Optimization Cheatsheet

Hongyi Guo*

December 4, 2020

1 Linear Programming

Definition 1.1 (Standard Form)**.**

$$\begin{aligned} \min \quad & c^\top x \\ \text{s.t.} \quad & \sum_{j=1}^{n} a_j x_j = b \\ & x \geq 0. \end{aligned}$$

Definition 1.2 (Convex set)**.** A set X is convex if $\forall x, y \in X$, $\lambda x + (1-\lambda)y \in X$, $0 \leq \lambda \leq 1$.

Definition 1.3 (Convex cone)**.** A convex cone C is a convex set s.t. $\forall x \in C$, $\lambda x \in C$, $\forall \lambda \geq 0$.

Definition 1.4 (Hyperplane, half-space)**.** The set $\{x \in \mathbb{R}^n : a^\top x = b\}$ is called a Hyperplane. The set $\{x \in \mathbb{R}^n : a^\top x \geq b\}$ is called a half-space.

Definition 1.5 (Polyhedron)**.** A polyhedron is a set that can be described as the intersection of finitely many inequalities (half spaces) and equalities (hyperplanes).

Definition 1.6 (Bounded)**.** A set $S \in \mathbb{R}^n$ is bounded if there exists a finite constant K s.t. $\forall x \in S, |x_i| \leq K, i = 1, 2, \ldots, n$.

Definition 1.7 (Dimension)**.** Given a polyhedron $P = \{x \in \mathbb{R}^n : Ax \geq b\}$, the dimension $\dim(P)$ is the smallest <u>affine</u> subspace that contains it. P is full dimensional if $\dim(P) = n$.

Theorem 1.8. Two approaches to get the dimension of a polyhedron $P = \{x \in \mathbb{R}^n : Ax \geq b\}$.

1. Let $A^=$ be the submatrix of A consisting of the rows a^i, the i-th row of A s.t. $a^i x = b_i$, $\forall x \in P$ for $i = 1, \ldots, m$. Then $\dim(P) = n - \text{rank}(A^=)$.

*Northwestern University; hongyiguo2025@u.northwestern.edu

2. If the maximum number of affinely independent points in P is k, then $\dim(P) = k - 1$.

Remark 1.9. Normally, Theorem 1.8.1 gives us an upper bound of $\dim(P)$, while Theorem 1.8.2 gives us a lower bound of $\dim(P)$. By showing that they meet each other, we obtain the exact $\dim(P)$.

Definition 1.10. The definition of Linear, affine, conic, and convex combinations.

(a). Linear combination $\text{span}(S)$: $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i$.

(b). Affine combination $\text{aff}(S)$: $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i, \sum_{i=1}^{k} \lambda_i = 1$.

(c). Conic combination $\text{cone}(S)$: $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i, \lambda_i \geq 0, \forall i$.

(d). Convex combination $\text{conv}(S)$: $\bar{x} = \sum_{i=1}^{k} \lambda_i x^i, \sum_{i=1}^{k} \lambda_i = 1, 0 \leq \lambda \leq 1, \forall i$.

Definition 1.11 (Linearly/affinely independent)**.** The vectors $x^1, x^2, \ldots, x^k$ are linearly independent if $\sum_{i=1}^{k} \lambda_i x^i = 0$ has a unique solution $\lambda_i = 0, \forall i$. The vectors $x^1, x^2, \ldots, x^k$ are affinely independent if $\sum_{i=1}^{k} \lambda_i x^i = 0, \sum_{i=0}^{k} \lambda_i = 0$ has a unique solution $\lambda_i = 0, \forall i$.

Proposition 1.12. $x^1, \ldots, x^k$ are affinely independent iff. $x^1 - x^k, \ldots, x^{k-1} - x^k$ are linearly independent.

Proposition 1.13. $x^1, \ldots, x^k$ are affinely independent iff. $(1, x^1), \ldots, (1, x^k)$ are linearly independent.

Theorem 1.14. The intersection of convex sets is convex.

Theorem 1.15. Every polyhedron is a convex set.

Theorem 1.16. A convex combination of a finite number of vectors in a convex set also belongs to that set.

Theorem 1.17. The convex hull of a finite number of vectors is a convex set.

Definition 1.18 (Ray)**.** A ray is given by $x_0 + \lambda d, \lambda \geq 0$, for a given point x_0 and a direction vector $d \in \mathbb{R}^n$ s.t. $d \neq \mathbf{0}$.

Definition 1.19 (Recession direction)**.** A direction $d \in \mathbb{R}^n, d \neq \mathbf{0}$ is called a recession direction of a convex set S, if $\forall x \in S$, we have that $x + \lambda d \in S, \forall \lambda \geq 0$.

Definition 1.20 (Extreme point)**.** A point $x \in S$ is an extreme point of S, if it's not a <u>convex</u> combination of any $x^1, x^2 \in S, x^1 \neq x^2 \neq x$.

Definition 1.21 (Extreme ray)**.** A direction d is an extreme ray of S if it is not a <u>conical</u> combination of any $d^1 \neq d^2 \neq d$ that are rays of S.

Definition 1.22 (Active/binding/tight)**.** Consider $P \in \mathbb{R}^n$ is defined by constraints that are inequalities and equalities. If a given point $\bar{x}$ satisfies a constraint at equality, then the constraint is active/binding/tight at $\bar{x}$.

Definition 1.23 (Basic solution (b.s.) & basic feasible solution (b.f.s.))**.** Let $P \in \mathbb{R}^n$ be a polyhedron. The vector $\bar{x}$ is a b.s. to P if <u>all equality constraints</u> are active at $\bar{x}$, and there are <u>n linearly independent constraints</u> active at $\bar{x}$. Furthermore, if $\bar{x} \in P$, then $\bar{x}$ is also a b.f.s. to P.

Definition 1.24 (Degenerate)**.** A b.s. $\bar{x}$ is degenerate if there are more than n constrains active at $\bar{x}$.

Definition 1.25 (Adjacent)**.** Two distinct b.s. $\bar{x}, \bar{y}$ are adjacent if the $n-1$ linearly independent constraints active at $\bar{x}$ and $\bar{y}$ are common, and one active constraint is not common.

Theorem 1.26. Let $P = \{x \in \mathbb{R}^n : Ax \geq b\}, P \neq \emptyset$. Then $\bar{x} \in P$ is an extreme point of P iff. $\bar{x}$ is a b.f.s.

Definition 1.27 (Line)**.** A polyhedron $P \subset \mathbb{R}^n$ contains a line, if there exists a vector $x \in P$ and direction $d \in \mathbb{R}^n, d \neq \mathbf{0}$ s.t. $x + \lambda d \in P, \forall \lambda \in \mathbb{R}$.

Theorem 1.28. Let $P \subset \mathbb{R}^n$ be a polyhedron and $P \neq \emptyset$. Then P does not contain a line iff. P has at least one extreme point.

Theorem 1.29. If there exists an optimal solution to an LP that has an extreme point (i.e., the LP is feasible and it has a bounded optimal solution), then there exists an optimal solution that is an extreme point.

2 The Simplex Method

Theorem 2.1. Consider $P = \{x \in \mathbb{R}^n : Ax = b, x_j \geq 0, j = 1, \ldots, n\}$, where A is an $m \times n$ matrix with $m \leq n$ and $\text{rank}(A) = m$. Suppose $P \neq \emptyset$. We say that $\bar{x}$ is a b.s. iff. $\underline{A\bar{x} = b}$ and there exists <u>m linearly independent columns</u> $A^{B(1)}, \ldots, A^{B(m)}$ of A, with $\underline{\bar{x}_j = 0, \forall j \notin \{B(1), \ldots, B(m)\}}$. The variables $x_j, j \in \{B(1), \ldots, B(m)\}$ are called <u>basic</u> variables, and the other variables are called <u>nonbasic</u>.

Definition 2.2 (Convex)**.** A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if for every $x, y \in \mathbb{R}^n, \lambda \in [0, 1]$, we have $f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda) f(y)$.

Definition 2.3 (Convex optimization problems)**.** Convex optimization problem is to minimize a <u>convex function</u> over a <u>convex set</u>.

Definition 2.4 (Local/global optima)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function, and $S \subset \mathbb{R}^n$ be a convex set. A given point $x^* \in S$ is a local optimum for minimizing $f(x)$ over $x \in S$ if $\exists \epsilon > 0$ s.t. $\forall x \in S$ with $\|x^* - x\| \leq \epsilon$, we have $f(x^*) \leq f(x)$. The point x^* is also a global optimum if $f(x^*) \leq f(x), \forall x \in S$.

Definition 2.5 (Feasible direction)**.** Let $x \in P$. A vector $d \in \mathbb{R}^n, d \neq \mathbf{0}$ is called a feasible direction at x if $\exists \theta > 0$ for which $x + \theta d \in P$.

Definition 2.6 (Simplex Method)**.** The simplex method is as follows.

Algorithm 1 Simplex Method

Input: A standard form LP: min $z = c^\top x$ s.t. $Ax = b, x \geq 0$. Initial b.f.s. x.

while TRUE **do**
- Let $x = (x_B, x_N)$, $c = (c_B, c_N)$, $d = (d_B, d_N)$, $B = [A_{B(1)}, \ldots, A_{B(m)}]$ be the basis matrix.
- Calculate the reduced cost of variable x_j: $\bar{c}_j = c_j - c_B^\top B^{-1} A_j$.
- Steepest descent rule: Choose the variable x_j with the smallest $\bar{c}_j (< 0)$.
- **if** $\bar{c} \geq \mathbf{0}$ **then**
  - $z^* \leftarrow c^\top x$. // Optimal solution
  - **break**.
- **end if**
- Calculate the feasible direction: $d_B \leftarrow -B^{-1} A_j, d_N \leftarrow e_j$.
- **if** $d_B \geq \mathbf{0}$ **then**
  - $z^* \leftarrow -\infty$. // Unbounded
  - **break**.
- **end if**
- Ratio test: $\theta^* = \min_{i=1,\ldots,m: d_{B(i)} < 0} -x_{B(i)}/d_{B(i)}$.
- Update: $x \leftarrow x + \theta^* d$

end while

Output: Optimal solution x if exists.

Theorem 2.7 (Revised simplex)**.** $[B_i^{-1}|u] \to [B_{i+1}^{-1}|\mathbf{e}_\ell]$, where $u = -d_B = B_i^{-1} A_\ell$, and ℓ is s.t. $x_{B(\ell)}$ leaves the basis. Perform row operations on B_i^{-1}, with the ℓ-th row as the pivot row, until we turn u into $\mathbf{e}_\ell$. Notice that the order of variables in c_B must be the same as that of B^{-1}.

Remark 2.8 (Full Tableau Implementation of simplex)**.** The tableau is in the following form.

$-z$	x_B	x_N
$-c_B^\top B^{-1} b$	$\mathbf{0}$	$c_N^\top - c_B^\top B^{-1} N$
$x_B = B^{-1} b$	$\boldsymbol{I}$	$B^{-1} N$

Theorem 2.9. Assume that the LP is feasible, and every b.f.s. is nondegenerate. Then the simplex method stops after a finite number of iterations, with an optimal b.f.s. and a finite objective function value, or with a vector d s.t. $Ad = 0, d \geq 0$ and $c^\top d < 0$, and $-\infty$ as the optimal cost.

Definition 2.10 (Lexicographical larger (smaller))**.** A vector $u \in \mathbb{R}^n$ is said to be lexicographical larger (or smaller) than another vector $v \in \mathbb{R}^n$, if $v \neq u$ and the first nonzero component of $u - v$ is positive (or negative, respectively). We denote this by $u >_L v$ (or $u <_L v$, respectively).

Definition 2.11 (Bland's (smallest superscript) rule)**.** Find the smallest j with $\bar{c}_j < 0$ and let column A_j enter the basis. Out of all variables x_i that are tied for the ratio test, choose the one with the smallest i.

Definition 2.12 (Two-phase Method)**.** The two-phase method is as follows.

Algorithm 2 Two-phase Method

Require: A standard form LP: $\min\ z = c^\top x$, s.t. $Ax = b, x \geq 0, b \geq 0$.

1: Phase I: Add artificial variables y_i to row i, when necessary, i.e., $\min\ w = \sum y_i$, s.t. $Ax + y = b, x, y \geq 0$, which is easy to solve. (Remark: maybe you need one more iteration when facing degeneracy)

2: Phase II: If $w = 0$ (feasible), then solve the problem with original objective starting from the initial b.f.s. from Phase I.

Definition 2.13 (Big-M Method)**.** The big-M method is as follows.

Algorithm 3 Big-M Method

Require: A standard form LP: $\min\ z = c^\top x$, s.t. $Ax = b, x \geq 0, b \geq 0$.

1: $\min\ c^\top x + M \sum y_i$, s.t. $Ax + y = b, x, y \geq 0$.

---

3 Duality

Definition 3.1 (Dual of any LP)**.** See below.

$$
\begin{array}{rlcrl}
 & \text{Primal} & & & \text{Dual} \\
\min & c^\top x & \Leftrightarrow & \max & p^\top b \\
a_i x \geq b_i, & i \in A^{\geq} & \Leftrightarrow & p_i \geq 0, & i \in A^{\geq} \\
a_i x \leq b_i, & i \in A^{\leq} & \Leftrightarrow & p_i \leq 0, & i \in A^{\geq} \\
a_i x = b_i, & i \in A^{=} & \Leftrightarrow & p_i \text{ u.r.s.}, & i \in A^{=} \\
x_j \text{ u.r.s.} & j \in M^u & \Leftrightarrow & p^\top A_j = c_j, & j \in M^u \\
x_j \geq 0, & j \in M^{\geq} & \Leftrightarrow & p^\top A_j \leq c_j & j \in M^{\geq} \\
x_j \leq 0, & j \in M^{\leq} & \Leftrightarrow & p^\top A_j \geq c_j & j \in M^{\leq}
\end{array}
$$

Remark 3.2 (Dual form of the standard form)**.** If we consider the standard form, then we have

Primal		Dual	
min	$c^\top x$	max	$p^\top b$
subject to	$Ax = b$	subject to	$p^\top A \le c^\top$
	$x \ge 0$		

Remark 3.3. We have $u_i = p_i(a_i x - b_i) \ge 0, i = 1, \dots, m$, and $v_j = x_j(c_j - p^\top A_j) \ge 0, j = 1, \dots, n$.

Remark 3.4 (Economic interpretation)**.** Consider the primal problem: given the amount of resources available and required by producing n kinds of products, and the benefits of selling each unit of each product, to maximize the total profit.

The dual problem is then given by: Another company wants to buy all those resources, the minimum price should it pay to make the seller satisfied.

Theorem 3.5 (Weak Duality)**.** If x is a feasible solution to the primal (minimization) problem, and p is a feasible solution to the dual (maximization) problem, then $p^\top b \le c^\top x$.

Theorem 3.6 (Strong Duality)**.** If an LP has an optimal solution, so does its dual, and the respective costs are equal.

Remark 3.7 (Primal-Dual Relationship)**.** See below.

Dual \ Primal	Finite Optimum	Unbounded	Infeasible
Finite Optimum	✓		
Unbounded			✓
Infeasible		✓	✓

Theorem 3.8 (Complementary Slackness Theorem)**.** Let x and p be feasible solutions to the primal and the dual, respectively. Then the following two statements are equivalent:

(1) x and p are optimal solutions to the primal and the dual, respectively.

(2) $p_i(a_i x - b_i) = 0, i = 1, \dots, m$ and $x_j(c_j - p^\top A_j) = 0, j = 1, \dots, n$.

Lemma 3.9 (Farkas' Lemma)**.** Exactly one of the following must hold for the system $Ax = b, x \ge 0$:

(1) $\exists x \in \mathbb{R}^n : Ax = b, x \ge 0$,

(2) $\exists p \in \mathbb{R}^m : p^\top A \le \mathbf{0}$ and $p^\top b > 0$.

Lemma 3.10 (A Variant of Farkas' Lemma)**.** Exactly one of the following must hold for the system $Ax \leq b$:

(1) $\exists x \in \mathbb{R}^n : Ax \leq b$,

(2) $\exists p \in \mathbb{R}^m_+ : p^\top A = \mathbf{0}$ and $p^\top b < 0$.

Definition 3.11 (Recession Cone, recession direction (ray))**.** Let $P = \{x \in \mathbb{R}^n : Ax \geq b\} \neq 0$, then $P_0 = \{d \in \mathbb{R}^n : Ad \geq \mathbf{0}\}$ is called a recession cone of P. A direction $d \in P_0 \backslash \{\mathbf{0}\}$ is called a recession direction (ray) of P.

Theorem 3.12. The LP $\min c^\top x : Ax = b, x \geq 0$ is unbounded, iff $\exists d \in P_0$ with $c^\top d < 0$.

Theorem 3.13 (Representation of Polyhedra)**.** Let $P = \{x \in \mathbb{R}^n : Ax \geq b\} \neq \emptyset$ have at least one extreme point. Let $x^1, \ldots, x^k$ be all extreme points of P, and let $r^1, \ldots, r^\ell$ be all extreme rays of P. Let $Q = \{x : x = \sum_{i=1}^k \lambda_i x^i + \sum_{j=1}^\ell \mu_j r^j, \sum_{i=1}^k \lambda_i = 1, \lambda, \mu \geq 0\}$. Then $Q = P$.

Remark 3.14 (Sensitivity Analysis)**.** Changes we consider in sensitivity Analysis

(1) Objective function coefficient of a non-basic variable. If $\bar{c}_j = c_j + \Delta - c_B^\top B^{-1} A_j \geq 0$, current solution is still optimal. Otherwise, continue simplex with x_j entering the basis.

(2) Objective function coefficient of a basic variable. Recalculate $\bar{c}_N = c_N - (c_B + \Delta)^\top B^{-1} N$.

(3) The column of a non-basic variable. Recalculate $\bar{c}_j = \widehat{c}_j - c_B \top B^{-1} \widehat{A}_j$.

(4) The right-hand side b / Adding a new variable / Adding a new constraint. Dual Simplex.

Definition 3.15 (Dual simplex method)**.** The dual simplex method is given below.

Algorithm 4 Dual simplx method

Require: A primal infeasible, dual feasible solution.

1: **while** $\exists x_{B(\ell)} < 0$ **do**
2: Choose a primal infeasible basic variable $x_{B(\ell)} < 0$ to leave the basis (ℓ-th row is the pivot row).
3: Dual ratio test: Let $j = \text{argmin}_{i:u_i<0}\{-\bar{c}_i/u_i\}$, x_j enters the basis.
4: **end while**

4 Minimum Cost Network Flow Problem (MCNF)

Definition 4.1 (MCNF)**.** Given a directed graph $G = (V, A)$ where V is the set of nodes (vertices), and A is the set of arcs, the MCNF is formulated as

$$\begin{aligned} \min \sum_{(i,j)\in A} & c_{ij}x_{ij} \\ \text{s.t.} \sum_{j:(j,i)\in A} x_{ji} - & \sum_{j:(i,j)\in A} x_{ij} = b_i, \quad \forall i \in V \\ 0 \leq x_{ij} & \leq u_{ij}, \quad \forall (i,j) \in A. \end{aligned}$$

Here c_j denotes the unit cost of flow on arc $(i,j) \in A$, u_{ij} denotes the capacity of total flow on arc $(i,j) \in A$, and b_i denotes the amount of demand/supply at node $i \in V$ ($b_i > 0$: positive demand; $b_i < 0$: positive supply; $b_i = 0$: transshipment node).

Definition 4.2 (Transportation problem)**.** Every node is either a supply node or a demand node.

Definition 4.3 (Bipartite graph)**.** The only arcs are from a supply node to a demand node.

Definition 4.4 (Assignment problem)**.** Assigning n tasks to n workers. Every task must be assigned to exactly one worker and vice versa. We have $b_j = 1$ for each worker node j and $b_i = -1$ for each task node i

Definition 4.5 (Shortest path problem (SP))**.** Given a graph $G = (V, A)$, a source node s, a sink node t, and distances d_{ij} on arcs (i, j), find the shortest path from s to t.

Definition 4.6 (Maximum flow problem (MF))**.** Given a graph $G = (V, A)$, a source node s, a sink node t, and capacities u_{ij} on arcs i, j, find the maximum amount of flow that can be sent from s to t. It's formulated as $b_i = 0$ and $c_{ij} = -1$ if $(i, j) = (t, s)$ and $c_{ij} = 0$ otherwise.

Definition 4.7 (Totally unimodular matrix)**.** Matrix A is totally unimodular (TU) if every square submatrix has a determinant $0, 1$, or -1.

Theorem 4.8. If A is TU, then every b.f.s. of $Ax = b, x \geq 0$ is integral for integral b.

Theorem 4.9 (Crammer's Rule)**.** Consider $Ax = b$, where $A \in \mathbb{R}^{n\times n}, \det(A) \neq 0, x \in \mathbb{R}^n, b \in \mathbb{R}^n$. Then $x_i = \det(A^i)/\det(A), i = 1, 2, \ldots, n$, where A^i is the matrix formed by replacing the i-th column by the column vector b.

Theorem 4.10 (Laplace expansion)**.** Suppose $B \in \mathbb{R}^{n\times n}$ and choose a fixed $i \in \{1, 2, \ldots, n\}$. Then $\det(B) = \sum_{j=1}^{n}(-1)^{i+j} b_{ij} \det(M_{ij})$, where $\det(M_{ij})$ is the minor of element B_{ij}, i.e. the determinant of the submatrix M_{ij} formed by removing the i-th row and the j-th column of matrix B.

Lemma 4.11. If A is TU, then $A^\top$ is TU.

Theorem 4.12. If A is TU, then $\begin{bmatrix} A_{m\times n} \\ I_{m\times m} \; 0_{m\times(n-m)} \end{bmatrix}$ is also TU.

Theorem 4.13. The node-arc incidence matrix corresponding to the flow balance equations is TU.

Remark 4.14 (Other efficient (polynomial) algorithms for LP)**.** Although in practice, simplex works really well (average case), in the worst case (e.g. the pathological example), simplex could take an exponential number of iterations. There are other efficient (polynomial) algorithms for LP.

(1) Ellipsoid method runs for $O(n^6 \log(nU))$ iterations in the worse case, where U is the largest integer data in A, b, but the average case performance is its worst case performance.

(2) Interior point method is both theoretically and practically efficient.

5 Parametric LP

Remark 5.1 (Global dependence on the cost vector c)**.** Consider the parametric LP, where A and b are fixed, but the vector c varies. Let $P = \{x : x \in \mathbb{R}^n, Ax = b, x \geq 0\}$. The LP objective $z(c) = \min_{x\in P}\{c^\top x\}$ is a function of c.

Given the extreme points of P: $x_1, x_2, \ldots, x_\ell$, it holds that $z(c) = \min_{i=1,\ldots,\ell}\{c^\top x_i\}$.

Remark 5.2 (The dependence of the optimal cost on the vector b)**.** Let $P(b) = \{x : x \in \mathbb{R}^n, Ax = b, x \geq 0\}$ be the feasible set defining a standard form LP. Let $S = \{b : b \in \mathbb{R}^m, P(b) \neq \emptyset\}$ be the feasible domain of $P(b)$. For any $b \in S$, the objective function $z(b) = \min\{c^\top x : x \in P(b)\}$ is a function of b (useful in large-scale optimization and stochastic programming).

Suppose that the extreme points of the dual polyhedron are given by $p_1, p_2, \ldots, p_k$. Then $z(b) = w(b) = \max_{i=1,\ldots,k}\{p_i^\top b\}$.