Linear Programming Notes

Jingyuan Hu

Disclaimer: This note is entirely for personal purpose. Please feel free to email me at *juanhu96@gmail.com* if you find any mistakes or you have any suggestions. The majority of the contents/materials are based on Lecture Notes from UCLA EE236A (Linear Programming), UBC CPSC542F (Convex Analysis and Optimization), Cornell ORIE6300 (Mathematical Programming), along with my own understanding and comments. This note will be continuously updated.

Reference textbooks:
Convex Optimization (Stephen Boyd and L. Vandenberghe)
Introduction to Linear Optimization (Dimitris Bertsimas and John N. Tsitsiklis)

Contents

1 Introduction

A brief review on some basic concepts in optimization and linear algebra.

1.1 Linear Algebra Review

Definition 1.1. Subspace: A nonempty subset S of $\mathbb{R}^n$ is a subspace if $x, y \in S$, $\alpha, \beta \in \mathbb{R} \Rightarrow \alpha x + \beta y \in S$

Remark 1.1. This extends recursively to linear combinations of more than two vectors:

$$x_1, ..., x_k \in S, \alpha_1, ..., \alpha_k \in \mathbb{R} \Rightarrow \alpha_1 x_1 + ... + \alpha_k x_k \in S$$

Moreover, all subspaces contain the origin.

Definition 1.2. Range and null space: $range(A) := \mathcal{R}(A) = \{x \in \mathbb{R}^m | x = Ay \text{ for some y}\}$ is a subspace of $\mathbb{R}^m$, and $null(A) = \{x \in \mathbb{R}^n | Ax = 0\}$ is a subspace of $\mathbb{R}^n$

Definition 1.3. Linear independence: A nonempty set of vectors $\{v_1, v_2, ..., v_k\}$ is linearly independent if $\alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k = 0$ holds only for $\alpha_1 = \alpha_2 = \cdots = \alpha_k = 0$

Definition 1.4. Basis: $\{v_1, v_2, ..., v_k\} \subseteq S$ is a basis of a subspace S if (1) every $x \in S$ can be expressed as a linear combination of $v_1, v_2, ..., v_k$; (2) $\{v_1, v_2, ..., v_k\}$ is a linearly independent set

Note: : This is equivalent to say $x \in S$ is uniquely expressed by $x = \alpha_1 v_1 + \alpha_2 v_2 + \cdots + \alpha_k v_k$

Definition 1.5. Dimension: $dim(S)$: the number of vectors in a basis of S

Remark 1.2. All bases of a subspace have the same size

Example 1.1. $Ax = b$ with $A \in \mathbb{R}^{m \times n}$, then $range(A)$ determines existence of solutions (if $range(A) = \mathbb{R}^m$, there is at least one soln for every $b \in \mathbb{R}^m$) and $null(A)$ determines uniqueness of solutions (if $\hat{x}$ is a solution, then the complete soln set: $\{\hat{x} + v | Av = 0\}$)

Remark 1.3. If $null(A) = \{0\}$, there's at most one soln for every b, which can be shown be contradtion ($Ax_1 = b, Ax_2 = b \Rightarrow A(x_1 - x_2) = 0$)

Definition 1.6. Rank: The (column) rank of a matrix A is defined as $rank(A) := \dim \mathcal{R}(A)$. The row rank of A is $\dim \mathcal{R}(A^T)$.

Theorem 1.1. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$, then $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$ only in finite-dimensional vector spaces; $\mathcal{R}(A)^\perp = \mathcal{N}(A^T)$ holds for all (including infinite-dimensional) vector spaces.*

Proof. Take an arbitrary $x \in \mathcal{N}(A)$, then $Ax = 0$, and this is equivalent to $y^T Ax = 0$ for all y. Notice that $y^T Ax = (A^T y)^T x$, then $Ax = 0$ iff $x \perp A^T y$, i.e. $x \in \mathcal{R}(A^T)$. This implies $\mathcal{N}(A)^\perp = \mathcal{R}(A^T)$ and the rest are similiar. □

Theorem 1.2. *Let $A : \mathbb{R}^n \to \mathbb{R}^m$, then $\dim \mathcal{R}(A) = \dim \mathcal{N}(A)^\perp = \dim \mathcal{R}(A^T)$, i.e. row rank of A = column rank of A.*

Property 1.1. $rank(A) = rank(A^T)$

Property 1.2. $rank(A) \leq \min\{m, n\}$ (A is said to be full rank if equal)

Property 1.3. $\dim null(A) = n - rank(A)$

Definition 1.7. Left-invertible matrix: An $m \times n$ matrix A is left-invertible if: (TFAE)
There exists an X with $XA = I$ (X is called a left inverse of A)
$\iff$ the <u>columns</u> of A form a linearly independent set
$\iff$ $Ax = b$ has <u>at most</u> one solution for every r.h.s. b
$\iff$ $rank(A) = n \iff null(A) = \{0\}$

Remark 1.4. $A \in \mathbb{R}^{m\times n}$ left-invertible $\Rightarrow m \geq n$.

Definition 1.8. Right-invertible matrix: An $m \times n$ matrix A is right-invertible if: (TFAE)
There exists an Y with $AY = I$ (Y is called a right inverse of A)
$\iff$ the rows of A form a linearly independent set
$\iff$ $Ax = b$ has at least one solution for every r.h.s. b
$\iff$ $rank(A) = m \iff range(A) = \mathbb{R}^m$

Remark 1.5. $A \in \mathbb{R}^{m\times n}$ right-invertible $\Rightarrow m \leq n$.

Definition 1.9. Invertible matrix: A is invertible (nonsingular) if it is left- and right-invertible

Remark 1.6. From the definition of left- and right-invertible, we can easily conclude that A is necessarily square, and $Ax = b$ has exactly one solution for every r.h.s. b

Remark 1.7. If left inverse (X) and right inverses (Y) exist, they must be equal and unique. We use the notation A^{-1} for the left/right inverse of an invertible matrix. (i.e. $XA = I, AY = I \Rightarrow X = X(AY) = (XA)Y = Y$)

1.2 Optimization Review

Property 1.4. Cauchy-Schwarz inequality:

$$x^T y = \|x\| \, \|y\| \, cos\theta \Rightarrow -\|x\| \, \|y\| \leq x^T y \leq \|x\| \, \|y\|$$

Property 1.5. Projection (x on the line defined by nonzero y): the **vector** $\hat{t}y$ with

$$\hat{t} = \arg\min_t \|x - ty\| \Rightarrow \hat{t} = \frac{x^T y}{\|y\|^2} = \frac{\|x\| \, \|y\| \, cos\theta}{\|y\|^2} = \frac{\|x\| \, cos\theta}{\|y\|}$$

Definition 1.10. Hyperplane: Solution set of one linear equation with nonzero coefficient vector a: $a^T x = b$

Definition 1.11. Halfspace: Solution set of one linear inequality with nonzero coefficient vector a: $a^T x \leq b$

Note: a is the *normal vector* of hyperplane $G = \{x | a^T x = b\}$ and halfspace $H = \{x | a^T x \leq b\}$

Note: Vector $u = \frac{b}{\|a\|^2} a \Rightarrow a^T u = b$ (u is in the direction of a, which is the normal vector.)

Remark 1.8. x in hyperplane G if $a^T(x - u) = 0$, x in halfspace H if $a^T(x - u) \leq 0$

Definition 1.12. Polyhedron: Solution set of a *finite* number of linear inequalities

$$a_1^T x \leq b_1, a_2^T x \leq b_2, ..., a_m^T x \leq b_m$$

(Matrix notation: $Ax \leq b$ if A is a matrix with rows a_i^T)
A polyhedron is an intersection of a *finite* number of halfspaces

Note: Some linear inequalities in polyhedron can be equalities $Fx = g \iff Fx \leq g, -Fx \leq -g$.

Definition 1.13. Linear Function: $f : \mathbb{R}^n \to \mathbb{R}$, $f(\alpha x + \beta y) = \alpha f(x) + \beta f(y), \forall x, y \in \mathbb{R}^n, \alpha, \beta \in \mathbb{R}$

Property 1.6. f is linear $\iff$ $f(x) = a^T x$ for some a

Proof. ($\Leftarrow$): $f(\alpha x + \beta y) = a^T(\alpha x) + a^T(\beta y) = \alpha(a^T x) + \beta(a^T y) = \alpha f(x) + \beta f(y)$
($\Rightarrow$): $f(x) = f(x_1 e_1 + \cdots + x_n e_n) = x_1 f(e_1) + \cdots + x_n f(e_n) = x_1 a_1 + \cdots + x_n a_n = a^T x$ $\square$

Definition 1.14. Affine Function: $f : \mathbb{R}^n \to \mathbb{R}$, $f(\alpha x + (1-\alpha)y) = \alpha f(x) + (1-\alpha) f(y), \forall x, y \in \mathbb{R}^n, \alpha \in \mathbb{R}$

Property 1.7. f is affine $\iff f(x) = a^T x + b$ for some a, b

Proof. $\square$

Note:
(1) A linear function is affine
(2) A function is linear if and only if it preserves the linear (aka vector space) structure (i.e. operations of vector addition and multiplication by scalar)
(3) An affine function is a function composed of a linear function + a constant; It is linear if and only if it fixes the origin
(4) A function is affine if and only if it preserves the affine structure (i.e. transitive free action $(\nu, s) \to \nu + s$ by a vector space V)

Definition 1.15. Affine Set: A subset S of $\mathbb{R}^n$ is affine if $x, y \in S$, $\alpha + \beta = 1 \Rightarrow \alpha x + \beta y \in S$

Definition 1.16. The line through any two distinct points x, y in S is in S. The definition can be extend recursively to affine combinations of more than two vectors:

$$x_1, ..., x_k \in S, \alpha_1 + \cdots + \alpha_k = 1 \Rightarrow \alpha_1 x_1 + \cdots + \alpha_k x_k \in S$$

Property 1.8. Parallel Subspace: A nonempty set S is affine if and only if the set $L = S - \hat{x}$ with $\hat{x} \in S$ is a subspace.

Remark 1.9. The parallel subspace L is independent of the choice of $\hat{x} \in S$.

Remark 1.10. $\dim(S) = \dim(L)$

Remark 1.11. $S = \{x | Ax = b\}$ is an affine set. Moreover, all affine sets can be represented this way.

Remark 1.12. $S = \{x | x = Ay + c \text{ for some } y\}$ is an affine set. Moreover, all *nonempty* affine sets can be represented this way.

Definition 1.17. Affine hull: Affine hull of a set C is the smallest affine set that contains C, i.e. set of all affine combinations of points in C :

$$\text{aff}C = \{\alpha_1 v_1 + \cdots + \alpha_k v_k | k \geq 1, v_1, ..., v_k \in C, \alpha_1 + \cdots + \alpha_k = 1\}$$

Example 1.2. $C = \{(x, y, z) \in \mathbb{R}^3 | x^2 + y^2 = 1, z = 1\} \Rightarrow \text{aff}C = \{(x, y, z) \in \mathbb{R}^3 | z = 1\}$

Definition 1.18. Affine independence: a set of vectors $\{v_1, v_2, ..., v_k\}$ in $\mathbb{R}^n$ is affinely independent if

$$rank(\begin{bmatrix} v_1 & v_2 & ... & v_k \\ 1 & 1 & ... & 1 \end{bmatrix}) = k$$

Property 1.9. If $\{v_1, v_2, ..., v_k\}$ in $\mathbb{R}^n$ is affinely independent: the set $\{v_2 - v_1, v_3 - v_1, \ldots, v_k - v_1\}$ is linearly independent; the affine hull of $\{v_1, v_2, ..., v_k\}$ has dimension $k - 1$; $k \leq n + 1$

2 Piecewise-linear Optimization

2.1 Definitions and formulation

Definition 2.1. Piecewise-linear Function: $f : \mathbb{R}^n \to \mathbb{R}$ is (convex) **piecewise-lienar** (more accurately, *piecewise-affine*) if it can be expressed as $f(x) = \max_{i=1,\ldots,m} (a_i^T x + b_i)$

Example 2.1. Piecewise-linear Optimization
The optimization problem $\min f(x)$ where $f(x) = \max_{i=1,...,m}(a_i^T x + b_i)$ is equivalent to the linear programming problem

$$\min t$$

subject to

$$a_i^T x + b_i \leq t, i = 1, ..., m$$

(For fixed x the optimal t is $t = f(x)$)

Note: The corresponding matrix form: $\min \tilde{c}^T \tilde{x}$ subject to $\tilde{A}\tilde{x} \leq \tilde{b}$ where
$\tilde{x} = \begin{bmatrix} x \\ t \end{bmatrix}$, $\tilde{c} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, $\tilde{A} = \begin{bmatrix} a_1^T & -1 \\ \vdots & \vdots \\ a_m^T & -1 \end{bmatrix}$, $\tilde{b} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \end{bmatrix}$ (We put b on the RHS and the rest on the LHS)

2.2 Examples(ℓ_1, ℓ_∞)

Example 2.2. Minimizing a sum of piecewise-linear functions
Consider the optimization problem: $\min f(x)+g(x)$ where $f(x) = \max_{i=1,...,m}(a_i^T x+b_i)$, $g(x) = \max_{i=1,...,p}(c_i^T x+d_i)$.
The corresponding cost function is piecewise-linear (maximum of mp affine functions):

$$f(x) + g(x) = \max_{i=1,...,m;j=1,...,p}(a_i + c_j)^T x + (b_i + d_j)$$

Hence the equivalence LP ($m + p$ inequalities) $\min t_1 + t_2$ subject to

$$a_i^T x + b_i \leq t_1$$

$$c_j^T x + d_j \leq t_2$$

where

$$i = 1, ..., m, j = 1, ..., p$$

Matrix form: $\min \tilde{c}^T \tilde{x}$ subject to $\tilde{A}\tilde{x} \leq \tilde{b}$ where $\tilde{x} = \begin{bmatrix} x \\ t_1 \\ t_2 \end{bmatrix}$, $\tilde{c} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, $\tilde{A} = \begin{bmatrix} a_1^T & -1 & 0 \\ \vdots & \vdots & \\ a_m^T & -1 & 0 \\ c_1^T & 0 & -1 \\ \vdots & \vdots & \\ c_p^T & 0 & -1 \end{bmatrix}$, $\tilde{b} = \begin{bmatrix} -b_1 \\ \vdots \\ -b_m \\ -d_1 \\ \vdots \\ -d_p \end{bmatrix}$

Example 2.3. ℓ_∞-Norm (Cheybshev) Approximation
Consider the optimization problem: $\min \|Ax - b\|_\infty$ with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
(Recall that for $m-$vector y: $\|y\|_\infty = \max_i |y_i| = \max_i \max\{y_i, -y_i\}$)
The equivalent LP (with variables x and auxiliary scalar variable t) is $\min t$ subject to

$$-t\mathbf{1} \leq Ax - b \leq t\mathbf{1}$$

Matrix form: $\min \begin{bmatrix} 0 \\ 1 \end{bmatrix}^T \begin{bmatrix} x \\ t \end{bmatrix}$ subject to $\begin{bmatrix} A & -\mathbf{1} \\ -A & -\mathbf{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix}$ (notice that $\mathbf{1}$ is vector)

Example 2.4. ℓ_1-Norm Approximation
Consider the optimization problem: $\min \|Ax - b\|_1$ with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$
The equivalent LP is $\min \sum_{i=1}^{m} u_i$ subject to

$$-u \leq Ax - b \leq u$$

Matrix form: $\min \begin{bmatrix} 0 \\ \mathbf{1} \end{bmatrix}^T \begin{bmatrix} x \\ u \end{bmatrix}$ subject to $\begin{bmatrix} A & -I \\ -A & -I \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix}$

Note: ℓ_1-norm approximation is more robust against outliers.

2.3 Nullspace condition for exact recovery

Example 2.5. Sparse signal recovery via ℓ_1-Norm minimization
$\hat{x} \in \mathbb{R}^n$ is unkonwn (know to be very sparse), we make linear measurement y ($y = A\hat{x}$ with $A \in \mathbb{R}^{m\times n}, m < n$) then minimize $\|x\|_1$ subject to $Ax = y$ gives an estimate with the smallest ℓ_1-norm s.t it is consistent with the measurement (i.e. $Ax = y$)

Question: Subject to $Ax = y$, when is minimizing $card(x)$ equivalent to minimizing $\|x\|_1$?

Definition 2.2. Exact recovery
We say A allows **exact recovery** of k-sparse vectors (at most k non-zero entries) if $\hat{x} = \underset{Ax=y}{\arg\min} \|x\|_1$ when $y = A\hat{x}$ and $card(x) \leq k$ (a property of (the nullspace) of the measurement matrix A)

Remark 2.1. In other words, when the measurement is given ($y = A\hat{x}$) and the true solution is k-sparse ($card(x) \leq k$), the minimizer $\underset{Ax=y}{\arg\min} \|x\|_1$ coincides with the exact true solution $\hat{x}$, we say the measurement matrix A allows exact recovery of k-sparse vectors

Note: A depends on how we measure the original true signal $\hat{x}$

Notation 2.1. x has support $I \subseteq \{1, 2, ..., n\}$ if $x_i = 0$ for $i \notin I$

Notation 2.2. P_I: projection matrix on n-vectors with support I (i.e. P_I is diagonal, $(P_I)_{jj} = 1$ if $j \in I$ and $= 0$ otherwise)

Theorem 2.1. *Necessary and sufficient condition for exact recovery of k-sparse vectors:*

$$|z^{(1)}| + ... + |z^{(k)}| < \frac{1}{2} \|z\|_1, \forall z \in nullspace(A)\backslash\{0\}$$

where $z^{(i)}$ denotes component z_i in order of decreasing magnitude (i.e. $|z^{(1)}| \geq |z^{(2)}| \geq ... \geq |z^{(n)}|$)

Proof. Notice that

$$|z^{(1)}| + ... + |z^{(k)}| < \frac{1}{2} \|z\|_1, \forall z \in nullspace(A)\backslash\{0\}$$

is equivalent to $\|P_I z\|_1 \leq \frac{1}{2} \|z\|_1$ for all nonzero z in $null(A)$ and for all support set I with $|I| \leq k$ (since the inequality holds for largest k component, it must hold for arbitrary index set with cardinality $\leq k$). We call this **nullspace condition**. We want to show $\|P_I z\|_1 \leq \frac{1}{2} \|z\|_1 \iff \hat{x} = \underset{Ax=y}{\arg\min} \|x\|_1$.

($\Rightarrow$): Suppose A satisfies the nullspace condition, let $\hat{x}$ be a k-sparse vector with support I so we have $P_I\hat{x} = \hat{x}$, and define $y = A\hat{x}$ as usual. Consider any feasible x (i.e. satisfying $Ax = y$) that is different from $\hat{x}$. Denote $z = x - \hat{x}$, by assumption this implies $0 \neq z \in null(A)$. Then, we have

$$\begin{aligned} \|x\|_1 &= \|\hat{x} + z\|_1 \geq \|\hat{x} + (z - P_I z)\|_1 - \|P_I z\|_1 \\ &= \sum_{k\in I} |\hat{x}_k| + \sum_{k\neq I} |z_k| - \|P_I z\|_1 \\ &= \|\hat{x}\|_1 + \|z\|_1 - \|P_I z\|_1 - \|P_I z\|_1 \\ &= \|\hat{x}\|_1 + \|z\|_1 - 2\|P_I z\|_1 \\ &> \|\hat{x}\|_1 \end{aligned}$$

since the nullspace condition $\|P_I z\|_1 \le \frac{1}{2}\|z\|_1$ implies $\|z\|_1 - 2\|P_I z\|_1 \ge 0$, hence $\hat{x} = \arg\min_{Ax=y} \|x\|_1$.

($\Leftarrow$): We prove by showing the contrapositive, suppose A does not satisfy the nullspace condition, i.e. there exist some $0 \ne z \in null(A)$ with supp set $I, |I| \le k$, such that $\|P_I z\|_1 \ge \frac{1}{2}\|z\|_1$. We want to show A does not allow exact recovery. Pick an arbitrary k-sparse vector $\hat{x} = -P_I z$ and $y = A\hat{x}$, and denote $x = \hat{x} + z$, which satisfies $Ax = y$ (since $A(\hat{x}+z) = A\hat{x} = y$), then

$$\|x\|_1 = \|-P_I z + z\|_1 = \|z\|_1 - \|P_I z\|_1 \le 2\|P_I z\|_1 - \|P_I z\|_1 = \|\hat{x}\|_1$$

hence $\hat{x}$ is not the unique minimizer (i.e. $\hat{x} \ne \arg\min_{Ax=y} \|x\|_1$), which finishes the proof. □

Remark 2.2. $\frac{1}{2}\|z\|_1$ gives a bound on how concentrated nonzero vectors in *nullspace*(A) can be. This implies that $k < \frac{n}{2}$. (However, this is difficult to verify for general A)

2.4 Linear classification

Example 2.6. Given a set of points $\{v_1, ..., v_N\}$ with binary labels $s_i \in \{-1, 1\}$, need to find hyperplane that strictly separates the two classes, i.e.

$$\begin{aligned} a^T v_i + b > 0 \text{ if } s_i = 1 \\ a^T v_i + b < 0 \text{ if } s_i = -1 \end{aligned}$$

which is equivalent to $s_i(a^T v_i + b) \ge 1, i = 1, ..., N$. Therefore, we have the following optimization problem

$$minimize \sum_{i=1}^{N} \max\{0, 1 - s_i(a^T v_i + b)\}$$

The equivalent LP:

$$\begin{aligned} &\text{minimize } \sum_{i=1}^{N} u_i \\ &\text{subject to } 1 - s_i(a^T v_i + b) \le u_i \\ &\qquad\qquad 0 \le u_i \end{aligned}$$

In matrix form: $\min \begin{bmatrix} 0 \\ 0 \\ \mathbf{1} \end{bmatrix}^T \begin{bmatrix} a \\ b \\ u \end{bmatrix}$ subject to $\begin{bmatrix} -s_1 v_1^T & -s_1 & -1 & 0 & \cdots & 0 \\ -s_2 v_2^T & -s_2 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -s_N v_N^T & -s_N & 0 & 0 & \cdots & -1 \\ 0 & 0 & -1 & 0 & \cdots & 0 \\ 0 & 0 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} a \\ b \\ u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \le \begin{bmatrix} -1 \\ -1 \\ \vdots \\ -1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

Note: This is known as support vector machine.

3 Polyhedra

3.1 Polyhedron

Definition 3.1. Polyhedron: A polyhedron is the *solution set* of a *finite* number of linear inequalities

Note: The soln set can include equalities.

Remark 3.1. The solution of the infinite set of linear inequalities $a^T x \leq 1$ for all a with $\|a\|_1$ is the unit ball $\{x| \, \|x\| \leq 1\}$ and not a polyhedron.

Note: In this section we consider polyhedron in the form of $P = \{x|Ax \leq b, Cx = d\} \neq \emptyset$.

Definition 3.2. Lineality Space: The *lineality space* of P is $L = null(\begin{bmatrix} A \\ C \end{bmatrix})$ i.e. $\forall v \in L, Av = Cv = 0$.

Corollary 3.0.1. *If $x \in P$, then $x + v \in P$ for all $v \in L$.*

Definition 3.3. Pointed Polyhedron: A polyhedron with lineality space $\{0\}$ is called pointed. *A polyhedron is pointed if it does not contain an entire line (Goes off to inifinity in both directions).*

Remark 3.2. In other words, a polyhedron P is pointed if and only if $linspace(P) = \{0\}$, that is $linspace(P)$ has dimension 0.

Example 3.1. (Pointed polyhedra)
Probability simplex $P = \{x \in \mathbb{R}^n | \mathbf{1}^T x = 1, x \geq 0\}$
$P = \{(x, y, z) | |x| \leq z, |y| \leq z\}$

Example 3.2. (Not pointed polyhedra):
$P = \{x|a^T x \leq b\}$ $(n \geq 2)$: $L = \{x|a^T x = 0\}$
$P = \{x| -1 \leq a^T x \leq 1\}$ $(n \geq 2)$: $L = \{x|a^T x = 0\}$
$P = \{(x, y, z) | |x| \leq 1, |y| \leq 1\} : L = \{(0, 0, z)|z \in \mathbb{R}\}$

Note: For this first two example $n \geq 2$: otherwise $L = \{0\}$ and hence pointed.

3.2 Face

Definition 3.4. Face: Let $J \subseteq \{1, 2, ..., m\}$ an index set, define $F_J = \{x \in P | a_i^T x = b_i, \forall i \in J\}$. If $F_J \neq \emptyset$, it is called a *face* of P.

Property 3.1. F_J is a nonempty polyhedron, defined by the inequalities and equalities:

$$\{x \in P | a_i^T x \leq b_i, \forall i \notin J, a_i^T x = b_i, \forall i \in J, Cx = d\}$$

Property 3.2. The number of faces is finite and at least one (P itself is a face $P = F_\emptyset$)

Property 3.3. Faces of F_J are also faces of P.

Property 3.4. All faces have the same lineality space as P

Example 3.3. Consider the following problem:

$$\begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & 1 \\ -1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The solution set is a (non-pointed, hence include an entire line) polyhedron: $P = \{x \in \mathbb{R}^3 \big| |x_1 - x_2| + |x_3| \leq 1\}$.
The lineality space is the line: $L = \{(t, t, 0)|t \in R\}$.
Example of faces of the given polyhedron P:
Three-dimensional face: $F_\emptyset = P$
Two dim faces (surface):
$F_{\{1\}} = \{x|x_1 - x_2 + x_3 = 1, x_1 \geq x_2, x_3 \geq 0\}$ (only the first equality holds)
$F_{\{2\}} = \{x|x_1 - x_2 - x_3 = 1, x_1 \geq x_2, x_3 \leq 0\}$ (only the second equality holds)
$F_{\{3\}} = \{x| -x_1 + x_2 + x_3 = 1, x_1 \leq x_2, x_3 \geq 0\}$ (only the third equality holds)

$F_{\{4\}} = \{x | -x_1 + x_2 - x_3 = 1, x_1 \leq x_2, x_3 \leq 0\}$ (only the fourth equality holds)
One dim faces (line/edge):
$F_{\{1,2\}} = \{x | x_1 - x_2 = 1, x_3 = 0\}$ (first two equality: $x_1 - x_2 + x_3 = 1, x_1 - x_2 - x_3 = 1 \Rightarrow x_1 - x_2 = 1, x_3 = 0$)
$F_{\{1,3\}} = \{x | x_1 = x_2, x_3 = 1\}$
$F_{\{2,4\}} = \{x | x_1 = x_2 = 1, x_3 = -1\}$
$F_{\{3,4\}} = \{x | x_1 - x_2 = -1, x_3 = 0\}$
and F_J is empty for all other J.

Definition 3.5. Minimal face: A face of P is a *minimal face* if it does not contain another face of P.

Example 3.4. The minimal face of polyhedron in the above example: $F_{\{1,2\}}, F_{\{1,3\}}, F_{\{2,4\}}, F_{\{3,4\}}$. (Intuition: More contraints, smaller face)

Theorem 3.1. *If F is a face of P and $F' \subseteq F$, then F' is a face of P iff F' is a face of F.*

Theorem 3.2. *A face is minimal if and only if it is an affine set*

Proof. Let F_J be a face where $a_i^T x \leq b_i, \forall i \notin J$, $a_i^T x = b_i, \forall i \in J$ and $Cx = d$. Then, partition the inequalities $a_i^T x \leq b_i (i \notin J)$ into three groups:
(1) $i \in J_1$ if $a_i^T x = b_i$ for all x in F_J
(2) $i \in J_2$ if $a_i^T x < b_i$ for all x in F_J
(3) $i \in J_3$ if there exist points $\hat{x}, \tilde{x} \in F_J$ with $a_i^T \hat{x} < b_i$ and $a_i^T \tilde{x} = b_i$.
If J_3 is not empty then there exist some $j \in J_3$, $F_{J \cup \{j\}}$ is a proper face of F_J since
(a) $F_{J \cup \{j\}} \neq \emptyset$ ($\because \tilde{x} \in F_{J \cup \{j\}}$) and (b) $F_{J \cup \{j\}} \neq F_J$ ($\because \hat{x} \notin F_J$). Therefore, if F_J is a minimal face then $J_3 = \emptyset$ and F_J is the solution set of $a_i^T x = b_i$ for $i \in J_1 \cup J$ and $Cx = d$. (Unfinished) $\square$

Theorem 3.3. *All minimal faces are translates of the linearlity space of P. (Since all faces have the same linearity space)*

3.3 Extreme Point and Rank Test

Definition 3.6. Extreme Point (vertex): A minimal face of a pointed polyhedron.

Note: Question: Is a minimal face of a pointed polyhedron always a single point? Answer: A face of polyhedron P is an extreme point or a vertex if it has dimension 0.

Remark 3.3. Other definition: $x \in P$ is a vertex if there exist c such that $c^T x < c^T y$ for all $y \neq x$, where $y \in P$. Moreover, if P is convex, $x \in P$ is an extreme point of P if it cannot be written as $\lambda y + (1-\lambda) z$ for $y, z \in P$ where $y, z \neq x$, and $0 \leq \lambda \leq 1$.

Question: Given $\hat{x} \in P$, is it an extreme point?
Rank Test: Denote $J(\hat{x}) = \{i_1, i_2, ..., i_k\}$ be the indices of the **active contraints** at $\hat{x}$: $a_i^T \hat{x} = b_i$ for $i \in J(\hat{x})$ and $a_i^T \hat{x} < b_i$ for $i \notin J(\hat{x})$.
Note: $J(\hat{x})$ depends on the point $\hat{x}$ given.

Property 3.5. $\hat{x} \in P$ is a extreme point if $rank(\begin{bmatrix} A_{J(\hat{x})} \\ C \end{bmatrix}) = n$ where $A_{J(\hat{x})} = \begin{bmatrix} a_{i1}^T \\ \vdots \\ a_{ik}^T \end{bmatrix}$

Note: This property tells us to only care about the active contraints instead of all contratints in A. The intuition is the active contraints generate by some point leads to a full rank system and hence unique solution (extreme point).

Proof. Face $F_{J(\hat{x})}$ is defined as the set of points x satisfies $(*)$: $a_i^T x = b_i$ for $i \in J(\hat{x})$ and $a_i^T x \leq b_i$ for $i \notin J(\hat{x})$ (and $Cx = d$ as well). Then clearly by defintion we have $\hat{x} \in F_{J(\hat{x})}$. If the rank condition is satisfied (i.e. full rank), then we know that there is an unique solution, in this case which is given, $\hat{x}$. Therefore we have $F_{J(\hat{x})} = \{\hat{x}\}$ is a minimal face ($dim(F_{J(\hat{x})}) = 0$, does not contain any other face). If the rank condition does not hold, then there exist some $v \neq 0$ such that $a_i^T v = 0$ for $i \in J(\hat{x})$ and $Cv = 0$. This implies that $x = \hat{x} \pm tv$ is also in $F_{J(\hat{x})}$ for *small* positive and negative t, therefore the face $F_{J(\hat{x})}$ is not minimal. $\square$

Example 3.5. Consider $\begin{bmatrix} -1 & 0 \\ 2 & 1 \\ 0 & -1 \\ 1 & 2 \end{bmatrix} x \leq \begin{bmatrix} 0 \\ 3 \\ 0 \\ 3 \end{bmatrix}$, then obviously $\hat{x} = (1, 1)$ is in P since $\begin{bmatrix} -1 & 0 \\ 2 & 1 \\ 0 & -1 \\ 1 & 2 \end{bmatrix} \hat{x} = \begin{bmatrix} -1 \\ 3 \\ -1 \\ 3 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 3 \\ 0 \\ 3 \end{bmatrix}$, and the active constraints $J(\hat{x}) = \{2, 4\}$. (here C is empty since all inequalities) Then, $A_{J(\hat{x})} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ has rank 2 = # of cols, hence $\hat{x}$ is an extreme point.

Example 3.6. Consider $A : \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \leq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$ and $C : x_1 + x_2 + x_3 = 1$, then some extreme points are listed as follow:

$\hat{x} = (1, 0, 0) \Rightarrow J(\hat{x}) = \{2, 3\} \Rightarrow rank(\begin{bmatrix} A_{J(\hat{x})} \\ C \end{bmatrix}) = rank(\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}) = 3$ and

$\hat{x} = (0, 1, 0) \Rightarrow J(\hat{x}) = \{1, 3\} \Rightarrow rank(\begin{bmatrix} A_{J(\hat{x})} \\ C \end{bmatrix}) = rank(\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}) = 3$ and

$\hat{x} = (0, 0, 1) \Rightarrow J(\hat{x}) = \{1, 2\} \Rightarrow rank(\begin{bmatrix} A_{J(\hat{x})} \\ C \end{bmatrix}) = rank(\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 1 \end{bmatrix}) = 3$

Example 3.7. Consider nonempty pointed polyhedron $P = \{x \geq 0, Cx = d\}$, show: (1) $\hat{x}$ is an extreme point if $\hat{x} \in P$ and $rank([c_{i1}, c_{i2}, ...c_{ik}]) = k$ where c_j is column j of C and $\{i_1, i_2, ..., i_k\} = \{i|\hat{x}_i > 0\}$ (2) an extreme point $\hat{x}$ has at most $rank(C)$ nonzero elements

Proof. WLOG, assume $\{i_1, \ldots, i_k\} = \{1, \ldots, k\}$. Apply rank test to $\begin{bmatrix} -I \\ C \end{bmatrix} = \begin{bmatrix} -I_k & 0 \\ 0 & -I_{n-k} \\ D & E \end{bmatrix}$ where $D = [c_1 \ldots c_k]$ and $E = [c_{k+1} \ldots c_n]$. inequalities $k+1, ..., n$ are active at $\hat{x}$. $\hat{x}$ is an extreme point if the submatrix of active constraints has rank n, i.e.

$$rank(\begin{bmatrix} 0 & -I_{n-k} \\ D & E \end{bmatrix}) = n - k + rank(D) = n \Rightarrow rank(D) = k$$

$\square$

Definition 3.7. Doubly stochastic matrix: an $n \times n$ matrix X is doubly stochastic if

$$X_{ij} \geq 0, i, j = 1, ..., n, X\mathbf{1} = 1, X^T\mathbf{1} = 1$$

Note: nonnegative matrix with column and row sums equal to one, set of doubly stochastic matrices form is a pointed polyhedron in $\mathbb{R}^{n \times n}$

Definition 3.8. A **permutation matrix** is a doubly stochastic matrix with elements 0 or 1

Example 3.8. Birkhoffs theorem (Birkhoff von Neumann Theorem) The extreme points are the permutation matrices. (Set of $n \times n$ doubly stochastic matrices forms a convex polytope whose vertices are the $n \times n$ permutation matrices)

4 Convexity

4.1 Convex Hull

Definition 4.1. Convex combination: A convex combination of points $v_1, \ldots, v_k$ is a linear combination $x = \theta_1 v_1 + \theta_2 v_2 + ... + \theta_k v_k$ with $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$

Definition 4.2. Convex set: A set S is convex if it contains all convex combinations of points in S.

Example 4.1. Some examples of convex sets: affine sets $(Cx = d, Cy = d \Rightarrow C(\theta x + (1-\theta)y) = d, \forall \theta \in \mathbb{R})$ and polyhedra $(Ax \leq b, Ay \leq b \Rightarrow A(\theta x + (1-\theta)y) \leq b, \forall \theta \in [0,1])$

Definition 4.3. Convex hull: The convex hull of a set S: the set of all convex combinations of points in S, denote as $convS$

Definition 4.4. polytope: The convex hull $conv\{v_1, v_2, \ldots, v_k\}$ of a finite set of points.

Remark 4.1. $convS$ is the set of points x that can be expressed as $x = \theta_1 v_1 + \theta_2 v_2 + ... + \theta_k v_k$ with $\theta_i \geq 0$ and $\sum_{i=1}^k \theta_i = 1$, where $v_1, \ldots, v_k \in S$. (This directly follows from the defintion above)

Theorem 4.1. *(**Caratheodorys theorem**) If $S \subseteq \mathbb{R}^n$ then k can be taken less than or equal to $n+1$. (e.g in $\mathbb{R}^2$, every $x \in convS$ can be written as a convex combination of 3 points in S)*

Proof. Start from any convex decomposition of x:

$$\begin{bmatrix} x \\ 1 \end{bmatrix} = \begin{bmatrix} v_1 & v_2 & \ldots & v_m \\ 1 & 1 & ... & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$$

Let P be the set of vectors $\theta = (\theta_1, \theta_2, ..., \theta_m)$ that satisfies these conditions. Then P is a nonempty polyhedron described in 'standard form' from *Example* 3.7 (where $C = \begin{bmatrix} v_1 & v_2 & \ldots & v_m \\ 1 & 1 & ... & 1 \end{bmatrix}$ and $d = \begin{bmatrix} x \\ 1 \end{bmatrix}$ and $x = \theta \geq 0$). Hence, if $\hat{\theta} \in P$ is an extreme point, it must satisfy the rank test

$$rank(\begin{bmatrix} v_{i1} & v_{i2} & \ldots & v_{ik} \\ 1 & 1 & ... & 1 \end{bmatrix}) = k$$

where $\{i_1, \ldots, i_k\} = \{i | \hat{\theta}_i > 0\}$. This rank condition implies $k \leq n+1$.

4.2 Polyhedral Cone

Definition 4.5. Cone: S is a cone if $x \in S$ implies $\alpha x \in S$ for all $\alpha \geq 0$

Definition 4.6. Convex cone: A nonempty set S with the property $x_1, x_2, \ldots, x_k \in S$ and $\theta_1 \geq 0, \ldots, \theta_k \geq 0 \Rightarrow \theta_1 x_1 + ...\theta_k x_k \in S$

Note: In other words, all nonnegative combinations of points in S are in S. S is a convex set as well as a cone by the definition.

Example 4.2. *Examples of convex cone*: all subspaces; a polyhedral cone defined by $S = \{x|Ax \leq 0, Cx = 0\}$ (finite system of homogeneous linear inequalities).

Definition 4.7. Conic hull: A conic hull of set S is the set of all nonnegative combinations of points in S. This is also known as the cone genereated by S, denoted by $coneS$.

Definition 4.8. Finitely generated cone The conic hull $cone\{v_1, v_2, ..., v_k\}$ of a finite set.

Definition 4.9. Pointed polyhedral cone Consider a polyhedral cone $K = \{x \in \mathbb{R}^n|Ax \leq 0, Cx = 0\}$, the linearity space is $null(\begin{bmatrix} A \\ C \end{bmatrix})$. Hence, K is pointed if $rank(\begin{bmatrix} A \\ C \end{bmatrix}) = n$, in which case it has one extreme point (the origin). *Recall that a polyhedron with lineality space* $\{0\}$ *is called pointed.*

Remark 4.2. The one-dimensional faces are called **extreme rays**.

Definition 4.10. Recession(Asymptotic) cone: The recession/asymptotic cone of a polyhedron $P = \{x|Ax \leq b, Cx = d\}$ is $K = \{y|Ay \leq 0, Cy = 0\}$.

Property 4.1. (1)K has the same lineality space as P; (2)K is pointed iff P is pointed; (3) if $x \in P$ then $x + y \in P$ for all $y \in K$

4.3 Decomposition

Theorem 4.2. *Every polyhedron P can be decomposed as*

$$P = L + Q = L + conv\{v_1, \dots, v_r\} + cone\{w_1, \dots, w_s\}$$

where L is the lineality space and Q is a pointed polyhedron. Moreover, $\{v_1, \dots, v_r\}$ are the extreme points of Q and $\{w_1, \dots, w_s\}$ generate the extreme rays of the recession cone of Q.

Note: In general, it is extremely costly to compute from inequality description of P.

5 Alternatives

5.1 Alternatives for linear inequalities

Theorem 5.1. *For given A, b, exactly one of the following two statements is true:*
(1) there exists an x that satisfies $Ax \leq b$
(2) there exists an z that satisfies $z \geq 0, A^T z = 0, b^T z < 0$

Proof. First notice that (1) and (2) cannot be both true since (1): $Ax \leq b, z \geq 0 \Rightarrow z^T(Ax - b) \leq 0$ but (2): $A^T z = 0, b^T z < 0 \Rightarrow z^T(Ax - b) > 0$. It suffices to show the statements cannot be both false: whenever (1) is false, (2) has to be true. We'll do this by induction on the column dimension of A.
Basic Case: A has zero columns, and we have (1) $b \geq 0$, then (2) $z \geq 0, b^T z < 0$ cannot be true.
Induction Step: Assume holds for set of inequalities with $n-1$ variables ($n-1$ columns?). Now consider inequality $Ax \leq b$ where A is an $m \times n$ matrix, and divide it into three index groups: $I_+ = \{i|A_{in} > 0\}$, $I_0 = \{i|A_{in} = 0\}$, $I_- = \{i|A_{in} < 0\}$. Scale the inequalities with $A_{in} \neq 0$ to get an equivalent system:

$$\sum_{k=1}^{n-1} C_{ik}x_k + x_n \leq d_i \text{ for } i \in I_+$$

$$\sum_{k=1}^{n-1} C_{ik}x_k - x_n \leq d_i \text{ for } i \in I_-$$

$$\sum_{k=1}^{n-1} A_{ik}x_k \leq b_i \text{ for } i \in I_0$$

where $C_{ik} = \begin{cases} A_{ik}/A_{in} & i \in I_+ \\ -A_{ik}/A_{in} & i \in I_- \end{cases}$ and $d_i = \begin{cases} b_i/A_{in} & i \in I_+ \\ -b_i/A_{in} & i \in I_- \end{cases}$. The inequalities indexed by I_+ and I hold for some x_n iff

$$\max_{i \in I_-}(\sum_{k=1}^{n-1} C_{ik}x_k - d_i) \leq \min_{i \in I_+}(d_i - \sum_{k=1}^{n-1} C_{ik}x_k)$$

. Therefore $Ax \leq b$ is solvable iff there exist $(x_1, \ldots, x_{n-1})$ s.t.

$$\sum_{k=1}^{n-1}(C_{ik} + C_{jk})x_k \leq d_i + d_j \text{ for all } i \in I_-, j \in I_+$$

$$\sum_{k=1}^{n-1} A_{ik}x_k \leq b_i \text{ for all } i \in I_0$$

which is a system of inequalities with $n-1$ variables. If this system is infeasible, then there exist u_{ij} where $i \in I_-, j \in I_+$ and v_i where $i \in I_0$ such that

$$u_{ij} \geq 0 \text{ for } i \in I_-, j \in I_+$$

$$v_i \geq 0 \text{ for } i \in I_0$$

$$\sum_{i \in I_-, j \in I_+} (C_{ik} + C_{jk})u_{ij} + \sum_{i \in I_0} v_i A_{ik} = 0, k = 1, \ldots, n-1$$

$$\sum_{i \in I_-, j \in I_+} (d_i + d_j)u_{ij} + \sum_{i \in I_0} b_i v_i < 0$$

Define

$$z_i = \frac{1}{-A_{in}} \sum_{j \in I_+} u_{ij} \text{ for } i \in I_-$$

$$z_i = \frac{1}{A_{jn}} \sum_{i \in I_-} u_{ij} \text{ for } j \in I_+$$

$$z_i = v_i \text{ for } i \in I_0$$

to get a vector z that satisfies $z \geq 0, A^T z = 0, b^T z < 0$. □

Note: z in statement 2 is a **certificate** of infeasibility of $Ax \leq b$.

5.2 Farkas lemma and other variants

Lemma 5.2. *For given A, b, exactly one of the following two statements is true:*
(1) there exists an x that satisfies $Ax = b, x \geq 0$
(2) there exists an y that satisfies $A^T y \geq 0, b^T y < 0$

Proof. Apply previous thm of alternatives to the following system:

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

Then we know that this system is <u>infeasible</u> iff there exists (u, v, w) s.t. $u, v, w \geq 0, A^T(u - v) = w \geq 0$ and $b^T(u - v) < 0$. Denote $y = u - v$, then we have the desiring statement.

Note: Geometric interpretation (1) b is in the cone generated by the columns of A (2) the hyperplane $y^T z = 0$ separates b from $a_1, \ldots, a_m$ (since $A^T y \geq 0$ and $b^T y < 0$, which means that A and b is on the different side of the hyperplane).

5.3 Mixed inequalities and equalities

Lemma 5.3. *Now given A, C, b, d, exactly one of the following two statements is true:*
(1) there exists an x that satisfies $Ax \leq b, Cx = d$
(2) there exists an y, z that satisfies $z \geq 0, A^T z + C^T y = 0, b^T z + d^T y < 0$

Proof. Write $Cx = d$ as $Cx \leq d$ and $-Cx \leq -d$, then apply thm 5.1 to the linear system

$$\begin{bmatrix} A \\ C \\ -C \end{bmatrix} x \leq \begin{bmatrix} b \\ d \\ -d \end{bmatrix}$$

5.4 Strict inequalities

Lemma 5.4. *Now given A, B, b, c, exactly one of the following two statements is true:*
(1) there exists an x that satisfies $Ax < b, Bx \leq c$
(2) there exists an y, z that satisfies $y \geq 0, z \geq 0, A^T y + B^T z = 0$, and $b^T y + c^T z < 0$ or $b^T y + c^T z = 0$ (when $y \neq 0$)

Proof. Notice that (1) is equivalent to: $\exists u, t$ such that $Au \leq tb - \mathbf{1}$ and $Bu \leq tc, t \geq 1$.

6 Duality

6.1 Dual of LP

We start this section by two form of LP with the same parameters $A \in \mathbb{R}^{m \times n}, c \in \mathbb{R}^n, b \in \mathbb{R}^m$

Example 6.1. LP in 'inequality form' (**primal** problem)

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } Ax \leq b \end{aligned}$$

LP in 'standard form' (the **dual** of the first LP)

$$\begin{aligned} &\text{maximize } -b^T z \\ &\text{subject to } A^T z + c = 0 \\ &\qquad\qquad\quad z \geq 0 \end{aligned}$$

Notation 6.1. Primal and Dual optimal
p^*: primal optimal value; $p^* = +\infty$ if primal is infeasible, $p^* = -\infty$ if primal is unbounded.
d^*: dual optimal value; $d^* = -\infty$ if dual is infeasible, $d^* = +\infty$ if dual is unbounded.

Theorem 6.1. ***(Duality theorem)*** *If primal or dual problem is feasible, then $p* = d*$. Moreover, if $p^* = d^*$ is finite, then primal and dual optima are attained.*

Property 6.1. (**Lower bound property**) If x is primal feasible and z is dual feasible, then $c^T x \geq -b^T z$

Proof. Since x, z are feasible, $Ax \leq b$ and $A^T z + c = 0, z \geq 0$, then

$$0 \leq z^T(b - Ax) = z^T b - z^T Ax = b^T z - x^T A^T z = b^T z + x^T c = b^T z + c^T x$$

□

Note: $b^T z + c^T x$ is the **Duality Gap** associated with primal and dual feasible x, z

Theorem 6.2. ***(Weak Duality)*** $p^* \geq d^*$

Proof. Immediate comes from lower bound property. $\square$

Theorem 6.3. ***(Strong Duality)*** *If primal and dual problems are feasible, then there exist x^*, z^* that satisfy*

$$c^T x^* = -b^T z^*, Ax^* \leq b, A^T z^* + c = 0, z^* \geq 0$$

Combined with the lower bound property, this implies that
(a) This x^* and z^* is the primal/dual optimal; (b) The primal and dual optimal values are finite and equal (i.e. $p^* = c^T x^* = -b^T z^* = d^* < \infty$)

Proof. We show that there exist x^*, z^* that satisfy

$$\begin{bmatrix} A & 0 \\ 0 & -I \\ c^T & b^T \end{bmatrix} \begin{bmatrix} x^* \\ z^* \end{bmatrix} \leq \begin{bmatrix} b \\ 0 \\ 0 \end{bmatrix}$$

and

$$\begin{bmatrix} 0 & -A^T \end{bmatrix} \begin{bmatrix} x^* \\ z^* \end{bmatrix} = c$$

Notice that the last equation is $c^T x^* + b^T z^* \leq 0$, we do not need $c^T x^* + b^T z^* \geq 0$, since by lower-bound property any solution would necessarily satisfies this equation and hence satisfies $c^T x^* + b^T z^* = 0$. To proce such a solution exist we can show the alternative system

$$u \geq 0, t \geq 0, A^T u + tc = 0, Aw \leq tb, b^T u + c^T w < 0$$

has no solution. To show this, we discuss by cases. First, if $t > 0$, define $\tilde{x} = w/t, \tilde{z} = u/t$, we can rewrite the alternative system as

$$\tilde{z} \geq 0, A^T \tilde{z} + c = 0, A\tilde{x} \leq b, b^T \tilde{z} + c^T \tilde{x} < 0$$

Notice the last equation $b^T \tilde{z} + c^T \tilde{x} < 0 \Rightarrow c^T \tilde{x} < -b^T \tilde{z}$ contradicts with the lower bound property. Second, if $t = 0$: (1) $b^T u < 0$: $u \geq 0, A^T u = 0, b^T u < 0$ contradicts feasibility of $Ax \leq b$, hence $b^T u \geq 0$; (2) $c^T w < 0$: $Aw \leq 0, c^T w < 0$ contradicts feasibility of $A^T z + c = 0, z \geq 0$, hence $c^T w \geq 0$. Therefore $b^T u + c^T w \geq 0$, which means the alternative system has no solution. $\square$

Example 6.2. Primal Infeasible: If primal optimal value $p^* = +\infty$, then dual optimal value $d^* = +\infty$ or $d^* = -\infty$.

Proof. If primal is infeasible, then by *thm 5.1*, there exist w such that $w \geq 0, A^T w = 0, b^T w < 0$. Suppose the dual is feasible, there must be at least one feasibile point, denote z, which satisfies

$$z + tw \geq 0, A^T(z + tw) + c = A^T z + c + tA^T w = 0 + t \times 0 = 0$$

for all $t \geq 0$, hence $z + tw$ is dual feasibile for all $t \geq 0$. Take $t \to \infty$, we have

$$-b^T(z + tw) = -b^T z - tb^T w \to \infty$$

which implies the dual problem is unbounded above. (Then $d^* = +\infty$, but why have 'or $d^* = -\infty$'?) $\square$

Example 6.3. Dual Infeasible: If dual optimal value $d^* = -\infty$, then primal optimal value $p^* = -\infty$ or $p^* = +\infty$.

Proof. If dual is infeasible, there does not exist $z \geq 0$ such that $A^T z = -c$, then by *lemma 5.2*, there exist w such that $Aw \geq 0, -c^T w < 0$, i.e. exist $y = -w$ s.t. $Ay \leq 0, c^T y < 0$. Suppose the primal is feasible, there must be at least one feasibile point, denote x, which satisfies

$$A(x + ty) = Ax + tAy \leq b + 0 = b$$

for all $t \geq 0$, hence $x + ty$ is primal feasibile for all $t \geq 0$. Take $t \to \infty$, we have

$$c^T(x + ty) = c^T x + tc^T y \to -\infty$$

which implies the primal problem is unbounded below. $\square$

Example 6.4. Exception to Strong Duality: $p^* = +\infty, d^* = -\infty$ is possible. Consider the primal problem with one variable and one inequality (where A is zero matrix):

$$\begin{aligned} &\text{minimize } x \\ &\text{subject to } 0 \cdot x \leq -1 \end{aligned}$$

and it's dual problem

$$\begin{aligned} &\text{maximize } z \\ &\text{subject to } 0 \cdot z + 1 = 0 \\ &\qquad\qquad\quad z \geq 0 \end{aligned}$$

Then both system are infeasible, $p^* = +\infty, d^* = -\infty$.

Note: Summary table see UCLA236A lecture slide 6-11.

6.2 Variants and Examples

Example 6.5. Standard form LP I

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } Ax \leq b \end{aligned}$$

and

$$\begin{aligned} &\text{maximize } -b^T z \\ &\text{subject to } A^T z + c = 0 \\ &\qquad\qquad\quad z \geq 0 \end{aligned}$$

Example 6.6. Standard form LP II

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } Ax = b \\ &\qquad\qquad\quad x \geq 0 \end{aligned}$$

and

$$\begin{aligned} &\text{maximize } b^T y \\ &\text{subject to } A^T y \leq c \end{aligned}$$

Example 6.7. LP with inequality and equality constraints

$$\begin{aligned} &\text{minimize } c^T x \\ &\text{subject to } Ax \leq b \\ &\qquad\qquad\quad Cx = d \end{aligned}$$

and

$$\begin{aligned} &\text{maximize } -b^T z - d^T y \\ &\text{subject to } A^T z + C^T y + c = 0 \\ &\qquad\qquad\quad z \geq 0 \end{aligned}$$

Example 6.8. Piecewise-linear minimization
minimize $f(x) = \max_{i=1,\ldots,m}(a_i^T x + b_i)$.
The LP formulation is

$$\begin{aligned} &\text{minimize } t \\ &\text{subject to } \begin{bmatrix} A & -\mathbf{1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq -b \end{aligned}$$

the dual LP is

$$\begin{aligned} &\text{maximize } b^T z \\ &\text{subject to } A^T z = 0 \\ &\qquad\qquad\quad \mathbf{1}^T z = 1 \\ &\qquad\qquad\quad z \geq 0 \end{aligned}$$

Interpretation: for any $z \geq 0$ with $\sum_i z_i = 1$, $f(x) = \max_i(a_i^T x + b_i) \geq z^T(Ax + b)$ for all x. This provides a lower bound on the optimal value of the PWL problem, i.e.

$$\min_x f(x) \geq \min_x z^T(Ax + b) = \begin{cases} b^T \text{if } A^T z = 0 \\ -\infty \text{otherwise} \end{cases}$$

where the dual problem is to find the best lower bound of this type.
Note: Strong Duality tells us that the best lower bound is actually tight.

Example 6.9. l_∞−norm approximation
minimize $\|Ax - b\|_\infty$, the LP formulation is

$$\begin{aligned} &\text{minimize } t \\ &\text{subject to } \begin{bmatrix} A & \textbf{-1} \\ -A & \textbf{-1} \end{bmatrix} \begin{bmatrix} x \\ t \end{bmatrix} \leq \begin{bmatrix} b \\ -b \end{bmatrix} \end{aligned}$$

The (1) dual problem is

$$\begin{aligned} &\text{minimize } -b^T u + b^T v \\ &\text{subject to } A^T u - A^T u = 0 \\ &\qquad\qquad\quad 1^T u + 1^T v = 1 \\ &\qquad\qquad\quad u \geq 0, v \geq 0 \end{aligned}$$

and there is a (2) simpler equivalent dual:

$$\begin{aligned} &\text{minimize } t \\ &\text{subject to } A^T z = 0, \|z\|_1 \leq 1 \end{aligned}$$

To show the equivalence of the dual problem (assume A is $m \times n$): If u, z are feasible in the first

6.3 Complementary slackness