Notes about Mathematical Programming

Xinhao Nie

Aarhus University

August 14, 2025

1 Linear Programming

There are two different definitions about standard form of linear programming. One is

$$\begin{aligned} \text{maximize} \quad & c^\top \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \le b \\ & \mathbf{x} \ge 0, \end{aligned}$$

while another one is:

Definition 1.1 (Primal Problem)**.**

$$\begin{aligned} \text{minimize} \quad & c^\top \mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} = b \\ & \mathbf{x} \ge 0. \end{aligned}$$

Obviously, they are euqal. Actually, it doesn't really matter whether an LP is formulated as a minimization or maximization problem. However, whether the constraints are written with $\geq$ or $=$ does affect the way the problem is understood.

When we are trying to use simplex method, the first one is more suitable because we can easily import slack varibles and do pivot rule. Things go complicated when we try to prove strong duality theorem as what we will show next.

1.1 Duality

According to minimization or maximization in prime problem, we can give corresponding dual problem as follows:

Definition 1.2 (Dual Problem)**.**

$$\begin{aligned} \text{maximize} \quad & b^\top \mathbf{y} \\ \text{subject to} \quad & A^\top \mathbf{y} \le c. \end{aligned}$$

1.1.1 Weak Duality Theorem

Many people view it as one of the most remarkable and useful features of linear programming because of the relation between optimal value in prime problem and that in dual problem.

Theorem 1.1 (Weak Duality)**.** *The optimal value in corresponding dual problem is always less than the optimal value in prime problem.*

Proof.

$$\forall \mathbf{y}, \mathbf{x}. b^\top \mathbf{y} = (A\mathbf{x})^\top \mathbf{y} = \mathbf{x}^\top A^\top \mathbf{y} \le \mathbf{x}^\top c = c^\top \mathbf{x}.$$

□

1.1.2 Strong Duality Theorem

The proof of Strong Duality Theorem depends on the standard form LP, thus, it vaies from different definitions. The following proof is based on our definition. To this end, we need soem useful lemmas in discrete optimization.

Lemma 1 (Farkas' Lemma)**.** *Let $A \in \mathbb{R}^{m\times n}$ be a matrix and $b \in \mathbb{R}^m$ be a vector. The system $Ax = b, x \ge 0$ has a solution if and only if for all $\lambda \in \mathbb{R}^m$ with $\lambda^\top A \ge 0$ one has $\lambda^\top b \ge 0$.*

[**Xinhao Nie:** There are a number of slightly different formulations of the lemma in the literature. Notice that the original one is exactly one of the above two assertion is true. Intuitively (1) is weaker, however, $\lambda = 0$ is always a trivial solution of $\lambda^\top A \ge 0$. Thus, they are equivalent.]

Proof. [**ToLearn:** It's a special variant of Separation Theorem. As a result in geometry, I believe it's straightforward but hard to prove. TODO.] □

An interesting and easy variant of Farkas' lemma is either $Ax = b$ has a solution $x \in \mathbb{R}^n$, or $A^\top y = 0$ has a solution $y \in \mathbb{R}^m$ with $b^\top y \ne 0$. The proof is quite simple. If $Ax = b$ has a solution, $\operatorname{rank} A = \operatorname{rank}[A|b]$. Thus b is a linear combination of A. Then there exists some non-trivial numbers $\lambda_1, \lambda_2, ..., \lambda_n$ s.t. $\forall i \in [m]. b_i = \lambda_1 A_{i,1} + \lambda_2 A_{i,2} + \cdots + \lambda_n A_{i,n}$. If there exists y s.t. $\forall j \in [n], y_1 A_{1,j} + y_2 A_{2,j} + \cdots + y_m A_{m,j} = 0$, we have $\sum_{i\in[m]} y_i b_i = \sum_{i\in[m]} \sum_{j\in[n]} \lambda_j y_i A_{i,j} = \sum_{j\in[n]} \lambda_j \sum_{i\in[m]} y_i A_{i,j} = 0$. Conversely, suppose there is a solution x^*. $0 = y^\top A = y^\top A x^* = y^\top b \ne 0$ leads to a contradiction. It's named as *Fredholm alternative*.

Lemma 2 (Second Variant of Farkas' Lemma)**.** *Let $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$. The system $Ax \leq b$ has a solution if and only if for all $\lambda \in \mathbb{R}^m \geq 0$ with $\lambda^\top A = 0$ one has $\lambda^\top b \geq 0$.*

Proof. $\Rightarrow$: If there is a feasible x^* s.t. $Ax^* \leq b$, we have $\forall \lambda \geq 0. \lambda^\top Ax \leq \lambda^\top b$. It implies $\lambda^\top b \geq 0$.

$\Leftarrow$: Firstly, $Ax^+ - Ax^- + z = b, x^+, x^-, z \geq 0$ has a solution if and only if $Ax \leq b$ has a solution. We change it to matrix form as $[A| - A|I_m]X = b, X \geq 0$. Using Farkas' Lemma (1), we have it has a solution if and only if $\forall \lambda \in \mathbb{R}^m$ with $\lambda[A| - A|I_m] \geq 0$ one has $\lambda^\top b \geq 0$. It's equivalent to $\lambda A = 0$ and $\lambda \geq 0$. $\square$

Theorem 1.2. *If both prime problem and dual problem are feasible, then they have same optimal value.*

Suppose there are $\mathbf{x}^*$ and $\mathbf{y}^*$ which are optimal varibales in each problem. $A\mathbf{x}^* = b$ and $\mathbf{x}^* \geq 0$ hold. Let's consider the solution space of $A\mathbf{x}^* = b$. If A is full-rank, $\mathbf{x}^* = A^{-1}b$ is a unique solution because $\mathbf{x}^*$ is a solution. In this case, the optimal value of prime problem is $c^\top A^{-1}b$. Let $\mathbf{y}' = (A^{-1})^\top c$. We can verify that $b^\top \mathbf{y}' = c^\top A^{-1}b$ and $A^\top y = A^\top (A^{-1})^\top c \leq c$. Together with Weak Duality Theorem, there is no gap between two optimal values.

Now, we consider the case that A is not full-rank. Actually, it could be proved using simplex method. The result of choosing basic variables and non-basic variables is consist of a full-rank submatrix. Using a similar analysis as above, we can conclude it. But it depends on the relation between dual variables and prime variables which also relates to the definition. Here we can give a more mathematical proof using Farkas' Lemma.

Proof. Suppose the maximal value of dual problem is δ. There is no feasible solution in $b^\top y \geq \delta + \epsilon, A^\top y \leq c$. Change it into $\begin{bmatrix} -b^\top \\ A^\top \end{bmatrix} y \leq \begin{bmatrix} -\delta - \epsilon \\ c \end{bmatrix}$. Using the second variant of Farkas' lemma in the new problem, we know that there exists $\lambda \in \mathbb{R}^{m+1} \geq 0$ with $\lambda^\top \begin{bmatrix} -b^\top \\ A^\top \end{bmatrix} = 0$ one has $\lambda^\top \begin{bmatrix} -\delta - \epsilon \\ c \end{bmatrix} < 0$. Be carefully here! Generally the negation of $\geq$ is not $<$ because it's compenant-wise operation. But here it's just one number.

Suppose $\lambda = \begin{bmatrix} \lambda_1 \\ \lambda' \end{bmatrix}$. We can show that $\lambda_1 \geq 0$. Using the second variant of Farkas' lemma in original dual problem, we have for all $\lambda' \geq 0$ with $\lambda'^\top A^\top = 0$ one has $\lambda^\top c \geq 0$. If $\lambda_1 = 0$, it conflicts with previous result. Furthermode, by scaling, we can assume $\lambda_1 = 1$. One has $\lambda'^\top A^\top = b^\top$ and $\lambda' c < \delta + \epsilon$. Here λ' is a valid variable for primal problem with objective value less than $\delta + \epsilon$. Together with weak duality theorem, we get a infimum.

$\square$

1.1.3 Complementary Slackness

Theorem 1.3. *As a conclusion of strong and weak duality theorem, we have* $\forall i \in [n].x_i^*(c - A^\top \mathbf{y}^*)_i = 0$.

Proof. $b^\top \mathbf{y}^* = c^\top \mathbf{x}^* \geq \mathbf{y}^{*\top} A\mathbf{x}^* = \mathbf{y}^{*\top} b = b^\top \mathbf{y}^*$. Thus, $\forall i \in [n].x_i^*(c - A^\top \mathbf{y}^*)_i = 0$ must hold. □

2 Semidefinite Programming

2.1 Preliminary

2.1.1 Bilinear Form

In mathematics, a bilinear form is a bilinear map $V \times V \to K$ on a vector space V over a field K. In this chapter, we restrict it's on real number field.

Definition 2.1. *A bilinear form* $f : V \times V \to \mathbb{R}$ *is a map such that:*

- $f(\alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w}) = \alpha f(\mathbf{u}, \mathbf{w}) + \beta f(\mathbf{v}, \mathbf{w})$
- $f(\mathbf{w}, \alpha\mathbf{u} + \beta\mathbf{v}) = \alpha f(\mathbf{w}, \mathbf{u}) + \beta f(\mathbf{w}, \mathbf{v})$,

where $\mathbf{v}, \mathbf{u}, \mathbf{w} \in V$ *and* $\alpha, \beta \in \mathbb{R}$.

Once we selected the basis of vector space $V = \langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle$, the bilinear form f can be represented by a $n \times n$ matrix since the bilinear property.

$$f(\mathbf{x}, \mathbf{y}) = f(\sum_i x_i\mathbf{e}_i, \sum_j y_j\mathbf{e}_j) = \sum_i x_i f(\mathbf{e}_i, \sum_j y_j\mathbf{e}_j) = \sum_i \sum_j x_i y_j f(\mathbf{e}_i, \mathbf{e}_j) = \sum_{i,j} x_i y_j f_{ij}.$$

Thus, we can rewrite $f(\mathbf{x}, \mathbf{y})$ as $\mathbf{x}^\top F\mathbf{y}$. Moreover, given the change-of-basis matrix, we can calculate corresponding matrix F. Suppose the change-of-basis matrix is $A = (a_{ij}) : \forall j \in [n].\mathbf{e}'_j = \sum_{i=1}^n a_{ij}\mathbf{e}_i$. We have $\mathbf{x} = A\mathbf{x}'$.[**Xinhao Nie:** Be careful, it's a little bit confusing.] Then,

$$\mathbf{x}'^\top F'\mathbf{y}' = f'(\mathbf{x}', \mathbf{y}') = f(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top F\mathbf{y} = (A\mathbf{x}')^\top F(A\mathbf{y}') = \mathbf{x}'^\top A^\top FA\mathbf{y}'.$$

Finally, we get the relation $F' = A^\top FA$. We call it as *congruence*. One can observe that $\operatorname{rank} F = \operatorname{rank} F'$ by the rank inequility, namely $\operatorname{rank} AB \leq \min\{\operatorname{rank} A, \operatorname{rank} B\}$, and A is full-rank. As analyzed above, the rank is the first invariant we have observed that is independent of the choice of basis. Next, we present another invariant that is also independent of the choice of basis: symmetry/skew-symmetry.

Since $f(\mathbf{x}, \mathbf{y}) = \epsilon f(\mathbf{y}, \mathbf{x})$ and $\epsilon = \pm 1$, we have

$$\mathbf{x}^\top F\mathbf{y} = f(\mathbf{x}, \mathbf{y}) = \epsilon f(\mathbf{y}, \mathbf{x}) = \epsilon f^\top(\mathbf{y}, \mathbf{x}) = \epsilon\mathbf{x}^\top F^\top\mathbf{y}.$$

In other words, $F = \epsilon F^\top$. Following from congruence, $F' = A^\top FA = \epsilon A^\top F^\top A = \epsilon F'^\top$.

If $\operatorname{char} \mathbb{R} \neq 2$, the space of bilinear forms is the direct sum of symmetric subspace and skew-symmetric subspace. The proof follows directly from a simple fact $f(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(f(\mathbf{x}, \mathbf{y}) + f(\mathbf{y}, \mathbf{x})) + \frac{1}{2}(f(\mathbf{x}, \mathbf{y}) - f(\mathbf{y}, \mathbf{x}))$.

2.1.2 Quadratic Form

Definition 2.2. *$q : V \to \mathbb{R}$ is a quadratic form, if it meets two conditions:*

1. $q(\mathbf{v}) = q(-\mathbf{v}), \forall \mathbf{v} \in V$
2. $f(\mathbf{x}, \mathbf{y}) = \frac{1}{2}(q(\mathbf{x}+\mathbf{y}) - q(\mathbf{x}) - q(\mathbf{y}))$ *is a symmetric bilinear form.*

Additionally, define $\operatorname{rank} q = \operatorname{rank} f$.

One can get the quadratic form from the corresponding bilinear form and vice versa. Notice that

$$f(\mathbf{x}, -\mathbf{x}) = \frac{1}{2}(q(\mathbf{x}-\mathbf{x}) - q(\mathbf{x}) - q(-\mathbf{x})) = \frac{1}{2}q(0) - q(\mathbf{x}),$$

$$f(\mathbf{x}, \mathbf{x}) = -f(\mathbf{x}, -\mathbf{x}) = q(\mathbf{x}) - \frac{1}{2}q(0).$$

Meanwhile, $f(0,0) = \frac{1}{2}q(0) = 0$. We have $q(\mathbf{x}) = f(\mathbf{x}, \mathbf{x}) = \mathbf{x}^\top F \mathbf{x} = \sum_{i,j} f_{ij} x_i x_j$.

Definition 2.3 (Canonical Form)**.** *For a quadratic form q, select some basis $\langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle$ such that $\forall \mathbf{x} = \sum_{i=1}^n x_i \mathbf{e}_i \in V$ one has $q(\mathbf{x}) = \sum_{i=1}^n f_{ii} x_i^2$. Then, it's the* canonical form *of q under the canonical basis $\langle \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle$.*

Theorem 2.1. *For all symmetric bilinear form f or just a symmetric matrix F, there is a canonical basis.*

Proof. Prove it by induction! When the dimension is 1, it's obvious. Assume the Theorem holds if the dimension less than n. Now, if $f = 0$, it's also obvious because it's diagonal form under every basis. If $f \neq 0$, there must be some vector $\mathbf{e}_1$ such that $q(\mathbf{e}_1) \neq 0$, otherwise for all $\mathbf{x}, \mathbf{y}, f(\mathbf{x}, \mathbf{y}) = 0$ based on definition. $f(\mathbf{e}_1, \mathbf{e}_1) = q(\mathbf{e}_1) \neq 0$.

Consider the linear function $f_1 : \mathbf{x} \to f(\mathbf{x}, \mathbf{e}_1)$. We know $\dim \ker f_1 = n - 1$ because it's linear function. Restrict f on $\ker f_1$. Using the assumption, we have a canonical basis $\langle \mathbf{e}_2, \ldots, \mathbf{e}_n \rangle$. Because they span the kernel space, $f(\mathbf{e}_i, \mathbf{e}_1) = 0, \forall i \neq 1$. Thus, $f(\mathbf{e}_i, \mathbf{e}_j) = 0, \forall i \neq j$. [**Xinhao Nie:** For non-symmetric matrix, we cannot conclude the last statement.]

Finally, we should verify they are linearly independent. By contradiction, $\mathbf{e}_1$ is linear combination of $\mathbf{e}_2, \ldots, \mathbf{e}_n$, however, f is bilinear which means $f(\mathbf{e}_1, \mathbf{e}_1) = 0$. □

Based on above theorems, we can calculate the quadratic form under the canonical basis, namely $q(\mathbf{x}) = \lambda_1 x_1^2 + \cdots + \lambda_n x_n^2$. Unfortunately, λ_i is not an invariant independent of the choice of basis vectors. For example, we swap the $\mathbf{e}_1$ and $\mathbf{e}_2$ and $\lambda_1' = \lambda_2, \lambda_2' = \lambda_1$. For another example, if $\mathbb{R}$ is real number field, we can scale the basis $\sqrt{\lambda_i}\mathbf{e}_i$ if $\lambda_i \geq 0$ and $\sqrt{-\lambda_i}\mathbf{e}_i$ if $\lambda_i < 0$. In this case, $q(\mathbf{x}) = x_1^2 + \cdots + x_s^2 - x_{s+1}^2 - \cdots - x_r^2$ where r is the rank of f and a basis-independent invariant. Naturally, we are wondering s is an invariant or not.

Theorem 2.2 (Sylvester's Law of Inertia)**.** *As above, s is a basis-independent invariant.*

Proof. Suppose $q(\mathbf{x}) = (x_1')^2 + \cdots + (x_t')^2 - (x_{t+1}')^2 - \cdots - (x_r')^2$ under some basis s.t. $t \neq s$. Without loss of generality, we suppose $t < s$. Consider two space $L_s = \langle \mathbf{e}_1, \ldots, \mathbf{e}_s \rangle$ and $L_t = \langle \mathbf{e}_{t+1}', \ldots, \mathbf{e}_n' \rangle$. We claim $L_s \cap L_t \neq 0$ because $\dim L_s \cap L_t = \dim L_s + \dim L_t - \dim \{L_s + L_t\} \geq s + n - t - n > 0$. There exists a $\mathbf{x} = \sum_{i=1}^{s} x_i \mathbf{e}_i = \sum_{j=t+1}^{n} x_j' \mathbf{e}_j' \neq 0$. Then $q(\mathbf{x}) = q(\sum_{i=1}^{s} x_i \mathbf{e}_i) > 0$ and $q(\mathbf{x}) = q(\sum_{j=t+1}^{n} x_j' \mathbf{e}_j') \leq 0$ lead to a contradiction. □

Definition 2.4. *A non-degenerate quadratic form $q : V \to \mathbb{R}$ is called* ***positive definite*** *(or* ***negative definite****) if*

$$q(x) > 0 \quad (\textit{or } q(x) < 0)$$

for all nonzero vectors $x \neq 0$.

The quadratic form q is called ***positive semidefinite*** *(or* ***negative semidefinite****) if*

$$q(x) \geq 0$$

for all $x \in V$.

Finally, the quadratic form q is called ***indefinite*** *if it takes both positive and negative values.*

Theorem 2.3. *The quadratic form q corresponds to a symmetric bilinear form f and a symmetric matrix F. Equivalently, q is positive semidefine if the eigenvalue $\lambda_i(F)$ satisfy $\forall i \in [n], \lambda_i(F) \geq 0$ and is positive semidefine if the eigenvalue $\lambda_i(F)$ satisfy $\forall i \in [n], \lambda_i(F) > 0$.*

Thanks to my Linear Algebra contextbook.

Proof. First of all, we know the symmetric bilinear form is diagonalizable. According to theorem 2.5, it implies two facts. One is that all eigenvalues are real number and it's valid to compare to zero. The other is that all eigenvectors can span the whole space.

In other words, for all $\mathbf{x} \in V$, there exists a λ s.t. $F\mathbf{x} = \lambda \mathbf{x}$. Then $q(\mathbf{x}) = \mathbf{x}^\top F \mathbf{x} = \lambda \mathbf{x}^\top \mathbf{x}$, which means $q(\mathbf{x}) \geq 0$ if and only if $\lambda \geq 0$. Together with the second fact, $\forall \mathbf{x}. q(\mathbf{x}) \geq 0$ if and only if $\forall i \in [n]. \lambda_i \geq 0$. □

2.1.3 Eigenvector and Eigenvalue

Theorem 2.4. *Eigenvectors corresponding to different eigenvalues must be linearly independent. The sum of eigenvectors, namely $\sum_{\lambda_i \in \operatorname{Spec} \mathcal{A}} V^{\lambda}$ is direct sum. Generally, it's not the original space.*

Proof. Prove it by induction. Suppoe we select m different eigenvalues. If $m = 1$, it holds. Assume our statement holds for each $m < n$.

If it is not true when $m = n$, then select some linearly dependent vectors $\mathbf{e}_1, \ldots, \mathbf{e}_m$. We have,

$$\alpha_1\mathbf{e}_1 + \cdots + \alpha_m\mathbf{e}_m = 0,$$

where α_i is non-trivial.

Apple the operator $\mathcal{A}$ to both sides of the equation. We have

$$\lambda_1\alpha_1\mathbf{e}_1 + \cdots + \lambda_m\alpha_m\mathbf{e}_m = 0.$$

Subtract them:

$$(\lambda_1 - \lambda_m)\alpha_1\mathbf{e}_1 + (\lambda_2 - \lambda_m)\alpha_2\mathbf{e}_2 + \cdots + (\lambda_{m-1} - \lambda_m)\alpha_{m-1}\mathbf{e}_{m-1} = 0.$$

It leads to a contradiction that if we select $m-1$ different eigenvalues, they are linearly dependent. □

Theorem 2.5 (Diagonalizable)**.** *Suppose A is a linear operator on a finite-dimensional vector space V over the field $\mathbb{R}$. A necessary and sufficient condition for $\mathcal{A}$ to be diagonalizable is that the following two conditions are satisfied:*

1. *All roots of the characteristic polynomial $\chi_A(t)$ are in $\mathbb{R}$;*
2. *The geometric multiplicity of each eigenvalue λ equals its algebraic multiplicity.*

Proof. $\Rightarrow$: Suppose two conditions hold. Let $\lambda_1, \ldots,_m$ are the roots of the characteristic polynomial $\chi_A(t)$, while $k_1, \ldots, k_m$ are their multiplicity. We know $\dim V^{\lambda_1} + \cdots + V^{\lambda_m} = \dim V^{\lambda_1} + \cdots + \dim V^{\lambda_m}$, therefore, $V = V^{\lambda_1} \oplus \cdots \oplus V^{\lambda_m}$. All the basis of $V^{\lambda_1}, \ldots, V^{\lambda_m}$ consists the basis of V. It's obvious that $\mathcal{A}$ is diagonal under such basis.

$\Leftarrow$: I omit details because it's more straightforward. □

2.2 SDP Formulation

2.2.1 Spectrahedron

Definition 2.5 (LMI)**.** *A linear matrix inequality (LMI) has the form*

$$A_0 + \sum_{i=1}^{m} A_i x_i \succeq 0,$$

where $A_i \in \mathcal{S}^n$ are given symmetric matrices.

Definition 2.6 (Spectrahedron)**.** *A set $S \subset \mathbb{R}^m$ is a spectrahedron if it has the form*

$$S = \{(x_1, \ldots, x_m) \in \mathbb{R}^m : A_0 + \sum_{i=1}^{m} A_i x_i \succeq 0\},$$

for some given symmetric matrices $A_i \in \mathcal{S}^n$.

2.2.2 Primal SDP Formulation

Definition 2.7. An SDP problem in standard primal form is written as

$$\begin{array}{ll} \text{minimize} & \langle C, X \rangle \\ \text{subject to} & \langle A_i, X \rangle = b_i, \quad i = 1, \ldots, m \\ & X \succeq 0. \end{array}$$

where $C, A_i \in \mathcal{S}^n$, and $\langle X, Y \rangle := \mathrm{Tr}(X^\top Y) = \sum_{ij} X_{ij} Y_{ij}$.

2.2.3 Dual SDP Formulation

Definition 2.8. An SDP problem in dual form is written as

$$\begin{array}{ll} \text{maximize} & b^\top y \\ \text{subject to} & \displaystyle\sum_{i=1}^{m} A_i y_i \preceq C, \end{array}$$

where $b = (b_i, \ldots b_m)$, and $y = (y_1, \ldots, y_m)$ are the dual decision variables.