Optimization Theory and Algorithm II **September 22, 2022**

Lecture 5

Lecturer:Xiangyu Chang *Scribe: Xiangyu Chang*

Edited by: Xiangyu Chang

1 Standard Form of Linear Programming

Recall: Linear Programming.

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x}, \\ s.t. \quad & A\mathbf{x} = \mathbf{b}, \\ & \mathbf{x} \succeq 0. \end{aligned}$$

Why we need the standard form? 1. For designing algorithm uniformly. 2. Any linear programming can be transformed into the standard form.

For example, consider the following problem:

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x}, \\ s.t. \quad & A\mathbf{x} \preceq \mathbf{b}. \end{aligned}$$

It is equivalent to

$$\begin{aligned} \min_{\mathbf{x}} \quad & \mathbf{c}^\top \mathbf{x}, \\ s.t. \quad & A\mathbf{x} + \mathbf{s} = \mathbf{b}, \\ & \mathbf{s} \succeq 0. \end{aligned}$$

Then, let $\mathbf{x} = \mathbf{x}^+ - \mathbf{x}^-$, where $\mathbf{x}^+ = \max\{\mathbf{x}, 0\}$ and $\mathbf{x}^- = \max\{-\mathbf{x}, 0\}$. Finally, it has

$$\begin{aligned} \min_{\mathbf{x}} \quad & [\mathbf{c}^\top, -\mathbf{c}^\top, 0] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix} \\ s.t. \quad & [A, -A, I] \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix} = \mathbf{b}, \\ & \begin{bmatrix} \mathbf{x}^+ \\ \mathbf{x}^- \\ \mathbf{s} \end{bmatrix} \succeq 0. \end{aligned}$$

Remark 1

- *We say the linear program is infeasible if the feasible set is empty.*
- *The LP problem is unbounded if the objective function is unbouned below on the feasible region. That is there exists $\{\mathbf{x}^t\}$, such that $\mathbf{c}^\top \mathbf{x}^t \to -\infty$ as $t \to \infty$.*

- *Generally, assume $A \in \mathbb{R}^{m\times n}, m < n$ is full row rank.*

The Lagrangian is
$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = \mathbf{c}^\top \mathbf{x} + \boldsymbol{\nu}^\top (A\mathbf{x} - \mathbf{b}) - \boldsymbol{\lambda}^\top \mathbf{x}.$$
The KKT conditions of the standard linear programming are

$$\begin{aligned}
A^\top \boldsymbol{\nu} + \mathbf{c} &= \boldsymbol{\lambda}, \\
A\mathbf{x} &= \mathbf{b}, \\
\mathbf{x} &\succeq 0, \\
x_i \lambda_i &= 0, \\
\boldsymbol{\lambda} &\succeq 0.
\end{aligned}$$

The dual function is $g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) = -\boldsymbol{\nu}^\top \mathbf{b}$, such that $\mathbf{c} + A^\top \boldsymbol{\nu} = \boldsymbol{\lambda}$. Then the dual problem is

$$\max_{\boldsymbol{\nu}} \; -\boldsymbol{\nu}^\top \mathbf{b} \tag{1}$$
$$s.t. \; \mathbf{c} + A^\top \boldsymbol{\nu} \succeq 0. \tag{2}$$

This is equivalent to

$$\max_{\boldsymbol{\nu}} \; \boldsymbol{\nu}^\top \mathbf{b} \tag{3}$$
$$s.t. \; \mathbf{c} - A^\top \boldsymbol{\nu} \succeq 0. \tag{4}$$

Theorem 1
- *If either primal problem or dual problem of LP has a finite solution, then so does the other, and the objective value are equal (strong duality).*
- *If either primal or dual problem of LP is unbounded, then the other problem is infeasible.*

2 Geometry of the Feasible Set

Definition 1 *Feasible domain: $P = \{\mathbf{x} \in^n | \; A\mathbf{x} = \mathbf{b}, \mathbf{x} \succeq 0\}$.*

Remark 2
- *Hyperplane: $\mathbf{a}^\top \mathbf{x} = \beta$, with $\mathbf{a}, \mathbf{x} \in^n$. Closed Half space: $\mathbf{a}^\top \mathbf{x} \leq \beta$.*
- *Intersection of a finite number of hyperplanes is a polyhedron. Bounded polyhedron is a polytope.*
- *Feasible domain is a convex polyhedron due to the intersection of $\mathbf{a}_i^\top \mathbf{x} \geq b_i$, $\mathbf{a}_i^\top \mathbf{x} \leq b_i$ and $x_i \geq 0, i = 1, \ldots, m$.*

Definition 2 *Extreme point of P is the point that can not be expressed by the convex combination of other points of P.*

Theorem 2 *P is convex polyhedral and $\mathbf{x} \in P$ is a vertex if and only if $\mathbf{x}$ is a extreme point of P.*

Proposition 1 *The polyhedron P has only a finite number of extreme points.*

Definition 3 *A vector $\mathbf{d}$ is an extreme direction of P, if $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{x}^0 + \lambda \mathbf{d}, \lambda > 0\} \subset P$ for all $\mathbf{x}^0 \in P$.*

Theorem 3 (Resolution Theorem) *Let $V = \{\mathbf{v}_i \in \mathbb{R}^n \mid i \in I\}$ be the set of all extreme point of P and I is a finite index set. Then $\forall \mathbf{x} \in P$, we have*

$$\mathbf{x} = \sum_{i \in I} \lambda_i \mathbf{v}_i + \lambda \mathbf{d}, \tag{5}$$

where

$$\sum_i \lambda_i = 1, \lambda_i \geq 0, \lambda \geq 0,$$

and either $\mathbf{d} = 0$ or $\mathbf{d}$ is a extreme direction.

3 Managing Extreme Points Algebraically

Theorem 4 *$\mathbf{x} \in P$ is a extreme point of P if and only if columns of A with respect to positive x_i are linearly independent.*

Proof 1 *Denote that*

$$\mathbf{x} = \begin{bmatrix} \bar{\mathbf{x}} \\ 0 \end{bmatrix} \text{ with } \bar{\mathbf{x}} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix} > 0, \text{ and } \bar{A} = [A_1, \ldots, A_p]. \tag{6}$$

It is easy to check that $A\mathbf{x} = \bar{A}\bar{\mathbf{x}} = \mathbf{b}$.

Proof by contradiction. Assume that $\mathbf{x}$ is an extreme point but $\bar{A}$ is linearly dependent. Since $\bar{A}$ is linearly dependent, there exist a $\bar{\mathbf{w}} \neq 0$ such that $\bar{A}\mathbf{w} = 0$. Therefore, there exist a small number ϵ such that $\bar{\mathbf{x}} \pm \epsilon\bar{\mathbf{w}} \geq 0$ and $\bar{A}(\bar{\mathbf{x}} \pm \epsilon\bar{\mathbf{w}}) = \bar{A}\bar{\mathbf{x}} = \mathbf{b}$. Letting

$$\mathbf{y}_1 = \begin{bmatrix} \bar{\mathbf{x}} + \epsilon\bar{\mathbf{w}} \\ 0 \end{bmatrix}, \text{ and } \mathbf{y}_2 = \begin{bmatrix} \bar{\mathbf{x}} - \epsilon\bar{\mathbf{w}} \\ 0 \end{bmatrix}.$$

It is easy to check that $\mathbf{x} = \frac{\mathbf{y}_1 + \mathbf{y}_2}{2}$ and $y_1, y_2 \in P$. That is $\mathbf{x}$ can be expressed by the convex combination of $\mathbf{y}_1$ and $\mathbf{y}_2$, which contradicts with the fact that $\mathbf{x}$ is an extreme point of P.

Now we assume that $\bar{A}$ is linearly independent but $\mathbf{x}$ is not an extreme point of P. Then we can represent $\mathbf{x}$ as

$$\mathbf{x} = \lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2, \ \mathbf{y}_1 \neq \mathbf{y}_2 \ \lambda \in (0,1), \ \mathbf{y}_1, \mathbf{y}_2 \geq 0.$$

By the form of $\mathbf{x}$ shown in Eqn. (6), it holds that

$$\mathbf{y}_1 = \begin{bmatrix} \bar{\mathbf{y}}_1 \\ 0 \end{bmatrix}. \tag{7}$$

Now,

$$\mathbf{x} - \mathbf{y}_1 = \lambda\mathbf{y}_1 + (1-\lambda)\mathbf{y}_2 - \mathbf{y}_1 = -(1-\lambda)(\mathbf{y}_1 - \mathbf{y}_2) \neq 0 \tag{8}$$

where the last inequality is because $\mathbf{y}_1 \neq \mathbf{y}_2$ and $\lambda < 1$. Therefore

$$A(\mathbf{x} - \mathbf{y}_1) = \bar{A}(\bar{\mathbf{x}} - \bar{\mathbf{y}}_1) = \mathbf{b} - \mathbf{b} = 0,$$

which contradicts the assumption A is linearly independent.

Let A be an $m \times n$ matrix with, we say A has full rank (full row rank) if A has m linearly independent columns. In this, we can rearrange

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} \begin{matrix} \leftarrow \text{ basic variables} \\ \leftarrow \text{non-basic variables} \end{matrix} A = [\underbrace{B}_{\text{Basis}}, \underbrace{N}_{\text{non-basis}}]. \tag{9}$$

Definition 4 *If we set $\mathbf{x}_N$ to zero and $\mathbf{x}_B$ is the solution of $B\mathbf{x}_B = b$, then we say $\mathbf{x}$ is a basic solution. If $\mathbf{x}_B \geq 0$, then $\mathbf{x}$ is a basic feasible solution.*

Proposition 2 *A point $\mathbf{x}$ in P is an extreme point of P if and only if $\mathbf{x}$ is a basic feasible solution corresponding to some basis B.*

Theorem 5 *(Fundamental Theorem of LP) For a standard form LP, if its feasible domain P is nonempty, then the optimal objective value of $z = \mathbf{c}^\top \mathbf{x}$ over P is either unbounded below, or it is attained at (at least) an extreme point of P.*

Proof 2 *By the resolution theorem, there are two cases:*

Case 1, P has an extreme direction $\mathbf{d}$ such that $\mathbf{c}^\top \mathbf{d} < 0$. Then P is unbounded and $z \to -\infty$.

Case2, P does not have an extreme direction $\mathbf{d}$ such that $\mathbf{c}^\top \mathbf{d} < 0$. Then $\forall \mathbf{x} \in P$, either $\mathbf{x} = \sum_i \lambda_i \mathbf{v}_i$ or $\mathbf{x} = \sum_i \lambda_i \mathbf{v}_i + \lambda \bar{\mathbf{d}}$ with $\mathbf{c}^\top \bar{\mathbf{d}} \geq 0$.

In both cases, it holds that

$$\begin{aligned} \mathbf{c}^\top \mathbf{x} =& \mathbf{c}^\top \left(\sum_i \lambda_i \mathbf{v}_i \right) + \lambda \mathbf{c}^\top \bar{\mathbf{d}} \\ \geq& \sum_i \lambda_i (\mathbf{c}^\top \mathbf{v}_i) \\ \geq& \min_i \mathbf{c}^\top \mathbf{v}_i \\ =& \mathbf{c}^\top \mathbf{v}_{min}. \end{aligned}$$

3.0.1 Simplex Method

We know that the optimal points of LP are extreme points. And the extreme points are basic feasible points. So, we can use the property of BFP to build the simplex method. The basic idea is

- give a BFS $\mathbf{x}^t$.
- Find another BFS by $\mathbf{x}^{t+1} = \mathbf{x}^t + \lambda \mathbf{d}$.
- Check the objective function increasing or decreasing.

Basic idea: give a BFP $\mathbf{x}$, that is $A\mathbf{x} = \mathbf{b}, \mathbf{x} \succeq 0$. In addition, $\mathbf{x} = \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix}, A = [B, N], B\mathbf{x}_B = \mathbf{b}, \mathbf{x}_N = 0, B = [A_1, \ldots, A_m], N = [A_{m+1}, \ldots, A_n], \mathbf{x}_B = (x_1, \ldots, x_m)^\top$.

Denote the fundamental matrix of LP

$$M = \begin{bmatrix} B & N \\ 0 & I \end{bmatrix}. \tag{10}$$

So,

$$M^{-1} = \begin{bmatrix} B^{-1} & -B^{-1}N \\ 0 & I \end{bmatrix}. \tag{11}$$

It is easy to check that

$$M\mathbf{x} = \begin{bmatrix} B & N \\ 0 & I \end{bmatrix} \begin{bmatrix} \mathbf{x}_B \\ \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \mathbf{b} \\ 0 \end{bmatrix}. \tag{12}$$

Let $x_q, q \in [m+1, n]$ be one of the non-basis variables. A_q is the corresponding column of A, and e_q is the corresponding column of $I \in \mathbb{R}^{(n-m)^2}$.

Denote that

$$\mathbf{d}_q = \begin{bmatrix} -B^{-1}A_q \\ e_q \end{bmatrix}, \tag{13}$$

then we can check that $\mathbf{x}(\lambda) = \mathbf{x} + \lambda \mathbf{d}_q$ is a feasible point for $A\mathbf{x}(\lambda) = \mathbf{b}$. That is

$$A\mathbf{x}(\lambda) = A(\mathbf{x} + \lambda \mathbf{d}_q) = A\mathbf{x} + [B, N] \begin{bmatrix} -B^{-1}A_q \\ e_q \end{bmatrix} = A\mathbf{x} = \mathbf{b}. \tag{14}$$

Let choose a proper step size λ that can achieve the feasibility of $\mathbf{x}(\lambda) \succeq 0$. We know that

$$\mathbf{x}(\lambda) = (x_1 + \lambda d_{q1}, \ldots, x_n + \lambda d_{qn}).$$

Definition 5 (reduced cost) *The quantity of $r_q = \mathbf{c}^\top \mathbf{x}(\lambda) - \mathbf{c}^\top \mathbf{x} = \mathbf{c}^\top \mathbf{d}_q = \mathbf{c}_q - \mathbf{c}_B^\top B^{-1} A_q$ is called a reduced cost with respect to the variable $\mathbf{x}_q$.*

Theorem 6 *If $\mathbf{x} = [B^{-1}b; 0]$ is a basic feasible solution with B and $r_q < 0$, for some non-basic variable x_q, then $\mathbf{d}_q = [-B^{-1}A_q; e_q]$ leads to an improved objective function.*

Theorem 7 *If $\mathbf{x}$ is a basic feasible solution with $r_q \geq 0$ for all non-basic variables, then $\mathbf{x}$ is optimal solution.*

Proof 3 $\mathbf{x}$ *is local optimum. Since linear programming is a convex optimization problem, the local optimum is the global one.*

We need to refine our question: How to choose a proper step size λ that can achieve the feasibility of $\mathbf{x}(\lambda) \succeq 0$ when $r_q < 0$.

Then we have two cases:

- If $\mathbf{d}_q \succeq 0$, and it is a extreme directoin of P, we know that $r_q = \mathbf{c}^\top \mathbf{x}(\lambda) - \mathbf{c}^\top \mathbf{x} = \mathbf{c}^\top \mathbf{d}_q < 0$, then according to the resolution theorem, the LP is unbounded below.
- If there exists $i \in \{1, \ldots, m\}$, $d_{iq} < 0$, $\lambda = \min_{i \in \{1,\ldots,m\}} \left\{ \frac{x_i}{-d_{iq}} \mid d_{iq} < 0 \right\}$.

Revisit the Simplex method by an example:

$$\begin{aligned} \min\ & -4x_1 - 2x_2 && (15) \\ \text{s.t. } & x_1 + x_2 + x_3 = 5, && (16) \\ & 2x_1 + 1/2x_2 + x_4 = 8, && (17) \\ & \mathbf{x} \succeq 0. && (18) \end{aligned}$$

In this case, we have

$$A = \begin{bmatrix} 1 & 1 & 1 & 0 \\ 2 & 1/2 & 0 & 1 \end{bmatrix} \tag{19}$$

The basis $\mathcal{B} = \{3, 4\}$, for which we have

$$x_B = \begin{bmatrix} x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 5 \\ 8 \end{bmatrix} \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 1 \\ 2 & 1/2 \end{bmatrix}. \tag{20}$$

It holds that

$$-B^{-1}N = \begin{bmatrix} -1 & -1 \\ -2 & -1/2 \end{bmatrix} \quad d_1 = \begin{array}{cc} \begin{bmatrix} -1 \\ -2 \\ 1 \\ 0 \end{bmatrix} & \begin{array}{c} x_3 \\ x_4 \\ x_1 \\ x_2 \end{array} \end{array} \quad c^\top d_1 = -4 < 0, \quad \lambda = \min\left(\frac{8}{2}, \frac{5}{1}\right) = 4. \tag{21}$$

$$x_B = \begin{array}{cc} \begin{bmatrix} 4 \\ 1 \end{bmatrix} & \begin{array}{c} x_1 \\ x_3 \end{array} \end{array} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 0 \\ 1/2 & 1 \end{bmatrix} \quad -B^{-1}N = \begin{bmatrix} \frac{1}{4} & \frac{1}{2} \\ \frac{3}{4} & -\frac{1}{2} \end{bmatrix} \tag{22}$$

$$d_2 = \begin{bmatrix} -\frac{1}{4} \\ -\frac{3}{4} \\ 1 \\ 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} \quad c^\top d_2 = 1 - 2 < 0, \quad d_4 = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_3 \\ x_2 \\ x_4 \end{bmatrix} \quad c^\top d_4 = 2 > 0 \tag{23}$$

$$x_B = \begin{array}{cc} \begin{bmatrix} \frac{11}{3} \\ \frac{4}{3} \end{bmatrix} & \begin{array}{c} x_1 \\ x_2 \end{array} \end{array} \quad B = \begin{bmatrix} 1 & 1 \\ 2 & 1/2 \end{bmatrix} \quad N = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad B^{-1}N = \begin{bmatrix} -\frac{1}{3} & \frac{2}{3} \\ \frac{4}{3} & -\frac{2}{3} \end{bmatrix} \tag{24}$$

$$c^\top d_3 = 4/3 > 0 \quad c^\top d_4 = 4/3 > 0. \tag{25}$$

Done!

We summarize the simplex method of LP in Algorithm 1.

Algorithm 1 Simplex Method of LP

1: **Input:** A BFS $\mathbf{x} \in \mathbb{R}^n$ and $A = [B, N]$
2: Step 1: Compute $r_q = \mathbf{c}^\top \mathbf{d}_q$. If $r_q \geq 0$ for all non-basis x_q, then the current BFS is optimal. Otherwise, pick up one $r_q < 0$, go to the next step.
3: Step 2: If $d_q \succeq 0$, then LP is unbounded below. Otherwise, find

$$\lambda = \min_{i \in \{1,\ldots,m\}} \left\{ \frac{x_i}{-d_{iq}} \mid d_{iq} < 0 \right\}$$

.
4: $\mathbf{x} \leftarrow \mathbf{x} + \lambda \mathbf{d}_q$ is a new BFS.
5: Update B and N, and go to Step 2.

References