.

How can we judge IP formulations based on how closely P approximates P_I?

3.6.1 Dimension

Definition 3.6.1
$P \subseteq \mathbb{R}^n$ is a polyhedron, then

$$\dim P$$

is the number of affinely independent points in P less one!

Definition 3.6.2 (Valid Inequality)
Given $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$

$$a^T x \leq \alpha$$

is a valid inequality for P if

$$P \subseteq x \in \mathbb{R}^n : a^T x \leq \alpha$$

Definition 3.6.3 (Face)
$P_f \subseteq P$ is a face of P if

$$P_f = P \cap \{x \in \mathbb{R}^n : a^T x = \alpha\}$$

for some valid inequality

$$a^T x \leq \alpha$$

Every face of P is a polyhedron.

- $\dim \varnothing = -1$
- dim extreme points 0
- dim edges 1
- ...
- dim facets $\dim(P) - 1$

Definition 3.6.4 (Facet)
A face P_f of P is called a facet if

$$\lim P_f = \dim P - 1$$

Theorem 3.6.1
Let $P \subseteq \mathbb{R}^n, \dim P = n$ where P is a polyhedron.
Then every description of P in terms of linear inequalities contain at least one inequality for each facet and all such minimal descriptions have exactly one inequality per facet.
Moreover, minimal descriptions are unique up to scaling by positive constants.

Remark that facets of P_I help us describe the strongest valid inequalities for P_I, hence give us a tool in judging IP formulations

$$\max c^T x, x \in P$$

versus

$$\max c^T x, x \in P_I$$

While the facet description of a polyhedron $P \subseteq \mathbb{R}^n$, with $\dim P = n$ is minimal, if we are allowed to express P as a projection of another polyhedron

$$P_2 \subseteq \mathbb{R}^{n+m}$$

P_2 might have much fewer facets than P.

For example, some description in higher dimensions have less facets and can be projected down.

3.7 Special Combinatorial Optimization

We will now specialize to a class of combinatorial optimization problems.

3.7.1 Perfect Matchings in Bipartite Graphs

Definition 3.7.1 (Neighbour)
Let $G = (V, E)$ be bipartite with (W, J) a bipartition of V.
Given $S \subseteq V$, the neighbour set of S is

$$N(S) := \{u \in V : u \in S, v \notin S, uv \in E\}$$

Theorem 3.7.1 (Hall, 1939)
Let $G = (V, E)$ be bipartite, (W, J) a bipartition such that $|W| = |J| = \frac{|V|}{2}$.
G has a perfect matching if and only if for all $S \subseteq W$

$$|N(S)| \geq |S|$$

Proof
$(\neg \Leftarrow \neg)$ If there exists $S \subseteq W$ such that $|N(S)| < |S|$, then clearly there cannot be a perfect matching.

$(\neg \Rightarrow \neg)$ Suppose G does not have a perfect matching. Let M be a maximum matching and C a minimum cardinality vertex cover in G.

We have

$$\begin{aligned} |W| &> |M| \\ &= |C| && \text{König} \end{aligned}$$

Let $S := W \setminus C$. Then, since C is a vertex cover, S ONLY has neighbours in $J \cap C$.

$$\begin{aligned} |N(S)| &\leq |J \cap C| \\ &= |C| - |W \cap C| \\ &< |W| - |W \cap C| && \text{assumption} \\ &= |S| \end{aligned}$$

as desired.

Although this result is not overly useful for formulating an algorithm, it does provide a concise certificate that no perfect matching exists.

Definition 3.7.2 (Deficient Set)
A set $S \subseteq W$ such that

$$|N(S)| < |S|$$

is a deficient set.

3.7.2 Maximum Weight Perfect Matching Problem in Bipartite Graphs

Definition 3.7.3
Let the characteristic vector of a perfect matching M be

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

Consider the following (IP) formulation of Weighted Perfect Matching Problem.

$$\begin{aligned} \max \sum_{e \in E} c_e x_e \\ Ax = 1 \\ x \geq 0 \\ x \in \mathbb{Z}^E \end{aligned}$$

for $c_e \in \mathbb{R}$ and A the incidence matrix of a digraph.

By TU, we can relax the integrality condition to obtain the LP relaxation (P)

$$\begin{aligned} \max c^T x \\ Ax &= 1 \\ x &\geq 0 \end{aligned}$$

and then consider its dual (D)

$$\begin{aligned} \min 1^T y \\ A^T y &\geq c \end{aligned}$$

Let $\alpha := \max\{c_e\}$ we can obtain a trivial feasible solution for (D) with

$$\bar{y} := \alpha \frac{1}{2} 1_V$$

We will start with a dual feasible solution and maintain dual feasibility, complementary slacknesss conditions, and strive for primal feasibility.

Let $G(\bar{y}) := (V, E(\bar{y}))$ be the graph obtained from G such that

$$E(\bar{y}) := \{uv \in E : \bar{y}_u + \bar{y}_v = c_{uv}\}$$

Does $G(\bar{y})$ have a perfect matching? If so, we have an optimal $\bar{x}$ for (IP), (P). Moreoever, $\bar{y}$ is an optimal solution of (D).

If $G(\bar{y})$ does NOT have a perfect matching, by Hall's Theorem, there is

$$S \subseteq W, |N_{G(\bar{y})}(S)| < |S|$$

Now, let

$$\bar{y}_v := \begin{cases} \bar{y}_v - \epsilon, & v \in S \\ \bar{y}_v + \epsilon, & v \in N_{G(\bar{y})}(S) \\ \bar{y}_v, & \text{else} \end{cases}$$

where

$$\epsilon := \min\{y_u + y_v - c_{uv} : uv \in E, u \in S, v \notin N_{G(\bar{y})}(S)\}$$

If that set above is empty (ie the minimum is NOT well defined), S gives a certificate that that there is no perfect matching by Hall's Theorem ($N_{G(\bar{y})}(S) = N(S)$). Another way to see this is that we can let $\epsilon \to -\infty$ and still maintain dual feasibility. By Hall's Theorem, G has no perfect matching.

(P), (IP) infeasible and (D) is unbounded.

We can then terminate the algorithm and conclude there is no perfect matching.

Notice that for each iteration the objective value becomes

$$1^T \bar{y}_{\text{new}} - 1^T \bar{y} = -\epsilon(|S| - |N_{G(\bar{y})}(S)|) \leq -\epsilon \in \mathbb{Z}$$

3.8 Theorems of the Alternative

We wish to consider some of these theores in the context of Combinatorial Optimization and Integer Programming.

Theorem 3.8.1
A bipartite graph $G = (V, E)$ with bipartition (W, J), $|W| = |J|$ either

(i) has a perfect matching, or

(ii) has a deficient set

Recall the first lecture, now over $\mathbb{Q}$

Theorem 3.8.2 (Fundamental Theorem of Linear Algebra, Gauss)
Given $A \in \mathbb{Q}^{m\times n}$ and $b \in \mathbb{Q}^m$, exactly one of the following holds

(i) $\exists \bar{x} \in \mathbb{Q}^n$ such that $Ax = b$

(ii) $\exists y \in \mathbb{Q}^m$ such that $A^T y = 0, b^T y = 0$

Is there a suitable generalization to Integer Programming?

We first come up with the following:

Given $A \in \mathbb{Q}^{m\times n}$ and $b \in \mathbb{Q}^m$, exactly one of the following holds

(I) $\exists \bar{x} \in \mathbb{Z}^n$ such that $Ax = b, x \geq 0$

(II) $\exists y \in \mathbb{Q}^m$ such that $A^T y \in \mathbb{Z}^n_+, b^T y \notin \mathbb{Z}_+$

Notice this is NOT true.

Example 3.8.3
Consider $A = [2, 3], b = 1$

(I) there is $x \in \mathbb{Z}^2$ such that $2x_1 + 3x_2 = 1, x \geq 0$, NO SOLUTION

(II) there is $y \in \mathbb{Q}$ such that $2y, 3y \in \mathbb{Z}^2_+, b^T y = y \notin \mathbb{Z}_+$, NO SOLUTION

Let us lower our expectations and focus on solving systems of linear equations in integers.

Theorem 3.8.4 (Kronecker, 1800's)
Let $a \in \mathbb{Q}^n, b \in \mathbb{Q}$. Then exactly one of the following has a solution

I there is $x \in \mathbb{Z}^n$ such that $a^T x = b$

II there is $y \in \mathbb{Q}$ such that $ya \in \mathbb{Z}^n, yb \notin \mathbb{Z}$

Theorem 3.8.5 (Kronecker's Approximation Theorem, 1884)
Let $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$. Exactly one of the following holds

(I) for all $\epsilon > 0$, there is $x \in \mathbb{Z}^n$ such that $\|Ax - b\| < \epsilon$

(II) there is $y \in \mathbb{R}^m$ such that $A^T y \in \mathbb{Z}^n$ and $b^T y \notin \mathbb{Z}$

Corollary 3.8.5.1
Let $A \in \mathbb{Q}^{m\times n}, b \in \mathbb{Q}^m$. Then exactly one of the following holds

(I) there is $x \in \mathbb{Z}^n$ such that $Ax = b$

(II) there is $y \in \mathbb{Q}^m$ such that $A^T y \in \mathbb{Z}^n, b^T y \notin \mathbb{Z}$

Definition 3.8.1 (Unimodular)
$U \in \mathbb{Z}^{n\times n}$ is called unimodular if

$$|\det U| = 1$$

Definition 3.8.2 (Hermite Normal Form)
Not necessarily square matrix H such that

$$\begin{aligned} h_{ii} &> 0 && \forall i \\ 0 \leq -h_{ij} &< h_{ii} && \forall j < i \\ h_{ij} &= 0 && \forall j > i \end{aligned}$$

Proof
Given $A \in \mathbb{Z}^{m\times n}, \operatorname{rank} A = m$, we will write

$$A =: HU$$

where $U \in \mathbb{Z}^{n\times n}$ unimodular and H is in Hermite Normal Form.

We will repeatedly do one of the following

(i) swap two colums (multiply from right by permutation matrix)

(ii) multiply any column by (-1) (multiplication from left by identity slightly changed)

(iii) replace a column j by columns $j + p\cdot$(column i) for some $p \in \mathbb{Z}$ (multiplication from right by $I + p$ on one entry)

Moreover, we have $H \in \mathbb{Z}^{m\times n}$

$$A = HU$$

so

$$Ax = b \iff HUx = b$$

For every unimodular matrix $U \in \mathbb{Z}^{n\times n}$

$$U(\mathbb{Z}^n) = \mathbb{Z}^n$$

which tells us that $HUx = b$ has an integral solution if and only if

$$Hz = b$$

has a solution $z \in \mathbb{Z}^n$.

In general, for IPs, how do we construct theorems of the alternative?

$$S := \{x \in \mathbb{Z}^n : Ax = b, x \geq 0\}$$

We will deal conv S instead.

We will try to represent conv S in terms of contraints obtained from

$$\begin{aligned} Ax &= b \\ x &\geq 0 \\ x &\in \mathbb{Z}^n \end{aligned}$$

Then, we can express theorems of the alternative in terms of conv S.

4 Nonlinear Optimization (Continuous Optimization)

Nonlinear Optimization Problems are at least as hard as (IP)s. Indeed

$$\begin{aligned} x \in \{0,1\}^n &\iff x_j(1-x_j) = 0 && \forall j \in [n] \\ x \in \mathbb{Z}^n &\iff \sin(\pi x_j) = 0 && \forall j \in [n] \end{aligned}$$

Because there is so much freedom with formulating problems, it is incredibly important to study the structure of problems and formulate problems such that we gain additional information about the problem.

4.1 Definitions & Basic Results

Definition 4.1.1 (Open Euclidean Ball)
$B(\bar{x}, \delta) := \{x \in \mathbb{R}^n : \|x - \bar{x}\|_2 < \delta\}$ for $\delta > 0$.

Definition 4.1.2 (Interior)
For $S \subseteq \mathbb{R}^n$ the interior of S is

$$\operatorname{int} S := \{x \in S : B(x, \delta_x) \subseteq S\}$$

Definition 4.1.3 (Sequence)
$\{x^{(k)}\}$ is a sequence of points in $\mathbb{R}^n$

Definition 4.1.4 (Closure)
Given $S \subseteq \mathbb{R}^n$, the closure of S is

$$\operatorname{cl} S := \{\bar{x} \in \mathbb{R}^n : x^{(k)} \to \bar{x}, \{x^{(k)}\} \subseteq S\}$$

Definition 4.1.5 (Compact)
$S \subseteq \mathbb{R}^n$ is called compact if S is closed and bounded.

Remark that there are different notions of compactness in other spaces.

Theorem 4.1.1 (Bolzano-Weierstrass)
Let $S\{\mathbb{R}\}^n$ be compact.
Then every sequence $\{x^{(k)}\} \subseteq S$ has a convergent subsequence, say $\{x^{(l)}\}$ such that

$$x^{(l)} \to \bar{x} \in S$$

Definition 4.1.6 (Continuity)
Let $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}^m$.
We say f is continuous at $\bar{x} \in S$ if for every sequence $(x_k) \subseteq S$ such that

$$x_k \to \bar{x} \implies f(x_k) \to f(\bar{x})$$

Theorem 4.1.2
Let $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}^m$. Then TFAE

(i) f is continuous on S

(ii) $\forall \bar{x} \in S, \forall \epsilon > 0, \exists \delta, \|x - \bar{x}\| < \delta \implies \|f(x) - f(\bar{x})\| < \epsilon$

(iii) for every open set $U \subseteq \mathbb{R}^m$, $f^{-1}(U) \subseteq S$ is open in S

(iv) for every closed set $F \subseteq \mathbb{R}^m$, $f^{-1}(F)$ is closed in S

Definition 4.1.7 (Level Sets)
Level sets (sublevel sets) is

$$\text{Level}_\alpha(f) := \{x \in S : f(x) \leq \alpha\}$$

Notice that if $f : S \to \mathbb{R}$ is continuous with S being closed, then all Level Sets are closed.

Definition 4.1.8 (Infimum)
Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$.
The infimum of f over S is the largest $\alpha \in [\pm\infty]$ such that

$$\forall x \in S, f(x) \geq \alpha$$

Definition 4.1.9 (Supremum)
of f over S is the smallest $\beta \in [\pm\infty]$ such that

$$\forall x \in S, f(x) \leq \beta$$

Example 4.1.3
$\inf\{\frac{1}{x} : x \in \mathbb{R}_{++}\} = 0$ (not attained)

$\sup\{\frac{1}{x} : x \in \mathbb{R}_{++}\} = \infty$ (not attained)

By convention, we will define

$$\inf \varnothing = \infty, \sup \varnothing = -\infty$$

as they are consistent with the primal-dual relations we saw before!

Theorem 4.1.4 (Weierstrass)
Let $S \subseteq \mathbb{R}^n$ be a non-empty compact set.
Let $f : S \to \mathbb{R}$ be continuous on S. Then f attains both its infimum and supremum.

Definition 4.1.10 (Coercive)
Let $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$.
Then f is called coercive on S if the level sets

$$\{x \in S : f(x) \leq \alpha\}$$

of f are bounde for every $\alpha \in \mathbb{R}$.

Notice that f is coercive if and only if every sequence $(x_k) \subseteq S$ such that

$$\|x_k\| \to \infty \implies f(x_k) \to \infty$$

Theorem 4.1.5
Let $S \subseteq \mathbb{R}^n$ be a non-empty closed set and $f : S \to \mathbb{R}$ is continuous and coercive over S.
Then f attains its infimum over S.

Definition 4.1.11 (Symmetric)
Let $A \in \mathbb{R}^{n\times n}$ then A is symmetric if $A = A^T$.

Proposition 4.1.6
If A is symmetric, then there is $U \in \mathbb{R}^{n\times n}$ orthogonal and $A = UDU^T$ where

$$U := \{u_1, \ldots, u_n\}$$

eigenvectors of A.
Furthermore, D is diagonal consisting of the eigenvalues of A.

Example 4.1.7
Consider a quadratic fnction

$$f(x) = \delta + c^T x + x^T A x = \delta + c^T x + \sum_i \sum_j a_{ij} x_i x_j$$

for $\delta \in \mathbb{R}, c \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n}$

Note that we can assume A is symmetric as

$$x^T A x = x^T \left(\frac{A + A^T}{2} \right) x$$

Definition 4.1.12 (Positive-Semidefinite)
A symmetric matrix $A \in \mathbb{R}^{n \times n}$ is positive semidefinite if all eigenvalues are non-negative.

Note that an equivalent definition is that

$$h^T A h \geq 0$$

for every $h \in \mathbb{R}^n$.

Definition 4.1.13 (Positive-Definite)
Positive-semidefinite matrices are positive-definite if every eigenvalue is positive.

Note that we can require

$$h^T A h > 0$$

for every $h \in \mathbb{R}^m, h \neq 0$.

Proposition 4.1.8 (Rayleigh Quotient)
We have formulas for the largest and smallest eigenvalues of positive semi-definite matrices

$$\lambda_1(A) = \max_{h \neq 0} \frac{h^T A h}{h^T h}, \lambda_n(A) = \min_{h \neq 0} \frac{h^T A h}{h^T h}$$

Proposition 4.1.9
If A is positive definite, then

$$f(x) = \delta + c^T x + x^T A x$$

is coersive.

Proof
$f(x) = \delta + c^T x + x^T Ax \geq \delta + c^T x + \lambda_n(A)\|x\|_2^2$ given by the Rayleigh quotient.

Furthermore
$$f(x) \geq \delta - \|c\| \cdot \|x\| + \lambda_n(A)\|x\|_2^2$$
by the cauchy-schwartz inequality.

This shows that $\|x\| \to \infty \implies f(x) \to \infty$.

4.2 Convexity

Definition 4.2.1 (Convex)
Let $S \subseteq \mathbb{R}^n$ be a convex set.
$f : S \to \mathbb{R}$ is convex on S if

$$\forall u, v \in S, \forall \lambda \in [0, 1], f(\lambda u + (1 - \lambda)v) \leq \lambda f(u) + (1 - \lambda)f(v)$$

Definition 4.2.2 (Strictly Convex)
f above is strictly convex if for $\lambda \in (0, 1)$ and $u \neq v \in S$

$$f(\lambda u + (1 - \lambda)v) < \lambda f(u) + (1 - \lambda)f(v)$$

Definition 4.2.3 (Differentiable)
Let $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}, \bar{x} \in \text{int}(S)$.
Then if there is $\nabla f(\bar{x}) \in \mathbb{R}^n$ such that

$$\lim_{x \to \bar{x}} \frac{f(x) - [f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})]}{\|x - \bar{x}\|} = 0$$

then we say f is differentiable at $\bar{x}$, and $\nabla f(\bar{x})$ is the derivative of f at $\bar{x}$.

Proposition 4.2.1
If $\frac{\partial f}{\partial x}$ are continuous in an open neighbourhood of $\bar{x}$, then

$$\nabla f(\bar{x}) = \left[\frac{\partial f}{\partial x_i}(\bar{x})\right]$$

and f is continuously differentiable.

The vector above is also called the gradient of f. Note that the derivative is really a linear map while the gradient is an actual vector.

There is a simple but very useful correspondance between sets and functions (Analysis $\iff$ Geometry).

Definition 4.2.4 (Epi-graph)
For $f : \mathbb{R}^n \to \mathbb{R}$

$$\operatorname{epi}(f) := \{(\mu, x) \in \mathbb{R} \times \mathbb{R}^n : f(x) \leq \mu\}$$

Theorem 4.2.2
$f : \mathbb{R}^n \to \mathbb{R}$ is a convex function if and only if $\operatorname{epi}(f)$ is a convex set in $\mathbb{R}^{n+1}$.

Theorem 4.2.3
Let $S \subseteq \mathbb{R}^n$ be a convex set. Let $f : S \to \mathbb{R}$ be continuously differentiable on S. Then f is convex if and only if

$$f(v) \geq f(u) + \nabla f(u)^T (v - u)$$

for all $u, v \in S$.

Proof
($\Rightarrow$) Suppose f is convex and $u, v \in S$. Let $\lambda \in (0, 1]$

$$\begin{aligned}
\lambda f(u) + (1 - \lambda) f(v) &\geq f(\lambda u + (1 - \lambda) v) \\
f(v) + \lambda[f(u) - f(v)] &\geq f(v + \lambda(u - v)) \\
\frac{f(u) - f(u + \lambda(u - v))}{\lambda} &\geq f(v) - f(u)
\end{aligned}$$

We can take the limit as $\lambda \to 0^+$ and get that

$$-\nabla f(v)^T (u - v) \geq f(v) - f(u)$$

We used $g(\lambda) := f(v + \lambda(u - v)), g : [0, 1] \to \mathbb{R}$

$$g'(0) = \nabla f(v)^T (u - v)$$

($\Leftarrow$) Suppose for all $u, v \in S$

$$f(v) \geq f(u) + \nabla f(u)^T (v - u)$$

Let $u, v \in S$ be arbitrary and $\lambda \in [0,1]$. Let $z := \lambda u = (1-\lambda)v$. Then

$$\begin{aligned} f(u) &\geq f(z) + \nabla f(z)^T(u-z) \\ f(v) &\geq f(z) + \nabla f(z)^T(v-a) \\ \lambda f(u) + (1-\lambda)f(v) &\geq f(z) + \nabla v(z)^T[\lambda u + (1-\lambda)v - z] \\ &= f(z) \end{aligned}$$

Definition 4.2.5 (Differentiable)
Let $S \subseteq \mathbb{R}^n, f : S \to \mathbb{R}^m$. Also, let $\bar{x} \in \text{int}(S)$.
Then if there is a linear transformation $Df(\bar{x}) : \mathbb{R}^n \to \mathbb{R}^m$ such that

$$\lim_{x \to \bar{x}} \frac{\|f(x) - [f(\bar{x}) + Df(\bar{x})(x-\bar{x})]\|}{\|x - \bar{x}\|} = 0$$

then f is differentiable at $\bar{x}$ and $Df(\bar{x})$ is the derivative of f at $\bar{x}$.

Since $Df(\bar{x})$ is a linear transformation from $\mathbb{R}^n \to \mathbb{R}^m$, it has a matrix representation by some $A \in \mathbb{R}^{m \times n}$

$$Df(\bar{x})(x-\bar{x}) = A(x-\bar{x})$$

If we write

$$f(x) = [f_i(x)]$$

$f_i :\to \mathbb{R}$ for every i.

If f is continuously differentiable at $\bar{x}$, then the ij-th entry of A is

$$\frac{\partial f_i}{\partial x_j}$$

Definition 4.2.6 (Jacobian Matrix)
The underlying matrix is called the Jacobian matrix when $m = n$.

Definition 4.2.7 (Jacobian Determinant)
We are also interested in the $\det A$ called the Jacobian Determinant.

Definition 4.2.8 (Contour)
$f : \mathbb{R}^n \to \mathbb{R}, \alpha \in \mathbb{R}$
We define

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

as the contour of f.

Note that in some literature, we let level set be the contour set and sublevel set be the level set as we have defined them.

Example 4.2.4
$f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = x_1^2 + \frac{1}{16}x_2^2$$

The contour with $\alpha = 1$ is an ellipsoid.

Proposition 4.2.5
If $f : \mathbb{R}^n \to \mathbb{R}$ is twice continuously differentiable, then

$$\frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j} = \frac{\partial^2 f(\bar{x})}{\partial x_j \partial x_i}$$

for every i, j.

Definition 4.2.9 (Hessian Matrix)
With the same assumptions above

$$Hf(\bar{x}) = \left[\frac{\partial^2 f}{\partial x_i \partial x_j}\right]$$

where $i, j \in [n]$.

Notice that since f is twice continuously differentiable, the Hessian at $\bar{x}$ is a symmetric matrix.

Example 4.2.6
$f : \mathbb{R}^2 \to \mathbb{R}$

$$f(x) = x_1^2 + \frac{1}{16}x_2^2$$

then

$$Hf(x) = \begin{bmatrix} 2 & 0 \\ 0 & \frac{1}{8} \end{bmatrix}$$

is positive-definite for every $x \in \mathbb{R}^2$.

Theorem 4.2.7
Let $S \subseteq \mathbb{R}^n$ be a convex set. Suppose $f :\to \mathbb{R}$ is twice continuously differentiable. Then f is convex on S if and only if $Hf(x)$ is positive semi-definite for every $x \in S$.

Definition 4.2.10 (Strict Convexity)
$f : S \subseteq \mathbb{R}^n \to \mathbb{R}$ is strictly convex if f is convex and for every distinct pair $u, v \in S$ and for all $\lambda \in (0, 1)$

$$\lambda(u) + (1 - \lambda)f(v) < f(\lambda u + (1 - \lambda)v)$$

Theorem 4.2.8
Let $S \subseteq \mathbb{R}^n$ be a convex set. Suppose $f :\to \mathbb{R}$ is twice continuously differentiable. Then f is convex on S if $Hf(x)$ is positive semi-definite for every $x \in S$.

Notice that converse is false. Consider $f : \mathbb{R} \to \mathbb{R}$ given by

$$f(x) := x^4$$

which is clearly strictly convex.

But

$$f^{(2)}(x) = 12x^2, f^{(2)}(0) = 0$$

Theorem 4.2.9
Let $S \subseteq \mathbb{R}^n$ be convex with $f : S \to \mathbb{R}$ continuously differentiable on S. Then f is convex on S if and only if

$$[\nabla f(u) - \nabla f(v)]^T (u - v) \geq 0$$

for all $u, v \in S$.

In continuous optimization, we almost always pick a direction $d \in \mathbb{R}^n$, to iteratively "move along" to improve our current point $\bar{x}$.

Definition 4.2.11 (Search Direction)
d is referred to as the search direction

The new point is $\bar{x} + \alpha d$

Definition 4.2.12 (Step Size)
where $\alpha > 0$ is the "step size".

We wish to analyze the behavior of f.

$$\begin{aligned} g(\alpha) &:= f(\bar{x} + \alpha d) && \bar{x}, d \\ g'(\alpha) &= [\nabla f(\bar{x} + \alpha d)]^T d \\ g'(0) &= [\nabla f(\bar{x})]^T d \\ g''(\alpha) &= d^T [Hf(\bar{x} + \alpha d)] d \\ g''(0) &= d^T [Hf(\bar{x})] d \end{aligned}$$

4.3 Steepest Descent & Newton's Method

Assume we are at $\bar{x} \in \mathbb{R}^n$, we perform the update

$$\bar{x}' = \bar{x} + \alpha d$$

4.3.1 Steepest Descent

If we choose

$$d = \nabla f(\bar{x})$$

the method is called the Steepest Descent.

The line search comes at the fact that we must linearly search to not "overshoot".

4.3.2 Newton's Method

If we choose

$$d = [Hf(\bar{x})]^{-1} \nabla f(\bar{x})$$

the method is called Newton's Method and is asymptotically better but requires computation of the Hessian and inverting a matrix. This is too costly in practice.

Theorem 4.3.1
Let $S \subseteq \mathbb{R}^n$ be a non-empty, convex set and $f : S \to \mathbb{R}$ be convex.
Suppose f is of class C^1 at $\bar{x} \in S$. Then $\bar{x}$ is a minimizer of f over S if and only if

$$[\nabla f(\bar{x})]^T (x - \bar{x}) \geq 0$$

for all $x \in S$.

4.4 Separating & Supporting Hyperplane Theorems

Example 4.4.1
Consider $S \subseteq \mathbb{R}^n$, non-empty, convex. Let

$$f(x) := \|x - u\|_2^2$$

where $u \in \mathbb{R}^n$ fixed.

Since f is convex

$$Hf(x) = 2I$$

for all $x \in \mathbb{R}^n$.

Furthermore, $f(x)$ is continously differentiable at every $x \in \mathbb{R}^n$.

We wish to minimize $f(x)$ over S. Let us apply the previous theorem.

Then $\bar{x}$ is a minimizer if and only if

$$[\nabla f(\bar{x})]^T (x - \bar{x}) = 2(x - u)^T (x - \bar{x}) \geq 0$$

for all $x \in S$.

Corollary 4.4.1.1 (Kolmogorov Criterion)
Let $\varnothing \neq S \subseteq \mathbb{R}^n$ be closed and convex. Then there exists a closest point

$$\bar{x}(u)$$

in S to u, which is unique and satisfies

$$[u - \bar{x}(u)]^T [x - \bar{x}(u)] \leq 0$$

for all $x \in S$.

Proof
To show existence and uniqueness, note that

$$\|x - u\|_2^2$$

is strictly convex and coercive.

Theorem 4.4.2 (Separating Hyperplane)
Let $S \subseteq \mathbb{R}^n$ be a non-empty closed convex set. Then for every $u \in \mathbb{R}^n \setminus S$, there is $a \in \mathbb{R}^n \setminus \{0\}, \alpha \in \mathbb{R}$ such that

$$a^T x \leq \alpha$$

for all $x \in S$ and

$$a^T u > \alpha$$

In other words,

$$a^T x = \alpha$$

separates u, S.

Proof
Let S, u be as above. Then there is a closest $\bar{x}(u) \in S$ to u.

Moreoever, we know $u \notin S$ so

$$\|\bar{x}(u) - u\|_2 > 0$$

By the Kolmogorov Criterion

$$[u - \bar{x}(u)]^T [x - \bar{x}(u)] \leq 0$$

for every $x \in S$.

Expand

$$\underbrace{[u - \bar{x}(u)]}_{=:a}^T x \leq \overbrace{\bar{x}(u)^T (u - \bar{x}(u))}^{=:\alpha}$$

We need only show the second case.

$$\begin{aligned} a^T u - \alpha &= [u - \bar{x}(u)]^T u - \bar{x}(u)^T (u - \bar{x}(u)) \\ &= [u - \bar{x}(u)]^T [u - \bar{x}(u)] \\ &= \|u - \bar{x}(u)\|_2^2 \\ &> 0 \end{aligned}$$

Note that we can use this theorem to get another proof for the Farkas' Lemma and Strong Duality follows. We can generalize this to get Strong Duality for general NLP.

Recall Farkas' Lemma

Lemma 4.4.3
For every $A \in \mathbb{R}^{m \times n}, b \in \mathbb{R}^m$ exactly one of the following holds

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

(II) $A^T y \geq 0, b^T y < 0$

In geometric terms

(I) $b \in \text{cone}\{A_j\}$ remark that $\text{cone}\{A_j\}$ is a polyhedral cone and therefore is a closed convex set

(II) $b \notin \text{cone}\{A_j\}$ so there is a separating hyperplane $a \in \mathbb{R}^n, \alpha \in \mathbb{R}$ such that $a^T b > \alpha$

In the second case, we can take

$$\begin{aligned} y &:= -a \\ y^T a_j &\geq -\alpha && \forall j \\ y^T b &< -\alpha \end{aligned}$$

Definition 4.4.1 (Supporting Hyperplane)
$S \cap \{x \in \mathbb{R}^n : a^T x = \alpha\} \neq \varnothing$
Then the RS set is a supporting hyperplane for S.

Note that the hyperplane we found in the proof is a supporting hyperplane for S as $\bar{x}(u)$ is in the intersection and

$$S \subseteq \{x \in \mathbb{R}^n : a^T x \leq \alpha\}$$

4.5 Lagragians & Lagrangian Duality

Let $S \subseteq \mathbb{R}^n, f, g_1, \ldots, g_n : S \to \mathbb{R}$ be given where S is a "simple" set and consider (P)

$$\begin{aligned} \inf f(x) \\ g_i(x) &\leq 0 && \forall i \in [m] \\ x &\in S \end{aligned}$$

Definition 4.5.1 (Lagrangian)
$L : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ such that

$$L(x, \lambda) := f(x) + \lambda^T g(x)$$

Definition 4.5.2 (Lagrange Multipliers)
$\lambda_i, \dots, \lambda_m$ as above.

Definition 4.5.3 (Lagrangian Dual)
The lagrangian dual of (P) is (D)

$$\begin{aligned} \sup h(\lambda) \\ \lambda \geq 0 \end{aligned}$$

where

$$h(\lambda) := \inf_{x \in S}\{L(x, \lambda)\}$$

Theorem 4.5.1
Let $S \subseteq \mathbb{R}^n, \bar{x} \in S$.
Then $\bar{x}$ is an optimal solution of (P) if there is $\bar{\lambda} \in \mathbb{R}^m$ such that all the following hold:

(i) $g(\bar{x}) \leq 0$ (primal feasibility)

(ii) $\bar{\lambda} \geq 0$ and $f(\bar{x}) + \bar{\lambda}^T = \inf_{x \in S}\{L(x, \bar{\lambda})\}$ (dual feasibility)

(iii) $\bar{\lambda}_i \cdot g_i(\bar{x}) = 0$ for all $i \in [m]$ (complementary slackness)

Before the proof, let us note the connection to Weak Duality.

$$\sup_{\lambda \geq 0} \inf_{x \in S}\{L(x, \lambda)\} \leq \inf_{x \in S} \sup_{\lambda \geq 0}\{L(x, \lambda)\}$$

which holds as any feasible $\bar{x}, \lambda$ is a solution for each system.

So the theorem above is NOT analogous to the Strong Duality Theorem, but is more like a corollary of the Weak Duality Theorem.

Proof
Let

$$\begin{aligned} \bar{x} &\in S \\ g(\bar{x}) &\leq 0 \\ \bar{\lambda} &\geq 0 \\ f(\bar{x}) + \bar{\lambda}^T g(\bar{x}) &= \inf_{x \in S}\{L(x, \bar{\lambda})\} \\ \bar{\lambda}^T g(\bar{x}) &= 0 \end{aligned}$$

Let x be any arbitrary feasible solution of (P).

$$\begin{aligned} f(x) &\geq f(x) + \underbrace{\bar{\lambda}}_{\geq 0} \overbrace{g(x)}^{\leq 0} \\ &\geq f(\bar{x}) + \bar{\lambda} g(\bar{x}) && \text{dual feasibility} \\ &= f(\bar{x}) && \text{complementary slackness} \end{aligned}$$

Now, let us consider the special case when $S = \mathbb{R}^n$ and

$$f, g_1, \ldots, g_m$$

are convex functions.

Definition 4.5.4 (Slater Point)
$\hat{x} \in \mathbb{R}^n$ is a Slater point for (P)

$$\inf\{f(x) : x \in \mathbb{R}^n, g(x) \leq 0\}$$

if $g(\hat{x}) < 0$.

Theorem 4.5.2
Let $S = \mathbb{R}^n$ and

$$f, g_1, \ldots, g_m : S \to \mathbb{R}$$

convex.
Suppose (P) has a Slater point, then a feasible solution $\bar{x}$ of (P) is optimal if and only if there exist $\bar{\lambda} \geq 0$ such that

(i) $L(\bar{x}, \lambda) \leq L(\bar{x}, \bar{\lambda}) \leq L(x, \bar{\lambda})$ for all $x \in \mathbb{R}^n, \lambda \geq 0$ (saddle point condition)

(ii) $\bar{\lambda}^T g(\bar{x}) = 0$ (complementary slackness)

Theorem 4.5.3 (Karush 1931, Kuhn-Tucker 1950s)
Consider the same (P) as above. Suppose (P) has a Slater point.
Assume $\bar{x} \in \mathbb{R}^n$ satisfies $g(\bar{x}) \leq 0$ and

$$f, g_i, i \in J(\bar{x}) := \{i : g_i(\bar{x}) = 0\}$$

are differentiable at $\bar{x}$.
Then $\bar{x}$ is optimal in (P) if and only if

$$-\nabla f(\bar{x}) \in \operatorname{cone}\{\nabla g_i(\bar{x}) : i \in J(\bar{x})\}$$

Can we generalize this theorem when f, g_i are NOT convex?

Consider (P)

$$\begin{aligned} \inf f(x) & \\ g_i(x) \leq 0 & \qquad\qquad i \in [m] \\ h_j(x) = 0 & \qquad\qquad j \in [p] \\ x \in S & \end{aligned}$$

where $f, g_i, h_j : \mathbb{R}^n \to \mathbb{R}, S \subseteq \mathbb{R}^n$ is a "simple" set.

Theorem 4.5.4 (Mangasarian & Fromonitz Constraint Qualification)
Feasible $\bar{x}$ is optimal in (P) if and only if

(i) $\bar{x} \in \text{int}(S), g(\bar{x}) \leq 0, h(\bar{x}) = 0$

(ii) f, g_i, h_j are continuous on S and $f, g_i, i \in J(\bar{x})$ and h_j are differentiable at $\bar{x}$

(iii) $\{\nabla h_i(\bar{x}) : i \in [p]\}$ is linearly independent

(iv)

$$\{d \in \mathbb{R}^n : \nabla g_i(\bar{x})^T d < 0, i \in J(\bar{x}), \nabla h_j(\bar{x})^T d = 0, j \in [p]\} \neq \varnothing$$

We will write (MFCQ) for short.

Consider the special case $S = \mathbb{R}^n$, g_i are convex, and h_j are affine.

If such a convex optimization problem has a Slater point $\hat{x}$, then

$$0 \underbrace{>}_{\text{Slater point}} g_i(\hat{x}) \underbrace{\geq}_{\text{convexity}} \overbrace{g_i(\bar{x})}^{=0} + \nabla g_i(\bar{x})^T (\hat{x} - \bar{x}) = \nabla f_i(\bar{x})^T d$$

for all $i \in J(\bar{x})$.

This gives us that (iv) of MFCQ holds at $\bar{x}$.

Consider another condition (iii') which says

$$\{\nabla h_j(\bar{x}) : j \in [p]\} \cup \{\nabla g_i(\bar{x}) : i \in J(\bar{x})\}$$

are linearly independent and notice that

$$(iii') \implies (iii), (iv)$$