# CO 250 Introduction to Optimization

Keven Qiu
taught by Peter Nelson

# Contents

# Chapter 1

# Introduction

## 1.1 Linear Programs

**Def Affine Function**: A function $f : \mathbb{R}^n \to \mathbb{R}$ is an affine function if $f(x) = a^T x + \beta$, where $x$ and $a$ are vectors with $n$ entries and $\beta$ is a real number. If $\beta = 0$, then $f$ is a linear function. Thus, every linear function is affine, but the converse is not true.

**Def Linear Program (LP)**: A linear program is the problem of maximizing or minimizing an affine function

$$f(x) = f(x_1, \ldots, x_n)$$

subject to a finite number of linear constraints of the form $g(x) \leq c$, $g(x) \geq c$, or $g(x) = c$.

- the $x_i$ are real variables.
- the function $f$ is the objective function
- the expressions $g(x) \leq c$, $g(x) \geq c$, and $g(x) = c$ are constraints

Ex.

$$\begin{array}{rr} \min & 3x_1 + 2x_2 - x_3 + 5 \\ \text{s.t.} & x_1 + x_3 \leq 7 \\ & 2x_1 \geq 10 \\ & x_1 + x_3 = 5 \\ & x_1, x_2 \geq 0 \end{array}$$

Non Ex.

$$\begin{array}{rr} \min & x_1 + x_2 \\ \text{s.t.} & x_1 + 2x_2 = 7 \\ & x_1 - x_2 < 0 \end{array}$$

Ex. Employees at a company have salary demands

- Tom wants $\geq$ \$20000.
- Peter, Nina, Samir each want $\geq$ \$5000 more than Tom.
- Gary wants at least that of Tom and Peter combined.
- Linda wants $\geq$ \$200 more than Gary.
- Combined salary of Nina and Samir should be $\geq$ twice combined salary of Tom and Peter.
- Bob is at least as high as Peter and Samir.
- Bob and Peter should earn $\geq$ \$60000 combined.

LP that will determine salaries minimizing total salary expenses.
**Variables**: $x_T, x_P, x_N, x_S, x_G, x_L, x_B$ be salaries of people.

$$\begin{aligned} \min \quad & x_T + x_P + x_N + x_S + x_G + x_L + x_B \\ \text{s.t.} \quad & x_T \geq 20000 \\ & x_P \geq x_T + 5000 \\ & x_N \geq x_T + 5000 \\ & x_S \geq x_T + 5000 \\ & x_G \geq x_T + x_P \\ & x_L \geq x_G + 200 \\ & x_N + x_S \geq 2(x_T + x_P) \\ & x_B \geq x_P \\ & x_B \geq x_S \\ & x_B + x_P \geq 60000 \\ & x_T, x_P, x_N, x_S, x_G, x_L, x_B \geq 0 \end{aligned}$$

Say we want to minimize the highest paid employee/salary, we can use another variable $y$ and set constraints of $y$ to $y \geq x_1, \ldots, x_n$, then the objective function is $\min y$.

**Techniques for linear program constraints** Suppose we had the linear program $\max\{c^T x : a_1^T x \leq b_1, a_1^T x \leq b_2, x \geq 0\}$, then if we want at least one of two constraints to hold, i.e. or, we can let $M > 0$ be a big enough constant and the new LP is

$$\begin{aligned} \max \quad & c^T x \\ \text{s.t.} \quad & a_1^T x \leq b_1 + My \\ & a_2^T x \leq b_2 + M(1-y) \\ & x \geq 0 \\ & y \in \{0, 1\} \end{aligned}$$

If the inequality is $\geq$, then we subtract: $b_1 - My$ and $b_2 - M(1-y)$.

## 1.2 Integer Programs

**Def Integer Program (IP)**: An integer program is a linear program together with constraints specifying that some particular non-empty subset of the variables are integers.

A mixed integer program (MIP) is an IP where not all variables are constrained to be integers.
A pure integer program is one where all variables have integer constraints.

Ex.

$$\begin{aligned} \min \quad & 2.1x_1 + x_2 - 3x_3 + 4 \\ \text{s.t.} \quad & x_1 + 2.5x_2 \leq 7.3 \\ & x_1 - x_3 \geq 4 \\ & x_1, x_2, x_3 \geq 0 \\ & x_1, x_2 \text{ integers} \end{aligned}$$

Ex. For a desert trek, $3.6m^3$ of cargo space is available. There are 7 items available each having a value and a volume: Given that only one of each item is available and items cannot be split up,

| Item | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|
| Volume($m^3$) | 0.55 | 0.6 | 0.7 | 0.75 | 0.8 | 0.9 | 0.95 |
| Value ($) | 250 | 300 | 500 | 700 | 750 | 900 | 950 |

formulate an IP for the maximize the total value of goods.
For item $i$, let $x_i$ be a variable where $x_i = 1$ if we take item $i$, $x_i = 0$ otherwise.

$$\begin{aligned} \max \quad & 250x_1 + 300x_2 + 500x_3 + 700x_4 + 750x_5 + 900x_6 + 950x_7 \\ \text{s.t.} \quad & 0.55x_1 + 0.6x_2 + 0.7x_3 + 0.75x_4 + 0.8x_5 + 0.9x_6 + 0.95x_7 \leq 3.6 \\ & x_1, \ldots, x_7 \geq 0 \\ & x_1, \ldots, x_7 \leq 1 \\ & x_1, \ldots, x_7 \text{ integers} \end{aligned}$$

The last 3 constraints can be summarized to $x_1, \ldots, x_7 \in \{0, 1\}$. Alternatively, let $v_i$ be the volume of item $i$, and $n_i$ be the value.

$$\begin{aligned} \max \quad & \sum_{i=1}^{7} n_i x_i \\ \text{s.t.} \quad & \sum_{i=1}^{7} v_i x_i \leq 3.6 \\ & x_1, \ldots, x_7 \in \{0, 1\} \end{aligned}$$

Ex. The blocks represent resources. Say $v_i$ is the value of resource $i$ when mined. Suppose that time available only allows 7 blocks to be mined, and each block must be either completely mined or not. Suppose also that, to mine a block, the blocks above it must be mined first. Ex. to mine


block 12, blocks 3,4,5,8,9 must be mined. Formulate an IP to maximize the total value.
Let $x_i$ be a variable indicating whether $i$ is mined or not.

$$
\begin{aligned}
\max \quad & \sum_{i=1}^{12} v_i x_i \\
\text{s.t.} \quad & \sum_{i=1}^{12} x_i \leq 7 \\
& x_1 + x_2 - 2x_6 \geq 0 \\
& x_2 + x_3 - 2x_7 \geq 0 \\
& x_3 + x_4 - 2x_8 \geq 0 \\
& x_4 + x_5 - 2x_9 \geq 0 \\
& x_6 + x_7 - 2x_{10} \geq 0 \\
& x_7 + x_8 - 2x_{11} \geq 0 \\
& x_8 + x_9 - 2x_{12} \geq 0 \\
& x_1, \dots, x_{12} \in \{0, 1\}
\end{aligned}
$$

This works because the constraint $x_i + x_j - 2x_k \geq 0$ where $x_i, x_j, x_k \in \{0,1\}$ states that either $x_k = 0$ or $x_i = x_j = x_k = 1$. This translates to the statement that either block $k$ is not mined, or blocks $i$ and $j$ are also mined. This captures the rule that each block can only be mined if those above it are mined too.

Remark: constraint $x \in \{0, 2\} \implies x - 2z = 0,\ z \in \mathbb{Z}, z \geq 0, z \leq 1$.

**Techniques for integer program constraints**

- If $A$ then $B$ is $x_b \geq x_a$
- If $A$ then not $B$ is $1 - x_b \geq x_a$
- If not $A$ then $B$ is $x_b \geq 1 - x_a$
- $A$ if and only if $B$ is $x_a = x_b$
- If $A$ then $B$ and $C$ is $x_b \geq x_a, x_b \geq x_c$
- If $A$ then $B$ or $C$ is $x_b + x_c \geq x_a$
- If $B$ or $C$ then $A$ is $x_a \geq x_b, x_a \geq x_c$
- If $B$ and $C$ then $A$ is $x_a \geq x_b + x_c - 1$
- If $m$ or more of $n$ items $x_2, \dots, x_{n+1}$ then $A$ is $\frac{x_2 + x_3 + \cdots + x_{n+1} - m + 1}{n - m + 1}$

## 1.3 Optimization Problems on Graphs

**Def Graph**: A graph is a pair $G = (V, E)$ where $V$ is a finite set, and $E$ is a subset of the unordered pairs of distinct elements of $V$.

A cost/weight/length function on a graph $G = (V, E)$ is a function $c : E \to \mathbb{R}$. We often write $c_e$ instead of $c(e)$.

**Def Bipartite Graph**: A graph $G = (V, E)$ is bipartite if we can partition the vertices into two sets, say $A$ and $B$, such that every edge has one endpoint in $A$ and one endpoint in $B$.

**Def Matching**: A matching in a graph $G = (V, E)$ is a subset $M$ of $E$ such that no vertex is incident to more than one edge of $M$.

**Def Perfect Matching**: Matching $M$ is a perfect matching of $G$ if every vertex of $G$ is incident to exactly one edge in the matching $M$. ($|M| = \frac{1}{2}|V|$)

Given a matching $M$ in a graph $G$ and a cost function $c : E \to \mathbb{R}$, the cost of $M$, written $c(M)$ is defined as $c(M) = \sum\limits_{e \in M} c_e$.

**Minimum Cost Perfect Matching Problem**: Given a graph $G = (V, E)$ and a cost function $c : E \to \mathbb{R}$, find the minimum cost of a perfect matching of $G$.

Define binary variables $x_e : e \in E$ where for each edge $e \in E$, $x_e = 0$ if $e \notin M$ and $x_e = 1$ if $e \in M$. We think of the variables $x_e : e \in E$ as forming a vector $x$ indexed by edges, or equivalently a function $x : E \to \{0, 1\}$. $x$ is the indicator function/vector for set $M$.

The objective function is

$$\sum_{e \in E} c_e x_e = \text{total cost of edges in matching with indicator function x}$$

we write $c(x)$ for this expression.
For each vertex $v$, write $\delta(v)$ for the set of edges that have $v$ as an end (incident with $v$). The constraint that $M$ is a perfect matching is equivalent to saying that for each $v \in V$, $M$ contains exactly one edge in $\delta(v)$.

$$\sum_{e \in \delta(v)} x_e = 1$$

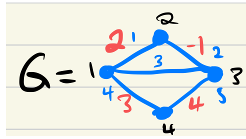
Translated to a statement about the vector $x$, this is saying that for each $v \in V$, we have

$$\sum_{e \in \delta(v)} x_e = 1 \iff x(\delta(v)) = 1$$

**IP**:

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e x_e \\ \text{s.t.} \quad & \sum_{e \in \delta(v)} x_e = 1 \qquad : v \in V \\ & \qquad x_e \in \{0, 1\} : e \in E \end{aligned}$$

Ex.




$$
\begin{aligned}
\min \quad & 2x_1 - x_2 + 0x_3 + 3x_4 + 4x_5 \\
\text{s.t.} \quad & x_1 + x_3 + x_4 = 1 \\
& x_1 + x_2 = 1 \\
& x_2 + x_3 + x_5 = 1 \\
& x_4 + x_5 = 1 \\
& x_1, \ldots, x_5 \in \{0, 1\}
\end{aligned}
$$

**Def Path**: A path in a graph $G = (V, E)$ is a sequence of edges of the form $v_1v_2, v_2v_3, \ldots, v_{k-1}v_k$ where $v_1, \ldots, v_k$ are distinct. We say this is a path from $v_1$ to $v_k$ and we often think of it as a set rather than a sequence.
If $l : E \to \mathbb{R}$ is a length function, then the length of a path $P$ is $\sum\limits_{e \in P} l_e$.

**Shortest Path Problem**: Given a graph $G$ and a length function $l : E \to \mathbb{R}$ and vertices $s, t$ of $G$, find a minimum length $st$-path. We solve this problem when every length is positive.

Let $G = (V, E)$ be a graph and let $U \subseteq V$ be a subset of the vertices. We generalize the definition of $\delta$. Let $\delta(U)$ denote the set of edges that have exactly one endpoint in $U$,

$$\delta(U) = \{uv \in E : u \in U, v \notin U\}$$

**Def *st*-cut**: An $st$-cut is a set of edges of the form $\delta(U)$, where $s \in U$ and $t \notin U$.

**Theorem 1.1** An $st$-path $P$ intersects every $st$-cut.

**Theorem 1.2** Let $S$ be a set of edges that contains at least one edge from every $st$-cut. Then there exists an $st$-path $P$ that is contained in the edges of $S$.

Variables: Introduce binary variable $x_e$ for every edge $e$, where $x_e = 1$ is when edge $e$ is selected to be part of our $st$-path, and $x_e = 0$ otherwise.

Constraints: By Theorem 1.2, in order to construct an $st$-path, it suffices to select one edge from every $st$-cut. Let $\delta(U)$ be an arbitrary $st$-cut of $G$. The number of edges we selected from the $st$-cut $\delta(U)$ is $\sum\limits_{e \in \delta(U)} x_e$. Since we wish to select at least one edge from $\delta(U)$, the constraint should be

$$\sum_{e \in \delta(U)} x_e \geq 1$$

Objective function: We wish to minimize the total length of the edges in the $st$-path $P$ we selected. If $e$ is an edge of $P$ then we will have $x_e = 1$ and we should contribute $c_e$ to the objective function.

$$\sum_{e \in E} c_e x_e$$

$$
\begin{aligned}
\min \quad & \sum_{e\in E} c_e x_e \\
\text{s.t.} \quad & \sum_{e\in\delta(U)} x_e \geq 1 : (U \subseteq V, s \in U, t \notin U) \\
& x_e \geq 0 : (e \in E) \\
& x_e \text{ integer}
\end{aligned}
$$

## 1.4 Nonlinear Programs

**Def Nonlinear Program (NLP)**: Consider functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g_i : \mathbb{R}^n \to \mathbb{R}$, for every $i \in \{1, 2, \ldots, m\}$. A nonlinear program is an optimization problem of the form

$$
\begin{aligned}
\min \quad & z = f(x) \\
\text{s.t.} \quad & g_1(x) \leq 0 \\
& g_2(x) \leq 0 \\
& \vdots \\
& g_m(x) \leq 0
\end{aligned}
$$

In order to maximize the objective function, we take $z = -f(x)$.

**Techniques for nonlinear program constraints**

- $x_1 \in \mathbb{Z}$ is $x_1 - \lceil x_1 \rceil = 0$
- $g_1(x) \leq 0$ or $g_2(x) \leq 0$ is $\min(g_1(x), g_2(x))$
- If $g_1(x) \leq 0$ then $g_2(x) \leq 0$. Define $f : \mathbb{R} \to \mathbb{R}$ such that

  $$f(y) = \begin{cases} 1 & (y \leq 0) \\ 0 & (y > 0) \end{cases}$$

  Then the constraint is $f(g_1(x)) \cdot g_2(x) \leq 0$.

# Chapter 2

# Solving Linear Programs

## 2.1 Possible Outcomes

There are three possible outcomes for an LP:

- Unbounded

$$\begin{aligned} \max \quad & x_1 \\ \text{s.t.} \quad & x_1 \geq 0 \end{aligned}$$

- Infeasible

$$\begin{aligned} \max \quad & x_1 \\ \text{s.t.} \quad & x_1 \geq 1 \\ & x_1 \leq -1 \end{aligned}$$

- Optimal (solution exists)

$$\begin{aligned} \max \quad & x_1 \\ \text{s.t.} \quad & x_1 \leq 4 \end{aligned}$$

Claim the following has no feasible solution ($\nexists x_1, x_2, x_3$ satisfying all the constraints).

$$\begin{aligned} \min \quad & x_1 + x_2 + 7x_3 \\ \text{s.t.} \quad & 2x_1 - x_2 + x_3 = 3 \quad A \\ & x_1 - 3x_2 - x_3 = 2 \quad B \\ & x_1 + x_2 + 3x_3 = 3 \quad C \\ & x_1, x_2, x_3 \geq 0 \end{aligned}$$

Consider the constraint $2A - 2B - C$. This says

$$2(2x_1 - x_2 + x_3) - 2(x_1 - 3x_2 - x_3) - (x_1 + x_2 + 3x_3) = 2(3) - 2(2) - 3 \Leftrightarrow x_1 + 3x_2 + x_3 = -1$$

but $x_1, x_2, x_3 \geq 0$ for every feasible solution so $0 \leq x_1 + 3x_2 + x_3 = -1 < 0$. This is a contradiction, so there is no feasible solution.

**Proposition** Consider the LP

$$\max\{c^T x : Ax = b, x \geq 0\}$$

If there is a vector $y$ such that $y^T A \geq 0$ and $y^T b < 0$, then the LP is infeasible (no $x$ that satisfies the constraints) and $y$ provides a certificate (proof of this fact).

**Proof**: Suppose that such a $y$ exists, and $x$ is a feasible solution. Then $Ax = b$ so $y^T Ax = y^T b$. But $y^T A \geq 0$ by assumption, and $x > 0$ because $x$ is feasible so $(y^T A)x \geq 0$ because it is a dot product of non-negative vectors. This is a contradiction, because $0 \leq (y^T A)x = y^T b < 0$.

From previous example, $y = \begin{pmatrix} 2 \\ -2 \\ -1 \end{pmatrix}$, then $y^T A = \begin{pmatrix} 1 & 3 & 1 \end{pmatrix} \geq 0$, $y^T b = -1 < 0$. So infeasible by proposition.

**Def Infeasible**: An LP is infeasible if it has no feasible solution.

**Def Unbounded**: An LP is unbounded if, for all $k$, there exists a feasible solution $x$ with object value better than $k$ (less for minimization, more for maximization).

Ex. Consider the LP

$$\max\{z(x) = c^T x : Ax = b, x \geq 0\}$$

where

$$A = \begin{pmatrix} 1 & 1 & -3 & 1 & 2 \\ 0 & 1 & -2 & 2 & -2 \\ -2 & -1 & 4 & 1 & 0 \end{pmatrix}, b = \begin{pmatrix} 7 \\ 2 \\ -3 \end{pmatrix}, c = \begin{pmatrix} -1 \\ 0 \\ 3 \\ 7 \\ 1 \end{pmatrix}, x = \begin{pmatrix} x_1 \\ \vdots \\ x_5 \end{pmatrix}$$

Let $\bar{x} = (2, 0, 0, 1, 2)^T$ and $d = (1, 2, 1, 0, 0)^T$. Note that $A\bar{x} = b$, so $\bar{x}$ is feasible. Note also that $Ad = 0$. Therefore, for all $t \in \mathbb{R}$, we have $A(\bar{x} + td) = A\bar{x} + tAd = b + 0 = b$.

Now $c^T \bar{x} = 7$ and $c^T d = 2$, so $c^T(\bar{x} + td) = 7 + 2t$. It follows that for all $t \geq 0$, the point $x(t) = \bar{x} + td$ is feasible with objective value $7 + 2t$. Since $7 + 2t \to \infty$ as $t \to \infty$, it follows that the LP is unbounded.

**Proposition** Consider the LP

$$\max\{c^T x : Ax = b, x \geq 0\}$$

If there exists a vector $d \geq 0$ with $Ad = 0$ and $c^T d > 0$, and there exists $\bar{x} \geq 0$ with $A\bar{x} = b$, then the LP is unbounded (for all $M$, there is a feasible solution $x'$ with $c^T x' > M$) and $\bar{x}, d$ provide a certificate of this.

**Proof**: If such an $\bar{x}, d$ exists, let $x(t) = \bar{x} + td$, for each $t \geq 0$. Since $\bar{x} \geq 0, d \geq 0$, we have $\bar{x} + td \geq 0$, so $x(t) \geq 0$ and $A(x(t)) = A(\bar{x} + td) = A\bar{x} + tAd = A\bar{x} = b$. So $x(t)$ is feasible for all $t \geq 0$. Now

$$c^T(x(t)) = c^T(\bar{x} + td) = c^T \bar{x} + t(c^T d)$$

This function is the objective value at $x(t)$, but since $c^T d > 0$, it approaches $\infty$ as $t \to \infty$. So the LP is unbounded.

A proof of optimality. Consider (L1)

$$\begin{aligned} \max \quad & (-3,0,-1,0,-2)x + 8 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & -1 & 2 & 3 \\ -2 & 3 & 4 & 0 & -2 \end{pmatrix} x = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{aligned}$$

Claim: $x = (0,1,0,2,0)$ is optimal. Proof: Note that $(0,1,0,2,0)$ gives an objective value of 8, and that any $x \geq 0$ satisfies $(-3,0,-1,0,-2)x \leq 0+0+0+8 = 8$, so no feasible solution can have a larger objective value. So $(0,1,0,2,0)$ is optimal.

Consider (L2)

$$\begin{aligned} \max \quad & (1,0,-5,8,10)x - 8 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & -1 & 2 & 3 \\ -2 & 3 & 4 & 0 & -2 \end{pmatrix} x = \begin{pmatrix} 4 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{aligned}$$

Claim: $x = (0,1,0,2,0)$ is optimal. Proof: If $x$ satisfy the constraints, then $x_1 - x_3 + 2x_4 + 3x_5 = 4$ (first row of matrix). So

$$\begin{aligned} (-3,0,-1,0,-2)x + 8 &= (-3,0,-1,0,-2)x + 8 + 4\overbrace{((1,0,-1,2,3)x - 4)}^{=0} \\ &= (1,0,-5,8,10)x - 8 && \text{(objective function)} \end{aligned}$$

So (L1) and (L2) have the same feasible solutions, and also every feasible solution has the same objective values for (L1) and (L2).

## 2.2 Standard Equality Form

**Def Standard Equality Form (SEF)**: A linear program is said to be in standard equality form if it is of the form

$$\max\{c^T x + \bar{z} : Ax = b, x \geq 0\}$$

where $\bar{z}$ denotes some constant. In other words, an LP in SEF satisfies these conditions

- it is a maximization problem
- other than the nonnegativity constraints, all constraints are equality constraints
- every variable has a nonnegativity constraint

Ex.

$$\begin{aligned} \max \quad & (2,-3,4)x \\ \text{s.t.} \quad & x_1 + x_2 = 7 \\ & x_2 + x_3 = 9 \\ & x \geq 0 \end{aligned}$$

**Def Equivalent Linear Programs**: Two LPs (P1) and (P2) are equivalent if

- (P1) is infeasible if and only if (P2) is infeasible
- (P1) is unbounded if and only if (P2) is unbounded
- Given any optimal solution for (P1), we can construct an optimal solution for (P2), and vice versa.

We argue that every LP is equivalent to some LP in SEF.

Let (P) be an arbitrary LP. If (P) is a minimization problem, then replacing the objective function with its negative and changing minimization to maximization gives an equivalent maximization problem.

To change inequality constraints to equality constraints, we introduce a slack variable $y$ representing how far the inequality $a^T x \le b$ is from being tight. The two LPs are equivalent because whenever $x$ is feasible for (P1), $(x, b - a^T x)$ is feasible for (P2) with the same objective value. and whenever $(x, y)$ is feasible for (P2), $x$ is feasible for (P1) with the same objective value.

$$\begin{aligned} \max \quad & (2,3,1)x + 1 \\ \text{s.t.} \quad & x_1 + x_2 = 7 \\ & x_1 - x_3 \le 6 \\ & x \ge 0 \end{aligned}$$

is the same as

$$\begin{aligned} \max \quad & (2,3,1)x + 1 \\ \text{s.t.} \quad & x_1 + x_2 = 7 \\ & x_1 - x_3 + y = 6 \\ & x, y \ge 0 \end{aligned}$$

Another one is if $x_k \le 0$ then we can change it to $x_k + y = 0, y \ge 0$.

Equivalent LPs for nonnegativity constraint:

$$\begin{aligned} \max \quad & x_1 + x_2 + x_3 \\ \text{s.t.} \quad & 2x_1 + x_3 = 17 \\ & x_1 - x_2 \le 5 \\ & x_1, x_3 \ge 0 \end{aligned}$$

is the same as

$$\begin{aligned} \max \quad & x_1 + (x_2' - x_2'') + x_3 \\ \text{s.t.} \quad & 2x_1 + x_3 = 17 \\ & x_1 - (x_2' - x_2'') \le 5 \\ & x_1, x_2', x_2'', x_3 \ge 0 \end{aligned}$$

For any feasible solution $x$ to (P1), if $x_2 \ge 0$, then $(x_1, x_2, 0, x_3)$ is feasible for $(P2)$ with same objective value and if $x_2 < 0$, then $x(x_1, 0, -x_2, x_3)$ is feasible for (P2) with the same objective value.

Conversely, if $(x_1, x_2', x_2'', x_3)$ is feasible for (P2), then $(x_1, x_2' - x_2'', x_3)$ is feasible for (P1) with same objective value. So (P1) and (P2) are equivalent.

**Changing an LP to SEF**: Given an arbitrary LP (P), we can find an equivalent LP (P') in SEF by

(1) Changing minimization to maximization and negativizing objective function if (P) is a minimization problem

(2) a. For each $\leq$ constraint, $a_i^T x \leq b_i$, introducing a slack variable $y_i$, and replacing $a_i^T x \leq b_i$ with $a_i^T x + y_i = b_i, y_i \geq 0$.

  b. For each $\geq$ constraint, $a_i^T x \geq b_i$, introducing a slack variable $y_i$, and replacing $a_i^T x \geq b_i$ with $a_i^T x - y_i = b_i, y_i \geq 0$.

(3) For each variable $x_i$ that does not have an $x_i \geq 0$ constraint, replacing it everywhere with $x_i' - x_i''$ for new variables $x_i', x_i''$, and adding constraints $x_i', x_i'' \geq 0$.

## 2.3 Bases and Canonical Forms

Consider the LP

$$\begin{array}{rc} \max & 2x_1 + 7x_3 + 6 \\ \text{s.t.} & \begin{pmatrix} 1 & 1 & 0 & 0 & -2 \\ -2 & 0 & 1 & 0 & 3 \\ 3 & 0 & 0 & 1 & 7 \end{pmatrix} x = \begin{pmatrix} 2 \\ 6 \\ 5 \end{pmatrix} \\ & x \geq 0 \end{array}$$

Note that $x = (0, 2, 6, 5, 0)^T$ is a feasible solution with objective value 6 because $2e_1 + 6e_2 + 5e_e = (2, 6, 5)^T$ from the identity matrix inside $A$.

Idea: Starting with $(0, 2, 6, 5, 0)^T$, try to increase the first entry, keep the fifth entry the same, and alter entries $2, 3, 4$ to remain feasible.

Let $t \geq 0$, and let $x = (t, x_2, x_3, x_4, 0)^T$; try to find $x_2, x_3, x_4$ to make this $x$ feasible. So we need $Ax = b$, i.e.

$$\begin{aligned} t + x_2 = 2 &\implies x_2 = 2 - t \\ -2t + x_3 = 6 &\implies x_3 = 6 + 2t \\ 3t + x_4 = 5 &\implies x_4 = 5 - 3t \end{aligned}$$

So $x = (t, 2 - t, 6 + 2t, 5 - 3t, 0)^T$ satisfies $Ax = b$ for each $t \geq 0$.

$$x = \begin{pmatrix} 0 \\ 2 \\ 6 \\ 5 \\ 0 \end{pmatrix} + t \begin{pmatrix} 1 \\ -1 \\ 2 \\ -3 \\ 0 \end{pmatrix}$$

For $x$ to be feasible, we need $x \geq 0$ also. This is equivalent to

$$\begin{aligned} 2-t \geq 0 &\Leftrightarrow t \leq 2 \\ 6+2t \geq 0 &\Leftrightarrow \text{always true for } t \geq 0 \\ 5-3t \geq 0 &\Leftrightarrow t \leq \frac{5}{3} \end{aligned}$$

So for all $t \leq \frac{5}{3}, x \geq 0$, $x$ is feasible. We want $t = x_1$ to be as large as possible, so we set $t = \frac{5}{3}$, and this gives

$$x = \begin{pmatrix} 0 \\ 2 \\ 6 \\ 5 \\ 0 \end{pmatrix} + \frac{5}{3} \begin{pmatrix} 1 \\ -1 \\ 2 \\ -3 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{5}{3} \\ \frac{1}{3} \\ \frac{28}{3} \\ 0 \\ 0 \end{pmatrix}$$

This $x$ has objective value $2x_1 + 7x_5 + 6 = \frac{10}{3} + 6 > 6$.

**Def Basis**: Given a real matrix $A$ with a set of column indices $\mathcal{I} = \{1, \ldots, k\}$, a basis of $A$ is a set $B \subseteq \mathcal{I}$ for which the submatrix $A_B$ is a square nonsingular matrix. $A_B$ denotes the submatrix obtained from $A$ by only including columns in $B$.

Ex. $A = \begin{pmatrix} 2 & 1 & 2 & 3 & 0 & 0 \\ 1 & 0 & 2 & 1 & 1 & 0 \\ -3 & 0 & 4 & -3 & 0 & 1 \end{pmatrix}$. $\{2,5,6\}$ is a basis in the matrix $A$ because $A_{\{2,5,6\}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is nonsingular and square.

Ex. $\{1,2,4\}$ is not a basis in $A$ because $A_{\{1,2,4\}}$ is singular. ($(1,1,-1)^T$ is in nullspace).

Note: If the number of rows of $A > \operatorname{rank}(A)$, then $A$ has no bases; otherwise $A$ has $\geq 1$ basis, and it can be found by row reductions.

**Proposition** If $A$ and $A'$ are row-equivalent and have the same number of rows, then they have the same bases.

**Proof**: Elementary row operations preserve whether linear combinations of columns are zero, and hence nonsingularity.

If $B \subseteq \mathcal{I}$ is a basis for $A$, then a vector $x$ is a basic solution to the equation $Ax = b$ if $x_i = 0$ for all $i \in \mathcal{I} \setminus B$.

Ex. $x = (0,2,0,0,3,4)^T$ is a basic solution to $Ax = (2,3,4)^T$, for the basis $\{2,5,6\}$ of $A$.

Ex. $x = (1,0,1,0,0,0)^T$ is a basic solution to $Ax = (4,3,1)^T$ for the basis $\{1,3,6\}$ and also for the basis $\{1,3,5\}$.

Given a basis $B \subseteq \mathcal{I}$, the indices in $B$ are basic and the indices outside $B$ are nonbasic. Similarly, the entries $x_i$ are basic/nonbasic. If $x$ is a basic solution to $Ax = b$ for some basis $B$, and also $x \geq 0$, then $x$ is a basic feasible solution to $\max\{c^T x : Ax = b, x \geq 0\}$.

**Proposition** For each basis $b$ of $A$ and vector $b$, there is exactly one basic solution to $Ax = b$. (It may or may not be feasible)

Proof: Solving $Ax = b$ with the restriction that $x_i = 0$ for all $i \notin B$ is equivalent to solving $A_B x = b$; this has a unique solution because $A_B$ is a nonsingular square matrix.

Recall: If $A, b, c, B \subseteq \{\text{cols of } A\}$ are such that

- $A_B$ is an identity matrix
- $c_i = 0$ for all $i \in B$ (equivalently $c_B = 0$)
- $x$ is a feasible solution that is basic for $B$

then there is a way to find a feasible solution to $\max\{c^T x + d : Ax = b, x \geq 0\}$ with a better objective value than $x$.

**Def Canonical Form**: Consider the LP in SEF:

$$\max\{c^T x + z : Ax = b, x \geq 0\}$$

Let $B$ be a basis of $A$. We say that the LP is in canonical form for $B$ if the following conditions are satisfied:

- $A_B$ is an identity matrix
- $c_B = 0$

Ex. Converting to canonical form

$$\begin{array}{ll} \max & x_1 + 2x_3 - x_3 + 2x_4 + 3x_5 + 6 \\ \text{s.t.} & \begin{pmatrix} 1 & 3 & 0 & 2 & 4 \\ 0 & 2 & 1 & 0 & 0 \\ -2 & -4 & -1 & 1 & -7 \end{pmatrix} x = \begin{pmatrix} 13 \\ 2 \\ -25 \end{pmatrix} \\ & \qquad\qquad\qquad\qquad\qquad x \geq 0 \end{array}$$

We want to convert(without changing the set of feasible solutions or the value of objective function) this to canonical form for the basis $B = \{1, 3, 5\}$.

Goal: Row-reduce to change $Ax = b$ into $A'x = b'$, where $A'_B = I_3$

$$\begin{aligned} \begin{pmatrix} 1 & 3 & 0 & 2 & 4 & | & 13 \\ 0 & 2 & 1 & 0 & 0 & | & 2 \\ -2 & -4 & -1 & 1 & -7 & | & -25 \end{pmatrix} &\sim \begin{pmatrix} 1 & 3 & 0 & 2 & 4 & | & 13 \\ 0 & 2 & 1 & 0 & 0 & | & 2 \\ 0 & 2 & -1 & 5 & 1 & | & 1 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & 3 & 0 & 2 & 4 & | & 13 \\ 0 & 2 & 1 & 0 & 0 & | & 2 \\ 0 & 4 & 0 & 5 & 1 & | & 3 \end{pmatrix} \\ &\sim \begin{pmatrix} 1 & -9 & 0 & -18 & 0 & | & 1 \\ 0 & 2 & 1 & 0 & 0 & | & 2 \\ 0 & 4 & 0 & 5 & 1 & | & 3 \end{pmatrix} \end{aligned}$$

So $Ax = b$ has the same solution set as $A'x = b'$ where $A' = \begin{pmatrix} 1 & -9 & 0 & -18 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 4 & 0 & 5 & 1 \end{pmatrix}$ and $b' = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$.

So what we did was $A_B^{-1} A x = A_B^{-1} b$

So an equivalent LP is

$$\begin{array}{lc}\max & x_1 + 2x_3 - x_3 + 2x_4 + 3x_5 + 6 \\ \text{s.t.} & \begin{pmatrix} 1 & -9 & 0 & -18 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 4 & 0 & 5 & 1 \end{pmatrix} x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{array}$$

We wish to change the form of the objective function $x_1 + 2x_3 - x_3 + 2x_4 + 3x_5 + 6$ to have the coefficients of $x_1, x_3, x_5$ to be equal to 0 without changing its value. We do this by using the equation $A'x = b'$ to eliminate $x_1, x_3, x_5$ for all feasible solutions.

Note that every feasible solution $x$ satisfies

$$\begin{aligned} x_1 - 9x_2 - 18x_4 &= 1 \\ 2x_2 + x_3 &= 2 \\ 4x_2 + 5x_4 + x_5 &= 3 \end{aligned}$$

So for feasible $x$, objective function is

$$\begin{aligned} x_1 + 2x_3 - x_3 + 2x_4 + 3x_5 + 6 &= x_1 + 2x_3 - x_3 + 2x_4 + 3x_5 + 6 \\ &\quad - 1(x_1 - 9x_2 - 18x_4 - 1) \\ &\quad + 1(2x_2 + x_3 - 2) \\ &\quad - 3(4x_2 + 5x_4 + x_5 - 3) \\ &= x_2 + 5x_4 + 14 \end{aligned}$$

So the canonical form LP is

$$\begin{array}{lc}\max & x_2 + 5x_4 + 14 \\ \text{s.t.} & \begin{pmatrix} 1 & -9 & 0 & -18 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 4 & 0 & 5 & 1 \end{pmatrix} x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \\ & x \geq 0 \end{array}$$

The basic feasible solution is $x = (1, 0, 2, 0, 3)^T$.

**Proposition** Suppose an LP

$$\max\{c^T x + \bar{z} : Ax = b, x \geq 0\}$$

and a basis $B$ of $A$ are given. Then the following LP is an equivalent LP in canonical form for the basis $B$, and $y = c_B^T A_B^{-1}$ (row vector), then:

$$\begin{array}{lc}\max & (c^T - yA)x + (yb + \bar{z}) \\ \text{s.t.} & A_B^{-1}Ax = A_B^{-1}b \\ & x \geq \mathbf{0} \end{array}$$

Proof: First note that, since $A_B$ is nonsingular, we have $Ax = b$ if and only if $A_B^{-1}Ax = A_B^{-1}b$, so feasible solutions are the same.

Second, note that $(A_B^{-1}A)_B = A_B^{-1}A_B = I$.
Third, for each feasible $x$, we have

$$\begin{aligned}(c^T - yA)x + yb + \bar{z} \\ &= c^T x - c_B^T A_B^{-1} x + c_B^T A_B^{-1} b + \bar{z} \\ &= c^T x + \bar{z} - \underbrace{c_B^T A_B^{-1}(Ax - b)}_{=0} \\ &= c^T x + \bar{z}\end{aligned}$$

Finally, $(c^T - yA)_B = (c^T - c_B^T A_B^{-1} A)_B = c_B^T - c_B^T A_B^{-1} A_B = 0$ so $c^T - yA$ has all its $B$-entries zero.

## 2.4 The Simplex Algorithm

Ex. Exploiting Canonical form

$$\begin{aligned}\max \quad & (0, 2, 0, -1, 0)x + 10 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 1 & -\frac{1}{2} & 0 \\ 0 & \frac{3}{2} & 0 & \frac{1}{2} & 1 \end{pmatrix} x = \begin{pmatrix} 5 \\ 1 \\ 9 \end{pmatrix} \\ & x \geq 0\end{aligned}$$

This is in canonical form for $B = \{1, 3, 5\}$ and $x = (5, 0, 1, 0, 9)$ is a basic feasible solution for $B$.

We want to find $t \geq 0$ and a vector $d = (-\frac{1}{2}t, t, -\frac{1}{2}t, 0, -\frac{3}{2}t)$ such that $x + d$ is feasible with better objective value than $x$. This amounts to finding a vector $x_B = (x_1, x_3, x_5)^T$ s.t.

$$x_B = \begin{pmatrix} 5 \\ 1 \\ 9 \end{pmatrix} - t \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{3}{2} \end{pmatrix} \geq 0$$

and $(5, 0, 1, 0, 9) + t(-\frac{1}{2}, 1, -\frac{1}{2}, 0, -\frac{3}{2}) \geq 0$. We want to find the largest $t$ s.t. this is true. This value is $t = 2$ where we have $1 - \frac{t}{2} = 0$ ($x_3$ becomes 0).

This gives a new value for $x$ of $(5, 0, 1, 0, 9) + 2(-\frac{1}{2}, 1, -\frac{1}{2}, 0, -\frac{3}{2}) = (4, 2, 0, 0, 6)$. This has objective value $14 > 10$.

$x_1 = (4, 2, 0, 0, 6)$ is not a basic solution for $B = \{1, 3, 5\}$ because it has a nonzero entry in position 2. But $x_1$ is a basic feasible solution for $B_1 = \{1, 2, 5\}$. So to try to increase objective value again, we convert the LP to canonical form for $B_1 = \{1, 2, 5\}$.

$$\begin{aligned}\left(\begin{array}{ccccc|c} 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 5 \\ 0 & \frac{1}{2} & 1 & -\frac{1}{2} & 0 & 1 \\ 0 & \frac{3}{2} & 0 & \frac{1}{2} & 1 & 9 \end{array}\right) &\sim \left(\begin{array}{ccccc|c} 1 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 5 \\ 0 & 1 & 2 & -1 & 0 & 2 \\ 0 & \frac{3}{2} & 0 & \frac{1}{2} & 1 & 9 \end{array}\right) \\ &\sim \left(\begin{array}{ccccc|c} 1 & 0 & -1 & 1 & 0 & 4 \\ 0 & 1 & 2 & -1 & 0 & 2 \\ 0 & 0 & -3 & 2 & 1 & 6 \end{array}\right)\end{aligned}$$

and

$$\begin{aligned}(0,2,0,-1,0)x+10 &= (0,2,0,-1,0)x+10\\ &\quad -2(\underbrace{(0,1,2,-1,0)x-2}_{=0})\\ &= (0,0,-4,1,0)x+14\end{aligned}$$

New LP:

$$\begin{aligned}\max \quad & (0,0,-4,1,0)x+14\\ \text{s.t.} \quad & \begin{pmatrix}1 & 0 & -1 & 1 & 0\\ 0 & 1 & 2 & -1 & 0\\ 0 & 0 & -3 & 2 & 1\end{pmatrix} x = \begin{pmatrix}4\\2\\6\end{pmatrix}\\ & x \geq 0\end{aligned}$$

new $x_1 = (4,2,0,0,6)^T$ is a basic feasible solution for $B_1 = \{1,2,5\}$.

As before, we want $x_1 + t(-1,1,0,1,-2) \geq 0$. $d = (1,-1,0,1,2)$ and this vector has to satisfy $Ad = 0$ so that $A(x+td) = Ax$.

$t = \min\{4/1, -, 6/2\} = 3$. This gives a better feasible solution

$$x_2 = (4,2,0,0,6)^T + 3(-1,1,0,1,-2)^T = (1,5,0,3,0)^T$$

This is a basic feasible solution for $B_2 = \{1,2,4\}$. Converting the problem into canonical form for $B_2$ gives

$$\begin{aligned}\max \quad & (0,0,-\frac{5}{2},0,-\frac{1}{2})+17\\ \text{s.t.} \quad & \begin{pmatrix}1 & 0 & \frac{1}{2} & 0 & -\frac{1}{2}\\ 0 & 1 & \frac{1}{2} & 0 & \frac{1}{2}\\ 0 & 0 & -\frac{3}{2} & 1 & \frac{1}{2}\end{pmatrix} x = \begin{pmatrix}1\\5\\3\end{pmatrix}\\ & x \geq 0\end{aligned}$$

We cannot continue by choosing a positive entry in $c$ because $c \leq 0$ but $x_2$ is optimal, because $c^T x_2 + 17 = 17$ and for all feasible $x$, $c^T x + 17 \leq 17$, because $c \leq 0, x \geq 0$ so $c^T x \leq 0$. So we can stop, as we have found an optimal solution.

Suppose that we have an LP in canonical form for a basis $B$

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

where $c_N$ is the non basic entries in $c$ and let $\bar{x}$ be the basic feasible solution for $B$, i.e. $\bar{x}_N = 0$.

**Proposition** If $c_N \leq 0$ (equivalently $c \leq 0$), then $\bar{x}$ is optimal.

Proof: $\bar{x}$ is feasible, and has objective function $c_N^T x_N + \bar{z} = 0 + \bar{z} = \bar{z}$ because $\bar{x}_N = 0$. For an arbitrary feasible solution $x$, we have $x \geq 0$ and $c_N \leq 0$, so $c_N^T x_N \leq 0$. So $x$ has objective value $\leq 0 + \bar{z} = \bar{z}$. So $\bar{x}$ is optimal.

**Proposition** If $k \in N$ ($N$ is non basis) is such that $c_k > 0$ and $A_k \le 0$, then the problem is unbounded.

Proof: Let $d \in \mathbb{R}^n$ with $d_k = 1$, $d_j = 0$ for all $j \in N \setminus \{k\}$, and $d_{b_i} = -A_{k,i}$ for $1 \le i \le m$, where $b_i$ is the $i$th entry of $B$. ex. $d = (0, 0, 0, \dots, 1, 0, 0, \dots, -a_1, -a_2, \dots, -a_m)$.

Since all entries of $A_k$ are $\le 0$, we have $d \ge 0$. Since $d_k = 1, c_k > 0$, and every entry other than $k$ is either 0 in $c$ or 0 in $d$, we have $c^T d = c_k \cdot 1 > 0$.

Finally,

$$\begin{aligned} Ad &= \underbrace{\sum_{i \in N} d_i A_i}_{d_k A_k} + \sum_{j \in B} d_j A_j \\ &= d_k A_k + \sum_{i=1}^{m} (-A_{k,i}) \cdot e_j \\ &= 1 \cdot A_k + (-A_k) \\ &= 0 \end{aligned}$$

Since $\bar{x}$ is feasible, $d \ge 0, c^T d > 0, Ad = 0$, then the problem is unbounded.

If the hypotheses for neither proposition holds, then there is some $k \in N$ for which $c_k > 0$, and $A_{j,k} > 0$ for some $j \in \{1, \dots, m\}$. If this is the case, then we can choose $j \in \{1, \dots, m\}$ such that $A_{j,k} > 0$ while $\frac{b_j}{A_{j,k}}$ is as small as possible, set $B' = (B \setminus \{j\}) \cup \{k\}$, convert into canonical form for $B'$, and get a feasible solution with objective value at least as large.

**One-Phase Simplex Algorithm**: Starting with a basic feasible solution for a basis $B_0$, convert to canonical form for $B_0$, If the hypotheses for proposition 1 or 2 hold, then STOP because either $\bar{x}$ is optimal (Prop 1) or the problem is unbounded (Prop 2). Otherwise, choose $j$ and $k$ as above, and set $B_1 = (B_0 \setminus \{j\}) \cup \{k\}$ and repeat.

**Theorem Bland's Rule** Choose the smallest allowed $j$ and $k$ for which $c_k > 0$ and $j$ for which the minimum $\frac{b_j}{A_{j,k}}$ occurs, then the algorithm terminates.

The one-phase simplex method requires a basis $B$ for which that basic solution $\bar{x}$ is feasible to get started.

Ex. (P):

$$\begin{aligned} \max \quad & (1, 6, 4)x + 2 \\ \text{s.t.} \quad & \begin{pmatrix} 7 & 1 & 3 \\ -1 & 2 & -4 \end{pmatrix} x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ & x \ge 0 \end{aligned}$$

(P')

$$\begin{aligned} \min \quad & x_4 + x_5 \\ \text{s.t.} \quad & \begin{pmatrix} 7 & 1 & 3 & 1 & 0 \\ -1 & 2 & -4 & 0 & 1 \end{pmatrix} x = \begin{pmatrix} 2 \\ 3 \end{pmatrix} \\ & x \ge 0 \end{aligned}$$

**Algorithm 1 Simplex Algorithm**

$$\max\{c^T x : Ax = b, x \geq \mathbf{0}\}$$

**Input**: Linear program (P) and feasible basis $B$

**Output**: An optimal solution of (P) or a certificate proving that (P) is unbounded.

1: Rewrite (P) so that it is in canonical form for the basis $B$. Let $\bar{x}$ be the basic feasible solution for $B$

2: **if** $c_N \leq \mathbf{0}$ **then**

3: **Stop** ($\bar{x}$ is optimal)

4: Select $k \in N$ such that $c_k > 0$

5: **if** $A_k \leq \mathbf{0}$ **then**

6: **Stop** ((P) is unbounded)

7: Let $r$ be any index $i$ where the following minimum is attained:

$$t = \min\left\{\frac{b_i}{A_{i,k}} : A_{i,k} > 0\right\}$$

8: Let $l$ be the $r$th basis element

9: Set $B = B \cup \{k\} \setminus \{l\}$

10: Go to step 1

## 2.5 Finding Feasible Solutions

Note that

- $B = \{4, 5\}$ is a basis in (P') and that $(0, 0, 0, 2, 3)^T$ is a corresponding basic feasible solution.
- No solution to (P') has objective value $< 0$ (so optimum $\geq 0$)
- If $(x_1, x_2, x_3)$ is feasible in (P), then $(x_1, x_2, x_3, 0, 0)$ is feasible in (P') and has objective value 0, so is optimal. Hence, if (P') has optimum $> 0$, then (P) is infeasible.
- If 0 is the optimum for (P'), and $(x_1, x_2, x_3, x_4, x_5)$ is an optimal solution that is basic for some basis $B$, then $x_4 = x_5 = 0$ and hence $(x_1, x_2, x_3)$ is a feasible solution for (P) that is basic for some basis.

**Two-Phase Simplex Algorithm**: Given an LP (P) in SEF,

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

construct (P') by scaling rows of $A$ and entries of $b$ to make $b \geq 0$, then writing

$$\begin{aligned} \min \quad & x_{n+1} + \cdots + x_{n+m} \\ \text{s.t.} \quad & \begin{pmatrix} A & \mid & I \end{pmatrix} \begin{pmatrix} x \\ x_{n+1} \\ \vdots \\ x_{n+m} \end{pmatrix} = b \\ & \qquad\qquad x \geq 0 \end{aligned}$$

Now $(0|b^T)$ is a basic feasible solution for (P'): run one-phase simplex on (P').

- If (P') has optimum $> 0$, then STOP: (P) is infeasible.
- If (P') has optimum 0 and $\bar{x}$ is an optimal solution given by simplex, then $\bar{x}_{n+1}, \ldots, \bar{x}_{n+m} = 0$; now $(\bar{x}_1, \ldots, \bar{x}_n)$ is a basic feasible solution for (P). Run one-phase simplex on (P) to either find an optimal solution, or output unbounded.

**Fundamental Theorem of Linear Programming** Let (P) be an LP problem. Then exactly one of the following holds:

(1) (P) is infeasible

(2) (P) is unbounded

(3) (P) has an optimal solution

## 2.6 Geometry

The feasible region of an LP is the set of feasible solutions if there are $n$ variables $x_1, \ldots, x_n$, this region is a subset of $\mathbb{R}^n$.

Ex.

$$\begin{aligned} a_1^T x &= b_1 \\ a_2^T x &\leq b_2 \\ a_3^T x &\geq b_3 \end{aligned}$$

Then the subsets $S_1 = \{x : a_1^T x = b_1\}, S_2 = \{x : a_2^T x \leq b_2\}, S_3 = \{x : a_3^T x \geq b_3\}$. Then the feasible region is $S_1 \cap S_2 \cap S_3$. $S_1$ is a plane through $\mathbb{R}^3$.

**Def Euclidean Norm**: The Euclidean norm of $x$ is defined as

$$\|x\| = \sqrt{x_1^2 + \cdots + x_n^2}$$

**Proposition** Let $a, b \in \mathbb{R}^n$. Then $a^T b = \|a\| \, \|b\| \cos(\theta)$, where $\theta$ is the angle between $a$ and $b$. Therefore, for every pair of nonzero vectors $a, b$, we have:

- $a^T b = 0$ if and only if $a, b$ are orthogonal.

- $a^T b > 0$ if and only if the angle between $a, b$ is less than $90°$.
- $a^T b < 0$ if and only if the angle between $a, b$ is larger than $90°$.

**Def Hyperplane and Halfspace**: Let $a \in \mathbb{R}^n$ be a nonzero vector and let $b \in \mathbb{R}$, we define:

(1) $H = \{x \in \mathbb{R}^n : a^T x = b\}$ is a hyperplane.

(2) $F = \{x \in \mathbb{R}^n : a^T x \leq b\}$ is a halfspace.

**Def Feasible Region**: The feasible region of an optimization problem is the set of feasible solutions.

**Proposition** The feasible region of an LP is a polyhedron or equivalently the intersection of a finite number of halfspaces.

Proof: For each constraint $a^T x \leq b$, let $F = \{x : a^T x \leq b\}$. For each constraint $a^T x \geq b$, let $F = \{x : -a^T x \leq -b\}$. For each constraint $a^T x = b$, let $F_1 = \{x : a^T x \leq b\}$ and $F_2 = \{x : -a^T x \leq -b\}$. Taking the intersection of $F$ (or $F_1, F_2$) over all constraints expresses the feasible region as a finite intersection of halfspaces.

**Def Polyhedron**: A polyhedron is a finite intersection of halfspaces. (Note: this differs from various more informal notions of polyhedron).

**Remark**: A hyperplane in $\mathbb{R}^n$ is a translation of an $n-1$-dimensional subspace.

Proof: Given $H = \{x \in \mathbb{R}^n : a^T x = b\}$, let $H_0$ be the subspace of $\{x \in \mathbb{R}^n : a^T x = 0\}$ of dim $= n-1$. Fix $x_0 \in H$. Then $x \in H$ iff $a^T x = b$ iff $a^T x = a^T x_0$ iff $a^T(x - x_0) = 0$ iff $x - x_0 \in H_0$. So $x \in H$ iff $x - x_0 \in H_0$; thus $H$ is a translation of $H_0$ by $x_0$.

Given a halfspace $F = \{x \in \mathbb{R}^n : a^x \leq b\}$, fix $x_0 \in \mathbb{R}$ such that $a^T x_0 = b$, then $x \in F$

- iff $a^T x \leq b$
- iff $a^T x \leq a^T x_0$
- iff $a^T(x - x_0) \leq 0$
- iff $\|a\| \, \|x - x_0\| \cos(\theta) \leq 0$ where $\theta$ is the angle between $a$ and $x - x_0$, $0° \leq \theta < 180°$
- iff $\cos(\theta) \leq 0$
- iff $\theta \geq 90°$

Thus, $F$ is the set of all points at angle $\geq 90°$ from $a$, translated by $x_0$ (stuff on one side of the hyperplane $H$).

**Proposition** Let $\bar{x} \in H$.

- $H$ is the set of points $x$ for which $a$ and $x - \bar{x}$ are orthogonal.
- $F$ is the set of points $x$ for which $a$ and $x - \bar{x}$ form an angle of at least $90°$.

**Def Convex**: A set $S \subseteq \mathbb{R}^n$ is said to be convex if for every pair of points $x, y \in S$, the line segment with ends $x, y$ is included in $S$.

$$\{\lambda x + (1-\lambda)y : 0 \leq \lambda \leq 1\}$$

**Proposition** If $\mathcal{S}$ is a (possibly infinite) collection of convex sets, then the intersection of the sets in $\mathcal{S}$, $\bigcap\limits_{s \in \mathcal{S}} s$ is convex. Note that $\bigcap \mathcal{S} = \bigcap\limits_{s \in \mathcal{S}} s$.

Proof: Let $x, y \in \bigcap \mathcal{S}$, and let $L$ be the line segment from $x$ to $y$. For each $s \in \mathcal{S}$, we have $x, y \in s$ and $s$ is convex, so $L \subseteq s$. So $L \subseteq s$ for all $s \in \mathcal{S}$, so $L \subseteq \bigcap \mathcal{S}$, therefore $\bigcap \mathcal{S}$ is convex.

**Proposition** Halfspaces are convex.

Proof: Let $F = \{x \in \mathbb{R}^n : a^T x \leq b\}$ be a halfspace. Let $x, y \in F$, and $0 \leq \lambda \leq 1$; we want to show that $\lambda x + (1-\lambda)y \in F$, or equivalently that $a^T(\lambda x + (1-\lambda)y) \leq b$.

$$\begin{aligned} a^T(\lambda x + (1-\lambda)y) &= \lambda a^T x + (1-\lambda)a^T y && (a^T x \leq b, a^T y \leq b) \\ &\leq \lambda b + (1-\lambda)b = b \end{aligned}$$

**Corollary** Polyhedra (or equivalently, the feasible region of LPs) are convex.

**Def Properly Contained**: We say that a point $x$ is properly contained in a line segment if it is in the line segment but is distinct from its ends.

A square has 4 extreme points, a semi circle has an infinite amount of extreme points.

**Def Extreme Point**: Given a set $S$, a point $x \in S$ is <u>not</u> an extreme point of $S$ if and only if there exists distinct $x_1, x_2 \in S$, and $0 < \lambda < 1$, for which

$$x = \lambda x_1 + (1-\lambda)x_2$$

Equivalently, a point $x$ is an extreme point in $S$ if it does not lie in the interior of a nontrivial line segment that is contained in $S$.

**Proposition** If $S = \{v \in \mathbb{R}^2 : \|v\| \leq 1\}$, then the extreme points are the points $x$ for which $\|x\| = 1$.

Proof: To prove convexity, we want $(\lambda x_1 + (1-\lambda)x_2)^2 + (\lambda y_1 + (1-\lambda)y_2)^2 \leq 1$.

$$\begin{aligned} &(\lambda x_1 + (1-\lambda)x_2)^2 + (\lambda y_1 + (1-\lambda)y_2)^2 \\ &= \lambda^2 x_1^2 + (1-\lambda)^2 x_2^2 + 2\lambda(1-\lambda)x_1x_2 + \lambda^2 y_1^2 + (1-\lambda)^2 y_2^2 + 2\lambda(1-\lambda)y_1y_2 \\ &= \lambda^2(x_1^2 + y_1^2) + (1-\lambda)^2(x_2^2 + y_2^2) + 2\lambda(1-\lambda)(x_1x_2 + y_1y_2) \\ &\leq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda)(x_1, y_1)^T \cdot (x_2, y_2)^T \text{ Cauchy-Schwarz Inequality} \\ &\leq \lambda^2 + (1-\lambda)^2 + 2\lambda(1-\lambda) \\ &= (\lambda + (1-\lambda))^2 \\ &= 1 \end{aligned}$$

We now show that if $x_1^2 + y_1^2 < 1$, then $(x, y)$ is not an extreme point. To do this, let $(x, y)$ be such a point, and choose $\varepsilon > 0$ small. Let $(x_1, y_1) = (x - \varepsilon, y)$ and $(x_2, y_2) = (x + \varepsilon, y)$.

Since $\varepsilon > 0$, we know $(x_1, y_1) \neq (x_2, y_2)$ and $(x, y) = \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)$ for $\lambda = \frac{1}{2}$. So we need to show $(x_1, y_1), (x_2, y_2) \in S$.

To show $(x_1, y_1) \in S$

$$\begin{aligned} x_1^2 + y_1^2 &= (x-\varepsilon)^2 + y^2 \\ &= x^2 + y^2 - 2\varepsilon x + \varepsilon^2 \\ &\leq x^2 + y^2 + 2\varepsilon\,|x| + \varepsilon && (\varepsilon^2 \leq \varepsilon) \\ &\leq x^2 + y^2 + 3\varepsilon && (|x| \leq 1) \\ &< 1 \end{aligned}$$

To show $(x_2, y_2) \in S$

$$\begin{aligned} x_1^2 + y_1^2 &= (x+\varepsilon)^2 + y^2 \\ &= \cdots \\ &\leq x^2 + y^2 + 3\varepsilon && (|x| \leq 1) \\ &< 1 \end{aligned}$$

So $(x, y)$ not extreme.

Now we need to show that if $x^2 + y^2 = 1$, then $(x, y)$ is extreme in $D$. Suppose not; then there exist distinct $(x_1, y_1)$ and $(x_2, y_2)$ and $0 < \lambda < 1$ s.t. $(x, y) = \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2)$ while $x_1^2 + y_1^2 \leq 1$, $x_2^2 + y_2^2 \leq 1$. So

$$\begin{aligned} 1 &= x^2 + y^2 \\ &= (\lambda x_1 + (1-\lambda)x_2)^2 + (\lambda y_1 + (1-\lambda)y_2)^2 \end{aligned}$$

Since the LHS and RHS of these chains of inequalities are both 1, all inequalities appearing must hold with equality.

- $\lambda^2(x_1^2 + y_1^2) = \lambda^2$, i.e. $x_1^2 + y_1^2 = 1$, $\lambda > 0$.
- $(1-\lambda)^2(x_2^2 + y_2^2) = (1-\lambda)^2$, i.e. $x_2^2 + y_2^2 = 1$, $\lambda < 1$.
- $\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} \cdot \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \|(x_1, y_1)\|\,\|(x_2, y_2)\|$ if $\cos(\theta) = 1$.

These facts imply that $(x_1, y_1), (x_2, y_2)$ are points on the unit circle and one is a positive scalar times the other, so they are the same. This contradicts the choice of them as different.

**Cauchy-Schwarz Inequality** Since $\langle u, v\rangle = \|u\|\,\|v\|\cos(\theta)$, then

$$|\langle u, v\rangle| \leq \|u\|\,\|v\|$$

Ex.

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 2 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \\ 1 & 2 & 3 \end{pmatrix}, b = \begin{pmatrix} 1 \\ 6 \\ 0 \\ 5 \\ 10 \end{pmatrix}, x = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$

then

$$A^= = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 3 & 0 & 1 \end{pmatrix} b^= = \begin{pmatrix} 1 \\ 0 \\ 5 \end{pmatrix}$$

**Theorem** Let $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ be a polyhedron and let $\bar{x} \in P$. Let $A^=$ be the set of rows $A_i$ of $A$ for which the constraint holds with equality at $\bar{x}$ (i.e. $A_i x = b_i$). Let $b^=$ be the corresponding subvector of $b$. So $A^= x = b^=$ is the set of tight constraints for $\bar{x}$. Then $\bar{x}$ is an extreme point of $P$ if and only if $\text{rank}(A^=) = n$.

Proof: Suppose first that $rank(A^=) = n$, and assume that $\bar{x}$ is not an extreme point. So there exist distinct $x_1, x_2 \in P$ and $0 < \lambda < 1$ such that $\bar{x} = \lambda x_1 + (1 - \lambda) x_2$.

So because $0 < \lambda < 1$ and $A^= x_i \leq b^=$ because $x_i \in P$,

$$\begin{aligned} b^= = A^= \bar{x} = A^=(\lambda x_1 + (1-\lambda)x_2) &= \lambda A^= x_1 + (1-\lambda) A^= x_2 \\ &\leq \lambda b^= + (1-\lambda) b^= \\ &= b^= \end{aligned}$$

Equality holds throughout, in particular, $A^= x_1 = b^=$ and $A^= x_2 = b^=$. Therefore, $A^=(x_1 - x_2) = b^= - b^= = 0$. Since $rank(A^=) = n$, the only vector $u \in \mathbb{R}^n$ for which $A^= u = 0$ is $u = 0$. Therefore, $x_1 - x_2 = 0$; thus $x_1 = x_2$, which contradicts their choice as distinct.

Conversely, suppose that $rank(A^=) < n$. This means that the columns of $A^=$ are linearly dependent, so there is a nonzero vector $d \in \mathbb{R}^n$ such that $A^= d = 0$.

Let $\varepsilon > 0$ be small, let $x_1 = \bar{x} - \varepsilon d, x_2 = \bar{x} + \varepsilon d, \lambda = \frac{1}{2}$. Clearly $x_1 \neq x_2$, and $0 < \lambda < 1$, while $\bar{x} = \lambda x_1 + (1 - \lambda) x_2$.

To show that $\bar{x}$ is not an extreme point, it remains to show that $x_1, x_2 \in P$, i.e. that $Ax_1 \leq b, Ax_2 \leq b$.

For each $i$, we have

$$\begin{aligned} A^= x_i &= A^=(\bar{x} + \varepsilon d) \\ &= A^= \bar{x} + \varepsilon A^= d \\ &= b + \varepsilon \cdot 0 = b \end{aligned}$$

So both $x_1, x_2$ satisfy all inequalities in the subsystem $A^= x \leq b^=$ (with equality). For each row $a_i$ of $A$ that is not a row of $A^=$, the inequality $a_i x \leq b_i$ holds strictly at $\bar{x}$; i.e. $a_i \bar{x} < b_i$.

Therefore, $a_i x_1 = a_i(\bar{x} - \varepsilon d) = a_i \bar{x} - \varepsilon a_i d$; since $a_i \bar{x} < b_i$, we have $a_i \bar{x} - \varepsilon a_i d < b_i$ for small enough $\varepsilon$. So $a_i x_1 < b_i$. Similarly, $a_i x_2 < b_i$.

So there is some $\varepsilon > 0$ s.t. $a_i x_1 < b_i$ and $a_i x_2 < b_i$ for every row $a_i$ of $A$ that is not a row of $A^=$. We have shown that $Ax_1 \leq b$ and $Ax_2 \leq b$, as required.

What are the extreme points if $P = \{x \in \mathbb{R}^n : Ax = b, x \geq 0\}$ is given in SEF? Rewrite $P$ to be in the form of the proposition:

$$\{x \in \mathbb{R}^n : Ax \leq b, -Ax \leq -b, -Ix \leq 0\} = \left\{ x \in \mathbb{R}^n : \begin{pmatrix} A \\ -A \\ -I \end{pmatrix} x \leq \begin{pmatrix} b \\ -b \\ 0 \end{pmatrix} \right\}$$

So $\bar{x}$ is extreme in $P$ iff the submatrix $(A, -A, -I)^T$ consisting of the inequalities that are tight for $\bar{x}$ for rank $n$.

The tight inequalities for $\bar{x}$ include all rows of $A$ and $-A$ (because $A\bar{x} = b$), together with the rows of $-I_n$ corresponding to the entries of $\bar{x}$ that are zero. Using these ideas, one can show that $\bar{x}$ is extreme in $P$ iff $\bar{x}$ is a basic feasible solution.

**Theorem** Let $A$ be a matrix where the rows are linearly independent and let $b$ be a vector. Let $P = \{x : Ax = b, x \geq 0\}$ and let $\bar{x} \in P$. Then $\bar{x}$ is an extreme point of $P$ if and only if $\bar{x}$ is a basic feasible solution of $Ax = b$.

Ex. Consider the LP

$$\begin{aligned} \max \quad & 2x_1 + 3x_2 \\ \text{s.t.} \quad & 2x_1 + x_2 \leq 10 \\ & x_1 + x_2 \leq 6 \\ & -x_1 + x_2 \leq 4 \\ & x_1, x_2 \geq 0 \end{aligned}$$

To solve this with simplex, we need slack variables $x_3, x_4, x_5$ such that

$$\begin{aligned} x_3 &= 10 - 2x_1 - x_2 && (C_3) \\ x_4 &= 6 - x_1 - x_2 && (C_4) \\ x_5 &= 4 + x_1 - x_2 && (C_5) \\ x_3, x_4, x_5 &\geq 0 && (C_6) \end{aligned}$$

So each point $(x_1, x_2) \in P$ corresponds to the point $(x_1, \ldots, x_5) \in \hat{P}$ where $\hat{P}$ is obtained from $P$ by adding new constraints above. Points $P$ and $\hat{P}$ are in correspondence and extreme points of $P$ and $\hat{P}$ correspond to each other.

**Theorem** $x$ is an extreme point for the feasible region of $(P)$ if and only if $x$ is an extreme point for the feasible region of $(\hat{P})$.

## 2.7 Simplex Tableau

For an LP:

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

the corresponding LP in tableau form is

$$\left(\begin{array}{c|c} -c^T & z \\ \hline A & b \end{array}\right)$$

Find the first negative term in first row ($R_0$), and make that the entering variable and proceed with the Simplex algorithm. The $z$ in the $R_0$ will be the current objective value.

# Chapter 3

# Duality Theory

Let (P) be the primal LP

$$\begin{aligned} \max \quad & 4x_1 - x_2 + 2x_4 \\ \text{s.t.} \quad & x_1 + x_2 + x_3 + x_4 \leq 7 \\ & 2x_1 - 3x_2 + x_4 \leq 3 \\ & x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 + x_4 \leq 0 \\ & x_1, x_2, x_3, x_4 \geq 0 \end{aligned}$$

$(1, 0, 3, 0)^T$ is feasible for (P) and has objective value 4, so the optimum is $\geq 6$. We argue that the optimum $\leq 10$, by taking a linear combination of the constraints

$$\begin{aligned} & (x_1 + x_2 + x_3 + x_4) \\ & + (2x_1 - 3x_2 + x_4) \\ & + 2(x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 + x_4) \leq 7 + 3 + 2 \cdot 0 = 10 \\ & = 5x_1 - x_2 + 4x_4 \leq 10 \end{aligned}$$

This implies that $4x_1 - x_2 + 2x_4 \leq 5x_1 - x_2 + 4x_4 \leq 10$ for all feasible solutions. So the optimum is $\leq 10$.

To make this work, we needed coefficients $y_1, y_2, y_3 \geq 0$ such that

- $1y_1 + 2y_2 + 1y_3 \geq 4$
- $1y_1 - 3y_2 + \frac{1}{2}y_3 \geq -1$
- $1y_1 + 0y_2 - \frac{1}{2}y_3 \geq 0$
- $y_1 + y_2 + y_3 \geq 2$

Given $y_1, y_2, y_3$ satisfying these constraints, we get the bound

$$\begin{aligned}
4x_1 - x_2 + 2x_4 \leq\ & (1y_1 + 2y_2 + 1y_3)x_1 \\
& + (1y_1 - 3y_2 + \frac{1}{2}y_3)x_2 \\
& + (1y_1 + 0y_2 - \frac{1}{2}y_3)x_3 \\
& + (y_1 + y_2 + y_3)x_4 \\
& = y_1(x_1 + x_2 + x_3 + x_4) \\
& + y_2(2x_1 - 3x_2 + x_4) \\
& + y_3(x_1 + \frac{1}{2}x_2 - \frac{1}{2}x_3 + x_4) \leq 7y_1 + 3y_2 + 0y_3
\end{aligned}$$

The best bound upper bound on the objective function that can be obtained by this technique can be expressed as an LP (P*) as the dual of (P):

$$\begin{aligned}
\min \quad & 7y_1 + 3y_2 + 0y_3 \\
\text{s.t.} \quad & 1y_1 + 2y_2 + 1y_3 \geq 4 \\
& 1y_1 - 3y_2 + \frac{1}{2}y_3 \geq -1 \\
& 1y_1 + 0y_2 - \frac{1}{2}y_3 \geq 0 \\
& y_1 + y_2 + y_3 \geq 2 \\
& y_1, y_2, y_3 \geq 0
\end{aligned}$$

We have constructed (P*) so that every feasible solution $y_1, y_2, y_3$ for (P*) fives an upper bound for the optimum for (P). Note that

- The variables $y_1, y_2, y_3$ of (P*) correspond to constraints in (P)
- The constraints in (P*) correspond to variables $x_1, x_2, x_3, x_4$ in (P)
- The RHS of the $\leq$ constraints in (P) gives the objective function in (P*)
- The objective function in (P) gives the RHS of constraints in (P*)
- The constraint matrix of (P*) is the transpose of that in (P)

## 3.1 Weak Duality

**Def Dual**: We define the dual of a linear program

$$\max\{c^T x : Ax \leq b : x \geq 0\} \implies \min\{b^T y : A^T y \geq c, y \geq 0\}$$

and

$$\min\{c^T x : Ax = b, x \geq 0\} \implies \max\{b^T y : A^T y \leq c, y \text{ free}\}$$

**Theorem Weak Duality** Let (P) and (P*) be a primal-dual pair where (P) is a maximization problem and (P*) a minimization problem. Let $\bar{x}$ and $\bar{y}$ be feasible solutions for (P) and (P*) respectively:

1. Then $c^T\bar{x} \leq b^T\bar{y}$.
2. If $c^T\bar{x} = b^T\bar{y}$, then $\bar{x}$ is an optimal solution to (P) and $\bar{y}$ is an optimal solution to (P*).

Proof:

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

Because $A^T\bar{y} \geq c, \bar{x} \geq 0$ and $A\bar{x} \leq b, \bar{y} \geq 0$. This means that, for every feasible $\bar{x}$ for (P) and $\bar{y}$ for (P*), $c^T\bar{x} \leq b^T\bar{y}$.

If $\bar{x}, \bar{y}$ exist, then $\bar{x}$ implies that (P) is not infeasible, and that fact that $c^T x \leq b^T\bar{y}$ for all feasible $x$ means that (P) is not unbounded. So (P) has an optimum $x^*$.

Now $c^T x^* \leq b^T\bar{y} = c^T\bar{x}$. It follows that $\bar{x}$ is also optimal, because its objective value is at least that of the optimum $x^*$.

**Finding the Dual** In general, an LP (P) is either minimization or maximization, has $\leq, \geq, =$ constraints, and has $\geq 0, \leq 0$ and free variables (unconstrained). We can take the dual of any such LP as follows (P):

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

(P*):

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

Where each ? constraint of (P) gives a variable of (P*) according to

- $\leq$ constraint $\iff$ $\geq 0$ variable.
- $\geq$ constraint $\iff$ $\leq 0$ variable.
- $=$ constraint $\iff$ free variable.

Each variable of (P) gives a constraint of (P*)

- $\geq 0$ variable $\iff$ $\geq$ constraint.
- $\leq 0$ variable $\iff$ $\leq$ constraint.
- free variable $\iff$ $=$ constraint.

Ex. (P)

$$\begin{aligned} \max \quad & 2x_1 - 3x_2 + x_3 + x_4 \\ \text{s.t.} \quad & x_1 + x_3 = 7 \\ & 2x_1 - x_2 + x_4 \leq 10 \\ & x_3 + x_4 \geq 1 \\ & x_1 \geq 0 \\ & x_2 \leq 0 \end{aligned}$$

The dual (P*) is

$$\begin{aligned} \min \quad & 7y_1 + 10y_2 + y_3 \\ \text{s.t.} \quad & 1y_1 + 2y_2 + 0y_3 \geq 2 \\ & 0y_1 - 1y_2 + 0y_3 \leq -3 \\ & 1y_1 + 0y_2 + y_3 = 1 \\ & 0y_1 + y_2 + y_3 = 1 \\ & y_2 \geq 0 \\ & y_3 \leq 0 \end{aligned}$$

(P,P*) is a primal-dual pair.

Ex. Weak duality when (P) is in SEF.

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

(P*)

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

Proof of weak duality: Let $\bar{x}, \bar{y}$ be feasible for P,P* respectively. Then

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

**Corollary** Let (P) and (P*) be a primal-dual pair. Then:

(1) If (P*) is feasible, then (P) is not unbounded.

(2) If (P) is unbounded, then (P*) is infeasible.

(3) If (P*) is unbounded, then (P) is infeasible.

(4) If (P) and (P*) are both feasible, then (P) and (P*) both have optimal solutions.

Proof: We assume (P) is a maximization problem (the min. case is similar). Suppose that (P*) is unbounded, but (P) is not infeasible. So (P) has a feasible solution $\bar{x}$. Since (P*) is unbounded, it has a feasible solution $\bar{y}$ with objective value < the objective value of (P) at $\bar{x}$. This contradicts weak duality.

## 3.2 Strong Duality

**Theorem Strong Duality** Let (P) and (P*) be a primal-dual pair. If there exists an optimal solution $\bar{x}$ of (P), then there exists an optimal solution $\bar{y}$ of (P*) with the same objective value.

Proof: We prove this when (P) is in SEF. Other cases can be reduced to this one. Let $\bar{x}$ be optimal for (P); we many assume that $\bar{x}$ is a basic feasible solution for some basis $B$. The canonical for (P) with respect to $B$ looks like

$$\max\{c_1^T x + \bar{z} : A_B^{-1} A x = A_B^{-1} b, x \geq 0\}$$

where $c_1 \leq 0$, $c_1^T \bar{x} = 0$, $c_1^T = c^T - \bar{y}^T A$ for some $\bar{y}$ ($c_1^T$ is obtained from $c^T$ by adding/subtracting multiples of rows of $A$).

We show that $\bar{y}$ is feasible for (P*) with the same objective value as $\bar{x}$, this will imply by weak duality that $\bar{y}$ is optimal for (P*), we have

$$\begin{aligned} A^T \bar{y} &= (\bar{y}^T A)^T \\ &= (c^T - c_1^T)^T && (c_1^T = c^T - \bar{y}^T A) \\ &= c - c_1 \\ &\geq c && (c_1 \leq 0) \end{aligned}$$

So $\bar{y}$ is feasible for (P*). Now

$$b^T \bar{y} = (A\bar{x})^T \bar{y} = \bar{x}^T (A^T \bar{y}) = \bar{x}^T (\bar{y}^T A)^T = \bar{x}^T (c^T - c_1^T)^T = c^T \bar{x} - c_1^T \bar{x} = c^T \bar{x}$$

**Theorem Strong Duality-Feasibility** Let (P) and (P*) be a primal-dual pair. If (P) and (P*) are both feasible, then there exists an optimal solution $\bar{x}$ of (P) and an optimal solution $\bar{y}$ of (P*) and the objective value of $\bar{x}$ in (P) equals the objective value of $\bar{y}$ in (P*).

## 3.3 Complementary Slackness

Consider the primal-dual pair of LPs in SEF. Let $\bar{x}, \bar{y}$ be feasible for (P) and (P*) respectively. Recall

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

By strong duality, $\bar{x}, \bar{y}$ are optimal if and only if $c^T \bar{x} = b^T \bar{y}$, i.e. equality holds. This holds if $c^T \bar{x} = (A^T \bar{y})^T \bar{x}$ or $\langle c, \bar{x} \rangle = \left\langle A^T \bar{y}, \bar{x} \right\rangle$.

Therefore, $\langle c, \bar{x} \rangle = \left\langle A^T \bar{y}, \bar{x} \right\rangle$ if and only if, for each $i$, either

- the $i$th entry of $\bar{x}$ is 0, or
- the $i$th entry of $c$ is equal to the $i$th entry of $A^T \bar{y}$.

By this reasoning, if $\bar{x}, \bar{y}$ are feasible for (P) and (P*) respectively, then $\bar{x}, \bar{y}$ are optimal if and only if for each $i$, either

- the $i$th entry of $\bar{x}$ is 0, or
- the $i$th entry inequality in the system $A^T y \geq c$ holds with equality at $\bar{y}$.

These are the complementary slackness (CS) conditions for (P) and (P*).

**Theorem Complementary Slackness - Special case** Let $\bar{x}$ be a feasible solution to $\max\{c^T x : Ax \le b, x \ge 0\}$ and let $\bar{y}$ be a feasible solution to the dual $\min\{b^T y : A^T y = c, y \ge 0\}$. Then $\bar{x}$ is an optimal solution to the primal and $\bar{y}$ is an optimal solution to the dual if and only if for every row index $i$ of $A$, constraint $i$ of the primal is tight for $\bar{x}$ or the corresponding dual variable $\bar{y}_i = 0$.

**Theorem Complementary Slackness** Let (P) and (P*) be an arbitrary primal-dual pair. Let $\bar{x}$ be a feasible solution to (P*) and let $\bar{y}$ be a feasible solution to (P*). Then $\bar{x}$ is an optimal solution to (P) and $\bar{y}$ is an optimal solution to (P*) if and only if the complementary slackness conditions hold.

- For every variable $\bar{x}_j$ of $(P_{\max})$, $\bar{x}_j = 0$ or the corresponding constraint $j$ of $(P_{\min})$ is satisfied with equality.
- For every variable $\bar{y}_i$ of $(P_{\min})$, $\bar{y}_i = 0$ or the corresponding constraint $i$ of $(P_{\max})$ is satisfied with equality.

Ex. (P)

$$\begin{array}{ll} \max & 5x_1 - 3x_2 - x_4 \\ \text{s.t.} & \begin{pmatrix} 1 & 1 & -1 & -1 \\ 4 & 2 & 1 & 2 \\ -1 & 4 & 1 & 3 \end{pmatrix} x \begin{array}{c} \le \\ = \\ \ge \end{array} \begin{pmatrix} 5 \\ 1 \\ -4 \end{pmatrix} \\ & x_1, x_2 \ge 0, x_3 \le 0, x_4 \text{ free} \end{array}$$

(P*)

$$\begin{array}{ll} \min & 5y_1 + y_2 - 4y_3 \\ \text{s.t.} & \begin{pmatrix} 1 & 4 & -1 \\ 1 & 2 & 4 \\ -1 & 1 & 1 \\ -1 & 2 & 3 \end{pmatrix} y \begin{array}{c} \ge \\ \ge \\ \le \\ = \end{array} \begin{pmatrix} 5 \\ -3 \\ 0 \\ -1 \end{pmatrix} \\ & y_1 \ge 0, y_2 \text{ free}, y_3 \le 0 \end{array}$$

We argue using complementary slackness that $\bar{x} = (1, 0, -3, 0)^T$ and $\bar{y} = (0, 1, -1)^T$ are optimal for (P) and (P*).

Note that $A\bar{x} = (4, 1, -4)^T, A^T\bar{y} = (5, -2, 0, -1)^T$. It is easy to check $\bar{x}, \bar{y}$ are feasible.

$\bar{x}_i \ne 0$ for $i \in \{1, 3\}$. Since $A^T\bar{y} = (5, -2, 0, -1)^T$ holds with equality in positions 1 and 3, CS holds for $\bar{x}$. Similarly $\bar{y}_j \ne 0$ for $j \in \{2, 3\}$. Since $A\bar{x} = (4, 1, -4)^T$ holds with equality in positions 2 and 3, so CS holds for $\bar{y}$. So by CS, $\bar{x}, \bar{y}$ are optimal.

## 3.4 Geometry of Optimality

Let $P = \left\{ x \in \mathbb{R}^2 : \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} x \le \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \right\}$ and let $\bar{x} = (1, 3)^T$. For which vector $c \in \mathbb{R}^2$ is $\bar{x}$ optimal for the LP $\max\{c^T x : x \in P\}$? $\bar{x}$ is optimal if $c = (1, 1)$ or $c = (1, 4)$ or $c = (0, 1)$. In general, if $c$ lies between the line spanned by $(1,1)$ and the line spanned by $(0,1)$, then $\bar{x}$ is optimal for $c$, otherwise it is not.

**Def Cone**: Let $a_1, \ldots, a_k$ be a set of vectors in $\mathbb{R}^n$. We define the cone generated by $a_1, \ldots, a_k$ to be the set

$$C = \text{cone}\{a_1, \ldots, a_k\} = \left\{ \sum_{i=1}^{k} \lambda_i a_i : \lambda_i \geq 0 \right\}$$

Ex. $cone \left\{ \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} : x_1 \geq 0, x_1 + x_2 \geq 1 \right\}$

Proof: (P)

$$\begin{array}{ll} \max & c^T x \\ \text{s.t.} & \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} x \leq \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} \end{array}$$

(P*)

$$\begin{array}{ll} \min & (2,3,4)y \\ \text{s.t.} & \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} y = c \\ & y \geq 0 \end{array}$$

Suppose $\bar{x} = (1,3)^T$ is optimal. By strong duality, (P*) has an optimal solution $\bar{y}$. By CS, for each inequality in $A\bar{x} \leq (2,3,4)^T$ which is not tight, the corresponding $\bar{y}_i$ is 0.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix} \bar{x} = \begin{pmatrix} 1 \\ 3 \\ 4 \end{pmatrix} \begin{matrix} < \\ = \\ = \end{matrix} \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix}$$

so $\bar{y}_1 = 0$ (constraint 1 of (P) is not tight).

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix} \bar{y} = c$$

this means that $\bar{y}_2 \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \bar{y}_3 \begin{pmatrix} 1 \\ 1 \end{pmatrix} = c$ for $\bar{y}_2, \bar{y}_3 \geq 0$. This is equivalent to $c \in cone((0,1)^T, (1,1)^T)$. Reversing this reasoning gives $\bar{x}$ is optimal for (P) iff $c \in cone((0,1)^T, (1,1)^T)$

**Theorem** Let $\bar{x}$ be a feasible solution to $\max\{c^T x : Ax \leq b\}$. Then $\bar{x}$ is optimal if and only if $c \in \text{cone}(a_1, \ldots, a_k)$, where $a_1, \ldots, a_k$ are the rows of $A$ corresponding to inequalities in $Ax \leq b$ which are tight at $\bar{x}$.

Proof: The dual to (P) is (P*) $\min\{b^T y : A^T y = c, y \geq 0\}$. By strong duality (P*) has an optimum $\bar{y}$. By CS, $\bar{x}$ is optimal if and only if for all constraints $a_i \bar{x} \leq b_i$ that are not tight at $\bar{x}$, the corresponding entry of $\bar{y}$ is 0. We have $A^T\bar{y} = c, \bar{y} \geq 0, \bar{y}_i = 0$ for all $a_i\bar{x} < b_i \iff$

$$c = \sum_{\substack{a_i^T \text{ row of } A \\ \text{tight for } \bar{x}}} y_i a_i, y_i \geq 0, \forall i$$

$\iff c$ lies in the cone of tight constraints.

## 3.5 Farkas' Lemma

Suppose that $Ax = b$ has no solution. Then $\begin{pmatrix} A & | & b \end{pmatrix} \sim \begin{pmatrix} \cdots & | & \cdots \\ \mathbf{0} & | & s \end{pmatrix}$ for some $s \neq 0$. The row $(\mathbf{0}|s)$ is a linear combination of the rows of $(A|b)$. This implies that $\exists y$ such that $y^T(A|b) = (\mathbf{0}|s)$ which implies that $\exists y$ such that $y^T A = 0$ and $y^T b \neq 0$.

Conversely, if $\exists y$ such that $y^T A = 0$ and $y^T b \neq 0$, then $Ax = b$.

$$\begin{aligned} y^T Ax &= y^T b \\ (0)x &= y^T b \\ 0 &\neq 0 \end{aligned}$$

**Lemma** Given $A$ and $b$, exactly one of the following is true

(i) $Ax = b$ has a solution

(ii) there exists $y$ such that $y^T A = 0$ and $y^T b \neq 0$ ($y$ is a certificate)

**Theorem Farkas' Lemma** Let $A$ be an $m \times n$ matrix and let $b$ be a vector with $m$ entries. Then exactly one of the following statements is true:

(1) The system $Ax = b, x \geq 0$ has a solution.

(2) There exists a vector $y$ such that $y^T A \geq 0$ ($A^T y \geq 0$) and $y^T b < 0$ ($b^T y < 0$).

Proof: We first show that (1) and (2) cannot occur simultaneously. If (2) holds for $\bar{y}$, and (1) holds for $\bar{x}$, then

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

This is a contradiction since $\bar{y}^T A \geq 0$ and $\bar{x} \geq 0$.

Suppose that $Ax = b, x \geq 0$ has no solution. Consider the LP (P)

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

is infeasible and (P*)

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

Note that (P*) is feasible. Therefore, (P*) is unbounded.

In particular, there is a feasible solution $\bar{y}$ to (P*) with objective value $b^T y < 0$. Now $A^T \bar{y} \geq 0$ by feasibility. and $b^T \bar{y} < 0$ by choice of $\bar{y}$. Then $\bar{y}^T A \geq 0, \bar{y}^T b < 0$ as required.

It is reasonable to want to construct $y$ in a given case, rather than to just prove $y$ exists.

Consider applying two-phase simplex to (P). The auxiliary problem is (P')

$$\begin{aligned} \max \quad & 0x_1 + \cdots + 0x_n - x_{n+1} - \cdots - x_{n+m} \\ \text{s.t.} \quad & \begin{pmatrix} A & \mid & I_m \end{pmatrix} x = b \\ & x \geq 0 \end{aligned}$$

Running simplex on (P') until an optimum is found gives an LP (P") with objective function

$$d = (0, \ldots, 0, -1, \ldots, -1)^T \implies d^T x - \bar{y}^T (Ax - b)$$

for some $\bar{y}$ where $(d^T - \bar{y}^T (A|I)) \leq 0$ because simplex terminated and $d^T x - \bar{y}^T ((A|I)x - b) < 0$ because by assumption, (P) was infeasible and so (P') had optimum $< 0$.

- $d^T - \bar{y}^T (A|I) \leq 0$
- $d^T x - \bar{y}^T ((A|I)x - b) < 0$
- $x \geq 0$

The first condition implies that $\bar{y}^T (A|I) \geq d^T$ so $\bar{y}^T A \geq 0$.

The second condition implies that $\bar{y}^T b < \bar{y}^T (A|I)x - d^T x = (y^T (A|I) - d^T)x$

**Outcomes to Linear Programs**

| P*/P | Optimal | Unbounded | Infeasible |
|---|---|---|---|
| Optimal | Yes | No | No |
| Unbounded | No | No | Yes |
| Infeasible | No | Yes | Yes |

**Matching IP Duality** Variables $x_e$ for each edge $e$.

$$\begin{aligned} \max \quad & \sum_{e \in E} x_e \\ \text{s.t.} \quad & \sum_{e \in \delta(v)} x_e \leq 1 : \forall v \in V \\ & x_e \geq 0 \\ & x_e \in \{0, 1\} \end{aligned}$$

The dual

$$
\begin{aligned}
\max \quad & \sum_{v \in V} y_v \\
\text{s.t.} \quad & y_u + y_v \geq 1 : \forall e = (u, v) \in E \\
& y \geq 0 \\
& y \in \{0, 1\}
\end{aligned}
$$

The dual gets you the vertex cover.

# Chapter 4

# Solving Integer Programs

## 4.1 Integer Programs vs. Linear Programs

**3-SAT**: Each constraint looks like $x_i + x_j + x_k \geq 1$ or $(1 - x_i) + x_j + x_k \geq 1$, etc.

$$\begin{aligned} \max \quad & \sum_{i=1}^{n} x_i \\ \text{s.t.} \quad & (1 - x_i) + x_j + x_k \geq 1 \\ & \vdots \\ & x_1, \ldots, x_n \in \{0, 1\} \end{aligned}$$

**Def Convex Hull**: Consider a set $S$ of points in $\mathbb{R}^n$. The convex hull of $S$, denoted $\operatorname{conv}(S)$, is defined as the smallest convex set that contains $S$. More precisely, it is the intersection of all convex sets in $\mathbb{R}^n$ that contain $S$.
$S = \operatorname{conv}(S)$ if and only if $S$ is a convex set.

**Fundamental Theorem of Integer Programming** Consider the following polyhedron $P = \{x \in \mathbb{R}^n : Ax \leq b\}$ where $A$ and $b$ have rational entries. Let $S$ be the set of all integer points in $P$. Then $\operatorname{conv}(S)$ is a polyhedron $Q = \{x \in \mathbb{R}^n : A'x \leq b'\}$ where $A'$ and $b'$ have rational entries.

**Theorem** Consider the IP $\max\{c^T x : Ax \leq b, x \text{ integer}\}$ and the LP $\max\{c^T x : A'x \leq b'\}$. Then the following hold

1. IP is infeasible if and only if LP is infeasible.
2. IP is unbounded if and only if LP is unbounded.
3. Every optimal solution to IP is an optimal solution to LP.
4. Every optimal solution to LP that is an extreme point is an optimal solution to IP.

**Strategy for Solving an IP** Let the IP be $\max\{c^T x : Ax \leq b, x \text{ integer}\}$.

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

- Let $S = \{\text{integer points in } P\}$
- Find the polyhedron $Q = \text{conv}(S)$ given by the fundamental theorem of integer programming
- Construct LP $\max\{c^T x : x \in Q\}$ and solve LP with simplex to solve the original IP

## 4.2 Cutting Planes

**Def Cutting Plane**: Suppose we have an optimal solution $\bar{x}$ to the LP relaxation. We want to find a constraint $\alpha^T x \leq \beta$ such that

- every feasible solution of the IP satisfies the constraint.
- $\bar{x}$ does not satisfy the constraint.

Consider the optimization problems (P1) $\max\{c^T x : x \in S_1\}$ and (P2) $\max\{c^T x : x \in S_2\}$.

**Def Relaxation**: If $S_2 \subseteq S_1$, then we say that (P2) is a relaxation of (P1).

**Proposition** Suppose (P2) is a relaxation of (P1), then

- If (P2) is infeasible, (P1) is infeasible.
- If $\bar{x}$ is optimal for (P2) and $\bar{x}$ is feasible for (P1), then $\bar{x}$ is optimal for (P1).
- If $\bar{x}$ is an optimal solution for (P2), then $c^T \bar{x}$ is an upper bound for (P1).

**Algorithm 2** Cutting Plane Algorithm

$$\max\{c^T x : Ax \leq b, x \geq 0, x \in \mathbb{Z}\}$$

**while** true **do**
    Let the (LP) denote $\max\{c^T x : Ax \leq b\}$
    **if** (LP) is infeasible **then**
        **Stop** (IP) is infeasible
    Let $\bar{x}$ be optimal solution to (LP)
    **if** $\bar{x}$ is integral **then**
        **Stop** $\bar{x}$ is an optimal solution to (IP)
    Find a cutting plane $a^T x \leq \beta$ for $\bar{x}$
    Add constraint $a^T x \leq \beta$ to the system $Ax \leq b$

**Finding the Cutting Plane**: Solve the LP relaxation and get the LP in a canonical form for $B$:

$$\begin{aligned} \max \quad & \bar{c}^T x + \bar{z} \\ \text{s.t.} \quad & x_B + A_N x_N = b \\ & x \geq 0 \end{aligned}$$

If $\bar{x}$ is not integer, then $b_i$ is fractional for some value $i$. We know that every feasible solution to the LP relaxation satisfies

$$x_{r(i)} + \sum_{j \in N} A_{ij} x_j = b_i \implies x_{r(i)} + \underbrace{\sum_{j \in N} \lfloor A_{ij} \rfloor x_j}_{\text{integer for all } x \text{ integer}} \leq b_i$$

Hence, every feasible solution to IP satisfies

$$x_{r(i)} + \sum_{j \in N} \lfloor A_{ij} \rfloor x_j \leq \lfloor b_i \rfloor$$

which is the cutting plane for $\bar{x}$ since $\bar{x}$ does not satisfy the above constraint.

# Chapter 5

# Nonlinear Optimization

**Fermat's Last Theorem** There do not exist integers $x, y, z \geq 1$ and integer $n \geq 3$ such that $x^n + y^n = z^n$.

## 5.1 Convexity

**Def Locally Optimal**: We say that $\bar{x}$ is locally optimal if for some positive integer $d \in \mathbb{R}$, we have that $f(\bar{x}) \leq f(x)$ for every $x \in S$ where $\|x - \bar{x}\| \leq d$.

**Def Convex Function**: We say that the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if for every pair of points $x, y \in \mathbb{R}^n$ and for every $\lambda \in [0, 1]$

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

Geometrically, $f$ is convex iff for all $x_1, x_2 \in \mathbb{R}^n$, the line segment from $(x_1, f(x_1))$ to $(x_2, f(x_2))$ lies above the graph of $f$.

Ex. Prove $f(x) = x^2$ is convex.

$$\begin{aligned} f(\lambda x + (1 - \lambda)y) &= (\lambda x + (1 - \lambda)y)^2 \\ &= \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 \end{aligned}$$

So we want

$$\begin{aligned} \lambda^2 x^2 + 2\lambda(1 - \lambda)xy + (1 - \lambda)^2 y^2 &\leq \lambda x^2 + (1 - \lambda)y^2 \\ \Leftrightarrow (\lambda^2 - \lambda)x^2 + 2\lambda(1 - \lambda)xy + ((1 - \lambda)^2 - (1 - \lambda))y^2 &\leq 0 \\ \Leftrightarrow \lambda(\lambda - 1)(x^2 - 2xy + y^2) &\leq 0 \end{aligned}$$

This is true because $\lambda \geq 0, \lambda - 1 \leq 0, x^2 - 2xy + y^2 = (x - y)^2 \geq 0$.

**Proposition** If $f : \mathbb{R} \to \mathbb{R}$ is twice-differentiable, then $f$ is convex if and only if $f''(x) \geq 0$ for all $x \in \mathbb{R}$.

Converse proof: Suppose $f$ is not convex; we argue that there is some $x$ such that $f''(x) < 0$. If $f$ is not convex, then $f(c) > \lambda f(a) + (1-\lambda)f(b)$. Use this and the mean value theorem to argue that there exist $a_0 \in [a, c]$ and $b_0 \in [c, b]$ such that $f'(a_0) > f'(b_0)$. It follows by Mean Value Theorem, there is a point $d$ between $a_0$ and $b_0$ such that $f''(d) < 0$.

**Def Epigraph**: Given $f : \mathbb{R}^n \to \mathbb{R}$, define the epigraph of $f$ as

$$\text{epi}(f) = \left\{ \begin{pmatrix} \mu \\ x \end{pmatrix} \in \mathbb{R} \times \mathbb{R}^n : f(x) \leq \mu \right\}$$

This is the set of all points on or above the graph of $f$.

**Proposition** Let $f : \mathbb{R}^n \to \mathbb{R}$, then $f$ is convex if and only if epi$(f)$ is a convex set.

Proof: Suppose that $f$ is convex. Let $\begin{pmatrix} \mu_1 \\ x_1 \end{pmatrix}, \begin{pmatrix} \mu_2 \\ x_2 \end{pmatrix} \in \text{epi}(f)$, and $\lambda \in [0, 1]$. We want to show that

$$\lambda \begin{pmatrix} \mu_1 \\ x_1 \end{pmatrix} + (1-\lambda) \begin{pmatrix} \mu_2 \\ x_2 \end{pmatrix} = \begin{pmatrix} \lambda\mu_1 + (1-\lambda)\mu_2 \\ \lambda x_1 + (1-\lambda)x_2 \end{pmatrix}$$

satisfies $\lambda\mu_1 + (1-\lambda)\mu_2 \geq f(\lambda x_1 + (1-\lambda)x_2)$. Since $f$ is convex, we have $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2) \leq \lambda\mu_1 + (1-\lambda)\mu_2$ because $(\mu_1, x_1)^T, (\mu_2, x_2)^T \in \text{epi}(f)$ and $\lambda, 1-\lambda \geq 0$.

Conversely, suppose that epi$(f)$ is a convex set. Let $x_1, x_2 \in \mathbb{R}^n$ and $\lambda \in [0, 1]$. We want to show that $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$. Note that $(f(x_1), x_1)^T, (f(x_2), x_2) \in \text{epi}(f)$ because $f(x_1) \geq f(x_1), f(x_2) \geq f(x_2)$. By convexity of epi$(f)$, it follows that $\lambda(f(x_1), x_1)^T + (1-\lambda)(f(x_2), x_2) \in \text{epi}(f)$ so

$$\begin{pmatrix} \lambda f(x_1) + (1-\lambda)f(x_2) \\ \lambda x_1 + (1-\lambda)x_2 \end{pmatrix} \in \text{epi}(f)$$

i.e. $f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda)f(x_2)$ by definition of epigraph.

**Def Level Set**: Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\beta \in \mathbb{R}$. We call the set

$$\{x \in \mathbb{R}^n : g(x) \leq \beta\}$$

a level set of the function $g$.

**Proposition** The level set of a convex function is a convex set.

Proof: Let $x_1, x_2 \in \{x \in \mathbb{R}^n : f(x) \leq \beta\}$ and let $0 \leq \lambda \leq 1$. We want to show that $\lambda x_1 + (1-\lambda)x_2 \in \{x \in \mathbb{R}^n : f(x) \leq \beta\}$. We have $f(x_1) \leq \beta, f(x_2) \leq \beta$, and we want $f(\lambda x_1 + (1-\lambda)x_2) \leq \beta$. By convexity of $f$,

$$\begin{aligned} f(\lambda x_1 + (1-\lambda)x_2) &\leq \lambda f(x_1) + (1-\lambda)f(x_2) \\ &\leq \lambda\beta + (1-\lambda)\beta \\ &= \beta \end{aligned}$$

because $\lambda, 1-\lambda \geq 0$.

**Lemma** If $g_1, \ldots, g_k$ are convex functions, then the function $h(x) = \max(g_1(x), \ldots, g_k(x))$ is convex.

Proof: Let $x_1, x_2 \in \mathbb{R}^n$ and $0 \leq \lambda \leq 1$. Then

$$\begin{aligned}
h(\lambda x_1 + (1-\lambda)x_2) &= \max_{i\in[k]} g_i(\lambda x_1 + (1-\lambda)x_2) \\
&\leq \max_{i\in[k]} \lambda g_i(x_1) + (1-\lambda)g_i(x_2) \\
&\leq \max_{i\in[k]}(\lambda g_i(x_1)) + \max_{i\in[k]}((1-\lambda)g_i(x_2)) \\
&= \lambda h(x_1) + (1-\lambda)h(x_2)
\end{aligned}$$

because $\lambda \geq 0, 1-\lambda \geq 0$, and the definition of $h$.

**Proposition** The feasible region of a convex NLP is a convex set.

## 5.2 Relaxing Convex NLPs

**Def Subgradient**: Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\bar{x} \in \mathbb{R}^n$. We say that $s \in \mathbb{R}^n$ is a subgradient of $f$ at $\bar{x}$ if for every $x \in \mathbb{R}^n$,

$$f(\bar{x}) + s^T(x - \bar{x}) \leq f(x)$$

Note that since $f(\bar{x})$ and $\bar{x}$ are constant, then the LHS is equal to $s^T x + (f(\bar{x}) - s^T\bar{x})$, which is an affine expression in $x$.

Ex. We want to compute the subgradients of $f(x) = \begin{cases} 1 & x < 2 \\ x-1 & x \geq 2 \end{cases}$ at $\bar{x} = 2$. We want to find all $s \in \mathbb{R}$ such that $f(2) + s(x-2) \leq f(x)$ for all $x \in \mathbb{R}$, i.e. we need $f(x) \geq sx + (1-2s)$ for all $x$. We prove that $s$ is a subgradient if and only if $0 \leq s \leq 1$.

$$\begin{aligned}
f(x) &\geq sx + 1 - 2s & \forall x \\
\Leftrightarrow 1 &\geq sx + (1-2x) & \forall x \leq 2 \\
x - 1 &\geq sx + (1-2x) & \forall x \geq 2
\end{aligned}$$

Suppose that $0 \leq s \leq 1$. Then $1-2s+sx = 1-s(2-x) \leq 1$ for all $x \leq 2$ and $sx+(1-2s)-(x-1) = s(x-2)+2-x = (s-1)(x-2) \leq 0$ since $s-1 \leq 0, x-2 \geq 0$, so $x-1 \geq sx+(1-2s)$.
If $s < 0$, then $s \cdot 0 + 1 - 2s > 1$, so $s$ is not a subgradient.
If $s > 1$, then $3s + 1 - 2s = s + 1 \geq 3 - 1$ so $s$ is not a subgradient.

This corresponds to the geometric fact that the lines through $(2, f(2))$ that lie below the graph of $f$ are precisely those with gradient between 0 and 1.

**Def Supporting Halfspace**: Consider a convex set $C \subseteq \mathbb{R}^n$ and let $\bar{x} \in C$. We say that the halfspace $F = \{x \in \mathbb{R}^n : s^T x \leq \beta, s \in \mathbb{R}^n, \beta \in \mathbb{R}\}$ is a supporting halfspace of $C$ at $\bar{x}$ if

(1) $C \subseteq F$

(2) $s^T\bar{x} = \beta$, i.e. $\bar{x}$ is on the hyperplane that defines the boundary of $F$

**Proposition** Let $g : \mathbb{R}^n \to \mathbb{R}$ be a convex function, let $\bar{x} \in \mathbb{R}^n$ such that $g(\bar{x}) = 0$, and let $s \in \mathbb{R}^n$ be a subgradient of $f$ at $\bar{x}$. Denote the level set $C = \{x \in \mathbb{R}^n : g(x) \leq 0\}$ and halfspace $F = \{x \in \mathbb{R}^n : g(\bar{x}) + s^T(x - \bar{x}) \leq 0\}$. Then $F$ is a supporting halfspace of $C$ at $\bar{x}$.

Proof: We can rewrite $F$ as $\{x \in \mathbb{R}^n : s^T x \leq s^T \bar{x}\}$ because $g(\bar{x}) = 0$. So clearly $s^T \bar{x} \leq s^T \bar{x}$ so $\bar{x}$ is on the boundary of $F$ which meets the second condition of supporting halfspace. We need to show $\{x \in \mathbb{R}^n : g(x) \leq 0\} \subseteq \{x \in \mathbb{R}^n : s^T(x - \bar{x}) \leq 0\}$ since $s$ is a subgradient, we have $0 + s^T(x - \bar{x}) \leq g(x)$ for all $x$. So if $g(x) \leq 0$, then $s^T(x - \bar{x}) = 0$, i.e. $C \subseteq F$.

**Corollary** Consider an NLP. Let $\bar{x}$ be a feasible solution and suppose that constraint $g_i(x) \leq 0$ is tight for some $i \in \{1, \dots, m\}$. Suppose $g_i$ is a convex function that has a subgradient $s$ at $\bar{x}$. Then the NLP obtained by replacing constraint $g_i(x) \leq 0$ by the linear constraint $s^T x \leq s^T \bar{x} - g_i(\bar{x})$ is a relaxation of the original NLP.
If $g_i(x) \leq 0$ is tight at $\bar{x}$, i.e. $g_i(\bar{x}) = 0$, then we have the linear constraint

$$s^T(x - \bar{x}) \leq 0$$

## 5.3 Karush-Kuhn-Tucker Theorem

Consider the NLP

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{aligned}$$

and the relaxation of the NLP and $J(\bar{x}) = \{i : g_i(\bar{x}) = 0\}$

$$\begin{aligned} \max \quad & -c^T x \\ \text{s.t.} \quad & s_i^T(x) \leq s_i^T \bar{x} - g(\bar{x}) : \forall i \in J(\bar{x}) \end{aligned}$$

Ex.

$$\begin{aligned} \min \quad & -x_1 - 2x_2 \\ \text{s.t.} \quad & |x_1 + x_2| - 4 \leq 0 \\ & x_2^2 - 9 \leq 0 \\ & x_1^2 + x_2 - 6 \leq 0 \end{aligned}$$

We show that $\bar{x} = (1, 3)^T$ is optimal.

$s_1 = (1, 1)$ is a subgradient for $g_1(x) = |x_1 + x_4| - 4$ at $(1, 3)$. We need that $g_1(x) \geq g_1(1, 3) + (1, 1)\begin{pmatrix} x_1 - 1 \\ x_2 - 3 \end{pmatrix}$ for all $(x_1, x_2) \in \mathbb{R}^2$.

$$\begin{aligned} g_1(1, 3) + (1, 1)\begin{pmatrix} x_1 - 1 \\ x_2 - 3 \end{pmatrix} &= |1 + 3| - 4 + x_1 + x_2 - 4 \\ &= x_1 + x_2 - 4 \\ &\leq |x_1 + x_2| - 4 = g_1(x) \end{aligned}$$

$s_2 = (0,2)$ is a subgradient for $g_2(x) = x_2^2 - 9$ at $(1,3)$. We need $g_2(1,3) + (0,6)\begin{pmatrix} x_1 - 1 \\ x_2 - 3 \end{pmatrix} \leq g_2(x)$ for all $x$, i.e.

$$\begin{aligned} (3^2 - 9) + 6x_2 - 18 &\leq x_2^2 - 9 \\ x_2^2 - 6x_2 + 9 &\geq 0 \\ (x_2 - 3)^2 &\geq 0 \end{aligned}$$

which is always true.

We now use these two subgradients to replace the first two constraints with linear constraints, and we drop the third constraint because $\bar{x} = (1,3)$ is not tight. This gives an LP (P)

$$\begin{aligned} \min \quad & -x_1 - 2x_2 \\ \text{s.t.} \quad & s_1^T(x - \bar{x}) \leq 0 \\ & s_2^T(x - \bar{x}) \leq 0 \end{aligned}$$

or equivalently,

$$\begin{aligned} \max \quad & x_1 + 2x_2 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq 4 \\ & \begin{pmatrix} 0 & 6 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \leq 18 \end{aligned}$$

Since (P) is a relaxation of the NLP, if we can prove that $\bar{x} = (1,3)$ is optimal for (P), then it is optimal for the NLP. By the lemma that $\bar{x}$ is optimal iff $c \in \text{cone}(a_{i_1}, \ldots, a_{i_t})$ where those are the tight constraints. By this lemma, $\bar{x} = (1,3)$ is optimal if and only if $(1,2) \in \text{cone}((1,1),(0,6))$ if and only if $\exists \beta_1, \beta_2 \geq 0$ such that $(1,2) = \beta_1(1,1) + \beta(0,6)$. This holds for $\beta_1 = 1, \beta_2 = \frac{1}{6}$. So $\bar{x}$ is optimal for (P) and is optimal for the NLP because $\bar{x}$ is feasible for the NLP and (P) is a relaxation of the NLP.

**Proposition** Consider the NLP above and assume $g_1, \ldots, g_m$ are convex functions. Let $\bar{x}$ be a feasible solution and suppose that for all $i \in J(\bar{x})$ we have a subgradient $s_i$ at $\bar{x}$. If $-c \in \text{cone}\{s_i : i \in J(\bar{x})\}$, then $\bar{x}$ is an optimal solution.

**Changing Nonlinear Objective Function to Linear**

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{aligned}$$

is equivalent to

$$\begin{aligned} \min \quad & z \\ \text{s.t.} \quad & f(x) - z \leq 0 \\ & g_1(x) \leq 0 \\ & \vdots \\ & g_m(x) \leq 0 \end{aligned}$$

Optimal solutions $(x, z)$ for the new NLP have the form $(\bar{x}, f(\bar{x}))$.

**Def Differentiable and Gradient**: Let $f : \mathbb{R}^n \to \mathbb{R}$ and $\bar{x} \in \mathbb{R}^n$ be given. If there exists $s \in \mathbb{R}^n$ such that

$$\lim_{h \to 0} \frac{f(\bar{x} + h) - f(\bar{x}) - s^T h}{\|h\|} = 0$$

we say that $f$ is differentiable at $\bar{x}$ and call the vector $s$ the gradient of $f$ at $\bar{x}$. We denote $s$ by $\nabla f(\bar{x})$.

$$\nabla f(x) = \left[\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n}\right]^T$$

**Proposition** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function and let $\bar{x} \in \mathbb{R}^n$. If the gradient $\nabla f(\bar{x})$ exists then it is a subgradient of $f$ at $\bar{x}$.

**Corollary** Let $\min\{c^T x : g_1(x) \leq 0, \ldots, g_m(x) \leq 0\}$ be a convex NLP and let $\bar{x}$ be feasible. Let $J = \{i : g_i(\bar{x}) = 0\}$. Suppose that for all $j \in J$, the function $g_j$ is differentiable with continuous first derivative at $\bar{x}$. If $-c \in \text{cone}(\nabla g_j(\bar{x}) : j \in J)$, then $\bar{x}$ is optimal.

**Def Slater Point**: An NLP has a Slater point if there exists $x'$ such that $g_i(x') < 0$ for every $i \in \{1, \ldots, m\}$.

**Karush-Kuhn-Tucker Theorem (KKT)** (KKT based on gradients) Consider a convex NLP that has a Slater point. Let $\bar{x} \in \mathbb{R}^n$ be a feasible solution and assume that $f, g_1, \ldots, g_m$ are differentiable at $\bar{x}$.

$$\bar{x} \text{ is optimal} \iff -\nabla f(\bar{x}) \in \text{cone}\, \nabla g_i(\bar{x}) : i \in J(\bar{x})$$

Ex. Consider the NLP

$$\begin{aligned} \min \quad & -7x_1 - 5x_2 \\ \text{s.t.} \quad & 2x_1^2 + x_2^2 + x_1 x_2 - 4 \leq 0 \\ & x_1^2 + x_2^2 - 2 \leq 0 \\ & -x_1 + \frac{1}{2} \leq 0 \end{aligned}$$

Prove that $\bar{x} = (1, 1)^T$ is optimal.

KKT requires that

- there exists $x'$ that strictly satisfies all constraints
- $-\nabla f(\bar{x}) \in \text{cone}(\nabla g_i(\bar{x}) : i \in J(\bar{x}))$ (tight constraints at $\bar{x}$)

Each $g_i$ is differentiable since they are all polynomial. To find a Slater point, use inspection. Guess $x_1 = 1$, $2 + x_2^2 + x_2 - 4 < 0, x_2^1 - 1 < 0, -\frac{1}{2} < 0$. $x_2 = 0$ satisfies all these. So $(1, 0)$ is a strictly feasible point.

The tight constraints at $\bar{x}$ are the first two: $g_1(\bar{x}) = 0$ and $g_2(\bar{x}) = 0$. We have

$$\nabla f(x) = (-7, -5), \nabla g_1(x) = (4x_1 + x_2, 2x_2 + x_1), \nabla g_2(x) = (2x_1, 2x_2)$$

at $\bar{x} = (1, 1)$,

$$\nabla f(\bar{x}) = (-7, -5), \nabla g_1(\bar{x}) = (5, 3), \nabla g_2(\bar{x}) = (2, 2)$$

So by KKT, $\bar{x}$ is optimal if and only if $(7,5) \in \operatorname{cone}((5,3),(2,2))$. Note that $(7,5) = 1(5,3)+1(2,2)$, so $\bar{x}$ is optimal.