$\boxed{\text{Topic 4}}$ Duality.

This topic introduces duality theory as a tool to analyze an optimization problem. We first discuss this theory for a general problem (without assuming convexity), then apply to a convex optimization problem.

1. <u>The Lagrange dual function:</u>

+) Consider a general (not necessarily convex) problem

$$\left.\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & f_i(x) \le 0 \;,\; i=1\ldots m \\ & h_i(x) = 0 \;,\; i=1\ldots p \end{array}\right\} \text{called the "primal" problem.}$$

where the domain D is nonempty

$$D = \bigcap_{i=0}^{m} \operatorname{dom} f_i \cap \bigcap_{i=1}^{p} \operatorname{dom} h_i$$

Denote the optimal value as p^*.

+) Lagrangian duality takes into account the constraints by augmenting the objective function with a weighted sum of the constraint functions: Define the Lagrangian as:

Lagrangian: $\mathcal{L}(x,\lambda,\nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$

$\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^p \to \mathbb{R}$, $\operatorname{dom} L = D \times \mathbb{R}^m \times \mathbb{R}^p$.

λ_i : Lagrange multiplier associated with i^{th} inequality constraint

ν_i : " " i^{th} equality constraint

vectors λ and ν ($\lambda \in \mathbb{R}^m$, $\nu \in \mathbb{R}^p$) are called dual variables (or Lagrange multiplier vectors)

+) Lagrange dual function:

This is a function of only the dual variables (λ, ν), obtained by finding the minimum of the Lagrangian over x.

Lagrangian $\mathcal{L}(x, \lambda, \nu)$: function of all primal & dual variables

x : primal variables

λ, ν : dual variables.

Dual function:

$$g(\lambda, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu)$$

$$= \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

Since $g(\lambda, \nu)$ is the infimum of a family of affine functi of (λ, ν), it is always concave even if the original problem is non-convex.

+) <u>Lower bound on the optimal value:</u>

$$\forall \lambda \succeq 0, \forall \nu : \quad g(\lambda, \nu) \leq p^*$$

This property can be easily verified:

Suppose $\tilde{x}$ is feasible $\rightarrow$ $f_i(\tilde{x}) \leq 0$, $i = 1 \ldots m$

$h_i(\tilde{x}) = 0$, $i = 1 \ldots p$

$$\rightarrow \quad \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0 \quad \forall \lambda_i \geq 0.$$

Then

$$\mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}) \quad \forall \tilde{x} \text{ feasible}$$

thus

$$g(\lambda, \nu) = \inf_{x \in D} \mathcal{L}(x, \lambda, \nu) \leq \mathcal{L}(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

$\forall \tilde{x}$ feasible

$$\rightarrow \quad g(\lambda, \nu) \leq p^*.$$

Note that $g(\lambda, \nu)$ could be $-\infty$, in which case the bound is vacuo

Example: (i) Iequality form LP.

$$\min \ c^T x$$
$$\text{s.t. } a_i^T x - b_i \le 0 \quad , i = 1 \dots m$$

$$\mathcal{L}(x,\lambda) = c^T x + \sum_{i=1}^{m} \lambda_i (a_i^T x - b_i)$$
$$= -b^T\lambda + (A^T\lambda + c)^T x$$

Then

$$g(\lambda) = \inf_x \mathcal{L}(x,\lambda) = \begin{cases} -b^T\lambda & \text{if } A^T\lambda + c = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

We obtain a non-trivial lower bound only when λ satisfies $A^T\lambda + c = 0$; then $-b^T\lambda$ is a lower bound on the optimal value of the LP.

(ii) Two way partitioning problem: (nonconvex)

$$\min \ x^T W x$$
$$\text{s.t. } x_i^2 = 1 \quad , i = 1, \dots, n \quad , W \in S^n.$$

This problem is nonconvex, combinatoric in x. The set of feasible points is finite (contains 2^n points).

We can find the optimal point by exhaustive search if $n \le 30$ but larger n ($n > 50$) the search is prohibitive (2^n operations)

Two-way partitioning interpretation: $x_i = -1$ or $x_i = +1$.

W_{ij}: cost of having elements i and j in the same partition

$-W_{ij}$: " different "

$\Rightarrow$ want to minimize total cost

The Lagrangian of this problem is

$$\mathcal{L}(x,\nu) = x^T W x + \sum_{i=1}^{n} \nu_i (x_i^2 - 1)$$
$$= x^T(W + \mathrm{diag}(\nu_i))x - 1^T\nu$$

The dual function is

$$g(\nu) = \begin{cases} -1^T\nu & \text{if } W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

From the dual function we can obtain a lower bound by taking a specific value of the dual variable ν, such as

$$\nu = -\lambda_{min}(W).1,$$

then

$$W + \mathrm{diag}(\nu) = W - \lambda_{min}(W).I \succeq 0.$$

which yields the lower bound

$$p^* \geq -1^T\nu = n\lambda_{min}(W).$$

+) Lagrange dual and the conjugate function:

The Lagrange dual is closely related to the conjugate function

$$f^*(y) := \sup_{x \in \mathrm{dom} f} (y^T x - f(x))$$

Consider

$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } Ax \preceq b \\ &\quad\;\; Cx = d \end{aligned}$$

then

$$\begin{aligned} g(\lambda,\nu) &= \inf_x \left(f_0(x) + \lambda^T(Ax-b) + \nu^T(Cx-d) \right) \\ &= -\lambda^T b - \nu^T d + \inf_x \left(f_0(x) + (\lambda^T A + \nu^T C)x \right) \\ &= -\lambda^T b - \nu^T d - f_0^*(-\lambda^T A - \nu^T C) \end{aligned}$$

This relation is useful if we readily know the conjugate function of $f_0(x)$.

The dual function is

$$g(\nu) = \begin{cases} -1^T\nu & \text{if } W + \mathrm{diag}(\nu) \succcurlyeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

From the dual function we can obtain a lower bound by taking a specific value of the dual variable ν, such as

$$\nu = -\lambda_{min}(W) . 1 ,$$

then

$$W + \mathrm{diag}(\nu) = W - \lambda_{min}(W) . I \succcurlyeq 0 .$$

which yields the lower bound

$$p^* \geq -1^T\nu = n \lambda_{min}(W) .$$

+) <u>Lagrange dual and the conjugate function:</u>
The Lagrange dual is closely related to the conjugate funct

$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^T x - f(x) \right)$$

Consider

$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } Ax \preccurlyeq b \\ &\qquad Cx = d \end{aligned}$$

then

$$\begin{aligned} g(\lambda, \nu) &= \inf_x \left(f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d) \right) \\ &= -\lambda^T b - \nu^T d + \inf_x \left(f_0(x) + (\lambda^T A + \nu^T C) x \right) \\ &= -\lambda^T b - \nu^T d - f_0^*(-\lambda^T A - \nu^T C) \end{aligned}$$

This relation is useful if we readily know the conjugate function of $f_0(x)$.

Example: Minimum volume covering ellipsoid

$$\min \; f_0(X) = \log\det X^{-1}$$
$$\text{s.t. } a_i^T X a_i \leq 1, \; i = 1 \ldots m$$

$\operatorname{dom} f_0 = S_{++}^n$.

The conjugate function of $f_0(X)$ is:

$$f_0^*(Y) = \log\det(-Y)^{-1} - n.$$

The original problem has inequality constraints that are linear in X, which can be expressed as

$$\operatorname{tr}(a_i a_i^T X) \leq 1$$

Then the dual function is:

$$g(\lambda) = \inf_X \left(\log\det(X^{-1}) + \sum_{i=1}^m \lambda_i [\operatorname{tr}(a_i a_i^T X) - 1] \right)$$

$$= -f_0^*\left(-\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) - 1^T\lambda$$

$$= \begin{cases} +\log\det\left(+\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) + n - 1^T\lambda & \text{if } \sum \lambda_i (a_i a_i^T) \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

Thus for any $\lambda \succeq 0$ such that $\sum_{i=1}^m (a_i a_i^T)\lambda_i \succ 0$, we have a lower bound for p^* as

$$\log\det\left(\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) + n - 1^T\lambda.$$

2. The Lagrange dual problem:

We want to know what is the best lower bound that can be obtained from the Lagrange dual function?

Example: Minimum volume covering ellipsoid

$$\min\ f_0(X) = \log\det X^{-1}$$
$$\text{s.t. } a_i^T X a_i \le 1,\ i=1\ldots m$$

$\operatorname{dom} f_0 = S_{++}^n$.

The conjugate function of $f_0(X)$ is:

$$f_0^*(Y) = \log\det(-Y)^{-1} - n.$$

The original problem has inequality constraints that are linear in X, which can be expressed as

$$\operatorname{tr}(a_i a_i^T X) \le 1$$

Then the dual function is:

$$g(\lambda) = \inf_X \left(\log\det(X^{-1}) + \sum_{i=1}^m \lambda_i [\operatorname{tr}(a_i a_i^T X) - 1] \right)$$

$$= -f_0^*\left(-\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) - 1^T\lambda$$

$$= \begin{cases} +\log\det\left(+\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) + n - 1^T\lambda & \text{if } \sum \lambda_i (a_i a_i^T) \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

Thus for any $\lambda \succeq 0$ such that $\sum_{i=1}^m (a_i a_i^T)\lambda_i \succ 0$, we have a lower bound for p^* as

$$\log\det\left(\sum_{i=1}^m \lambda_i (a_i a_i^T)\right) + n - 1^T\lambda.$$

cture 13:

2. The Lagrange dual problem:

We want to know what is the best lower bound that can be obtained from the Lagrange dual function?

This best bound lead to another opt problem:

$$\max \quad g(\lambda, \nu)$$
$$\text{s.t.} \quad \lambda \succeq 0$$

This problem is called the Lagrange dual problem. The original problem is then called the primal problem.

- Dual feasible: a pair (λ, ν) such that $\lambda \succeq 0$ <u>and</u> $g(\lambda, \nu) > -\infty$ is called dual feasible.

- (λ^*, ν^*): dual optimal (or optimal Lagrange multipliers) if they are optimal for the dual problem.

- The dual prolem is always convex regardless of the primal problem being convex or not.

- We often simplify the dual problem by making any implicit constraints that $(\lambda, \nu) \in \operatorname{dom} g$ to be explicit.

<u>Example</u>:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & a_i^T x \leq b \end{array} \iff \begin{array}{ll} \max & -b^T \lambda \\ \text{s.t} & A^T \lambda + c = 0. \\ & \lambda \succeq 0 \end{array}$$

+) Optimal dual value: d^*
We always have
$$d^* \leq p^* \longrightarrow \text{weak duality.}$$
The difference $p^* - d^*$ is called the duality gap.

<u>Example</u>: two-way partitioning

$$\begin{array}{ll} \min & x^T W x \\ \text{s.t.} & x_i^2 = 1, \ i = 1 \dots n \end{array} \iff \begin{array}{ll} \max & -1^T \nu \\ \text{s.t.} & W + \operatorname{diag}(\nu) \succeq 0 \end{array}$$

difficult to solve $(n \leq 30)$

easy to solve, even for large n $(n \cong 1000)$

+) <u>Strong duality</u>:

- If $d^* = p^*$ then we have strong duality.

  With strong duality, then (λ^*, ν^*) serves as certificate of optimality for optimal x^*.

- For convex problems, we usually (but not always) have strong duality.

There are many conditions that guarantee strong duality for convex problems. These conditions are called "constraints qualifications". One of them is Slater's condition.

+) Slater's condition: $\exists x$ "strictly feasible" such that

$$f_i(x) < 0 \ , \ i = 1 \dots m \ , \quad Ax = b .$$

- For affine inequalities, Slater's condition can be refined (relaxed) so that the affine inequalities do <u>not</u> have to hold with strict inequalities.

  Specifically, suppose $f_1, \dots, f_k$ are affine, then strong duality holds if

$$\begin{cases} f_i(x) \le 0 \ , \ i = 1 \dots k \ , \\ f_i(x) < 0 \ , \ i = k+1, \dots m \\ Ax = b \end{cases}$$

Slater's theorem: Strong duality holds if the problem is convex and Slater's condition holds (i.e. there exists a strictly feasible point).

Slater's condition also implies that if $d^* > -\infty$ then it is attained (achievable). That is, there exists a dual feasible point (λ^*, ν^*) such that $g(\lambda^*, \nu^*) = d^* = p^*$.

Example: (i) LP: Strong duality always hold as long as the problem is feasible ($p^* < \infty$).

$$\begin{aligned} &\min c^T x \\ &\text{s.t. } Ax \leq b \end{aligned} \quad \Longleftrightarrow \quad \begin{aligned} &\max -b^T\lambda \\ &\text{s.t. } A^T\lambda + c = 0 \\ &\quad\;\; \lambda \succeq 0 \end{aligned}$$

$p^* = d^*$ except when $p^* = \infty$, $d^* = -\infty$.
(both primal & dual are infeasible).

(ii) Minimum covering ellipsoid: Strong duality always holds.
(as the inequality constr. is linear in X: $\exists X$ s.t. $a_i^T X a_i \leq 1$).

(iii) Non-convex problem with strong duality.

$$\begin{aligned} &\min \; x^T A x + 2b^T x \\ &\text{s.t. } x^T x \leq 1 \end{aligned}$$

$A \in S^n$ but $A \not\succeq 0$ so the primal problem is non-convex

$$\begin{aligned} \mathcal{L}(x,\lambda) &= x^T A x + 2b^T x + \lambda(x^T x - 1) \\ &= x^T(A + \lambda I)x + 2b^T x - \lambda \end{aligned}$$

$$\rightarrow g(\lambda) = \begin{cases} -b^T(A+\lambda I)^{\dagger} b - \lambda & \text{if } \begin{cases} A + \lambda I \succeq 0 \\ b \in \mathcal{R}(A+\lambda I) \end{cases} \\ -\infty & \text{otherwise.} \end{cases}$$

where $(A+\lambda I)^{\dagger}$ is the pseudo-inverse of $A + \lambda I$.

Dual problem:

$$\begin{aligned} &\max \; -b^T(A+\lambda I)^{\dagger} b - \lambda \\ &\text{s.t. } A + \lambda I \succeq 0 \\ &\qquad b \in \mathcal{R}(A+\lambda I) \end{aligned} \quad \xrightarrow[\text{(reformulate)}]{\text{SDP}} \quad \begin{aligned} &\max \; -t - \lambda \\ &\text{s.t. } \begin{bmatrix} A+\lambda I & b \\ b^T & t \end{bmatrix} \succeq 0 \end{aligned}$$

Strong Duality holds (not easy to show).

+) <u>Geometric interpretation and sketch of a proof for strong duality.</u>

Define a set of function values:

$$\mathcal{A} = \{ (u, v, t) \mid \exists x \in D : \begin{array}{l} u_i \geq f_i(x), \; i = 1 \dots m \\ v_i = h_i(x), \; i = 1 \dots p \\ t \geq f_0(x) \end{array} \}$$

where $u \in \mathbb{R}_+^m$
$v \in \mathbb{R}^p$
$t \in \mathbb{R}_+$

The set $\mathcal{A}$ is convex for convex problems.

Then the set of feasible values in $\mathcal{A}$ is $\{(0, 0, t)\}$, and the optimal value can be expressed as

$$p^* = \inf \{ t \mid (0, 0, t) \in \mathcal{A} \}.$$

Also the dual function at (λ, ν) for $\lambda \succeq 0$ can be obtained as

$$g(\lambda, \nu) = \inf_{(u,v,t)} \{ (\lambda, \nu, 1)^T (u, v, t) \mid (u, v, t) \in \mathcal{A} \}.$$

since

$$g(\lambda, \nu) = \inf_x \left(\sum \lambda_i f_i(x) + \sum \nu_i h_i(x) + f_0(x) \right)$$

$$= \inf_{(u,v,t) \in \mathcal{A}} \left(\sum \lambda_i u_i + \sum \nu_i v_i + t \right)$$

$$= \inf_{(u,v,t) \in \mathcal{A}} \underbrace{(\lambda, \nu, 1)^T}_{a^T} \underbrace{(u, v, t)}_{w}$$

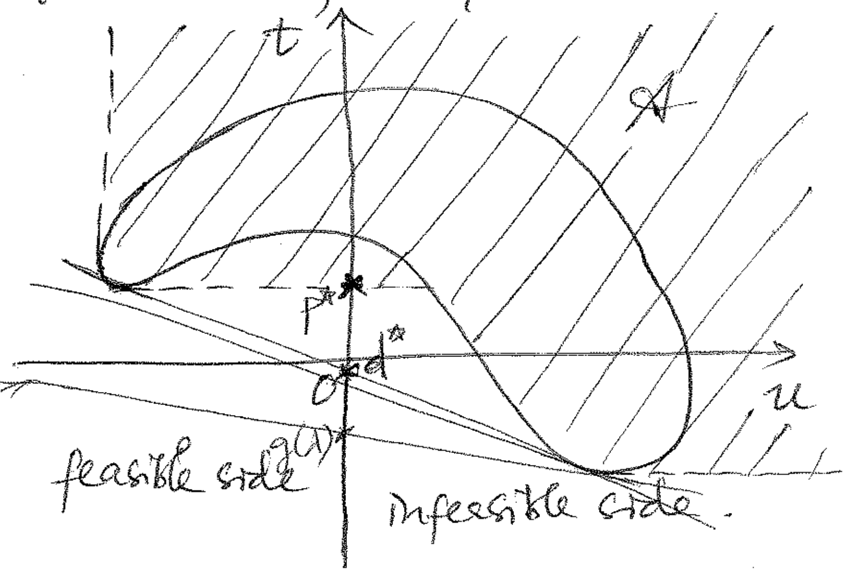
$$=' \inf_{w \in \mathcal{A}}$$

Thus $g(\lambda, \nu) \leq (\lambda, \nu, 1)^T (u, v, t) = a^T w \;\; \forall w \in \mathcal{A}$ is a hyperplane supporting set $\mathcal{A}$, which is in the form

$$a^T w \geq b \;\; \forall w \in \mathcal{A} \text{ where } \begin{cases} a = (\lambda, \nu, 1) \\ b = g(\lambda, \nu). \end{cases}$$

Plot the set $\mathcal{A}$ with the last element (t) as the "vertical axis. It is easier to visualize in a 2D plane, so lets consider a problem with only one inequality constraint, so that $u \in \mathbb{R}$, $v = \phi$.

This plot is for a generic (non-convex) problem, so set $\mathcal{A}$ is not convex.



hyperplane $a^T w = b$ or $\lambda u + t = g(\lambda)$ for a given λ.

<u>Weak duality</u>: We already have for each (λ, ν) then

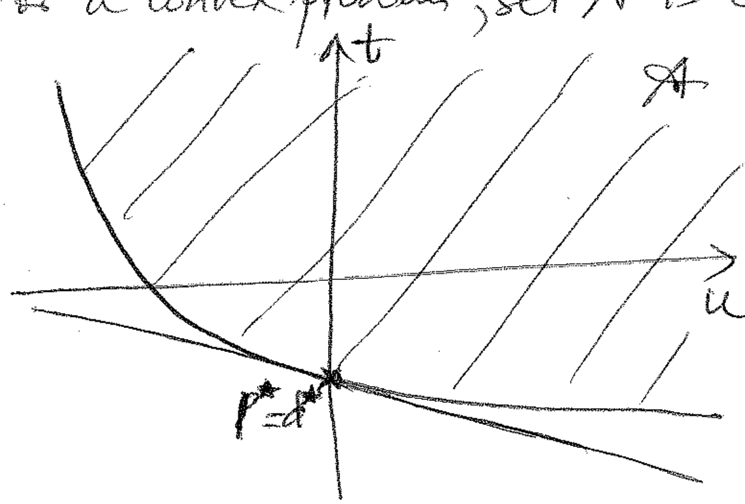
$$g(\lambda,\nu) \le (\lambda,\nu,1)^T(u,v,t) \quad \forall (u,v,t) \in \mathcal{A}$$

but $(0,0,p^*) \in \mathcal{A}$ thus

$$g(\lambda,\nu) \le (\lambda,\nu,1)^T(0,0,p^*) = p^* \quad \forall \lambda,\nu$$

<u>Strong duality</u>: For a convex problem, set $\mathcal{A}$ is convex.

This plot is for a convex set $\mathcal{A}$ resulting from a convex problem.



The idea is then to say that since $\mathcal{A}$ is convex, then the point $(0, 0, p^*)$ on the boundary of $\mathcal{A}$ must have a supporting hyperplane at it (we saw that this was not necessarily true for a non convex set $\mathcal{A}$).

Then Slater's condition is used to ensure that this supporting hyperplane at $(0, 0, p^*)$ is not vertical (the same as the t axis), which ensures that there is an intersection between this supporting hyperplane and the vertical axis t, so that $p^* > -\infty$.

Once $p^* > -\infty$ then the intersection is both p^* and d^*, therefore we get $p^* = d^*$.

+) <u>Max-min and saddle point interpretation:</u>

Consider a problem with only inequality constraints for simplicity (but analysis can be extended to problems with equality constraints).

The Lagrangian is $\mathcal{L}(\lambda, x) = \sum_{i=1}^{m} \lambda_i f_i(x) + f_0(x)$.

Now note that since $\lambda \succeq 0$, we have

$$\sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda) = \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$= \begin{cases} f_0(x) & \text{if } f_i(x) \leq 0,\ i = 1 \dots m \\ \infty & \text{otherwise} \end{cases}$$

Thus for all feasible x, then

$$f_0(x) = \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda) \qquad \forall x \in \mathcal{F} \text{ (feasible set)}$$

hence

$$p^* = \inf_{x \in \mathcal{F}} f_0(x) = \inf_{x} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda).$$

The idea is then to say that since $\mathcal{A}$ is convex, then the point $(0,0,p^*)$ on the boundary of $\mathcal{A}$ must have a supporting hyperplane at it (we saw that this was not necessarily true for a non convex set $\mathcal{A}$).

Then Slater's condition is used to ensure that this supporting hyperplane at $(0,0,p^*)$ is not vertical (the same as the t axis), which ensures that there's an intersection between this supporting hyperplane and the vertical axis t, so that $p^* > -\infty$.

Once $p^* > -\infty$ then the intersection is both p^* and d^*, therefore we get $p^* = d^*$.

+) Max-min and saddle point interpretation:

Consider a problem with only inequality constraints for simplicity (but analysis can be extended to problems with equality constraints).

The Lagrangian is $\mathcal{L}(\lambda, x) = \sum_{i=1}^{m} \lambda_i f_i(x) + f_0(x)$.

Now note that since $\lambda \succeq 0$, we have

$$\sup_{\lambda \succeq 0} \mathcal{L}(x,\lambda) = \sup_{\lambda \succeq 0} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) \right)$$

$$= \begin{cases} f_0(x) & \text{if } f_i(x) \le 0,\ i = 1 \dots m \\ \infty & \text{otherwise} \end{cases}$$

Thus for all feasible x, then

$$f_0(x) = \sup_{\lambda \succeq 0} \mathcal{L}(x,\lambda) \qquad \forall x \in \mathcal{F} \text{ (feasible set)}$$

hence

$$p^* = \inf_{x \in \mathcal{F}} f_0(x) = \inf_{x} \sup_{\lambda \succeq 0} \mathcal{L}(x,\lambda).$$

By the definition of the dual function, we also have

$$d^* = \sup_{\lambda \succeq 0} g(\lambda) = \sup_{\lambda \succeq 0} \inf_{x} \mathcal{L}(x, \lambda).$$

Thus weak duality implies

$$\sup_{\lambda \succeq 0} \inf_{x} \mathcal{L}(x, \lambda) \leq \inf_{x} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda)$$

In fact this inequality holds in general for every function $\mathcal{L}(x, \lambda)$. (max-min inequality)

Strong duality holds when we have equality:

$$\sup_{\lambda \succeq 0} \inf_{x} \mathcal{L}(x, \lambda) = \inf_{x} \sup_{\lambda \succeq 0} \mathcal{L}(x, \lambda).$$

+ Saddle point: For a given function $f(w, z)$, the pair $(\tilde{w}, \tilde{z})$ is called the saddle-point for f if

$$f(\tilde{w}, z) \leq f(\tilde{w}, \tilde{z}) \leq f(w, \tilde{z}) \quad \forall\, w, z.$$

$\tilde{w}$ minimizes f over w : $\quad f(\tilde{w}, \tilde{z}) = \inf_{w} f(w, \tilde{z})$

$\tilde{z}$ maximizes f over z : $\quad f(\tilde{w}, \tilde{z}) = \sup_{z} f(\tilde{w}, z)$

$$\longrightarrow \quad \inf_{w} \sup_{z} f(w, z) = \sup_{z} \inf_{w} f(w, z) = f(\tilde{w}, \tilde{z})$$

Thus strong duality holds and the saddle point is the common value.

+ Strong duality can also be interpreted as a zero-sum game between two players, one wants to minimize the payoff to the other, and the other wants to maximize this payoff.

3. Optimality conditions:

Again for this part we do not assume the original (primal) problem to be convex. So the theory applies to all optimization problems.

+) Certificate of (sub) optimality and stopping criterion:

- If we can always find a dual feasible point (λ, ν), then we can establish a lower bound on p^*:

$$p^* \geq g(\lambda, \nu).$$

$\to$ (λ, ν) provides a proof or certificate that $p^* \geq g(\lambda, \nu)$.

- If strong duality holds $\to$ (λ^*, ν^*) provides a certificate of optimality: $\quad p^* = g(\lambda^*, \nu^*)$

- Duality gap: define $\Delta = f_0(x) - g(\lambda, \nu)$, x, λ, ν feasible then

$$f_0(x) - p^* \leq \Delta$$

This establishes that point x is Δ-suboptimal.

- We can use this property to establish a non-heuristic stopping criterion for algorithms, so that we stop when

$$f(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \leq \varepsilon_{tolerance}$$

where (k) indicates the iteration number (much more on this later).

+) Complementary slackness:

Suppose that strong duality holds. Let x^* and (λ^*, ν^*) be the primal & dual optimal points.

We have:

$$f_0(x^*) = g(\lambda^*, \nu^*) \qquad \text{(strong duality)}$$
$$= \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x) \right)$$
$$\leq f_0(x^*) + \sum_{i=1}^{m} \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(x^*)}_{\leq 0} + \sum_{i=1}^{p} \nu_i^* \underbrace{h_i(x)}_{= 0}$$
$$\leq f_0(x^*)$$

which implies we must have equality through. Thus

$$\sum_{i=1}^{m} \lambda_i^* f_i(x^*) = 0$$

Since each term is non-positive (≤ 0), this implies

$$\lambda_i^* f_i(x_i^*) = 0 \quad \forall\, i = 1 \ldots m.$$

This condition is called complimentary slackness.

Complimentary slackness can also be expressed as:

$$\lambda_i^* > 0 \longrightarrow f_i^*(x) = 0 \quad \text{(tight constraint)}$$
$$\text{or } f_i^*(x) < 0 \longrightarrow \lambda_i^* = 0 \quad \text{(loose constraint)}.$$

+) KKT Conditions: Assume that $f_0, f_1 \ldots f_m, h_1 \ldots h_p$ are all differentiable, but we do not assume they are convex ye[t]

Let x^* and (λ^*, ν^*) be any primal and dual optimal poi[nts] with zero duality gap.

Since x^* minimizes $\mathcal{L}(x, \lambda^*, \nu^*)$ over x, it follows that

$$\nabla_x \mathcal{L}(x, \lambda^*, \nu^*) \Big|_{x = x^*} = 0$$
$$\Leftrightarrow \nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0.$$

Coupled with the fact that x^* is feasible, we then have the following set of conditions:

$$\begin{aligned} f_i(x^*) &\le 0 & i&=1\dots m \\ h_i(x^*) &= 0 & i&=1\dots p \\ \lambda_i^* &\geq 0 & i&=1\dots m \\ \lambda_i^* f_i(x^*) &= 0 & i&=1\dots m \end{aligned}$$

$$\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* h_i(x^*) = 0$$

This set of conditions is called the KKT conditions.

The KKT condition holds for any optimization problem that obtains strong duality and has differentiable objectives and constraint functions.

- KKT for convex problems:

For a primal convex problem, the KKT condition is also sufficient for x^*, (λ^*, ν^*) to be primal & dual optimal.

So for convex problems, KKT is both necessary and sufficient.

That is, if the point $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ that satisfy the KKT condition for a convex problem ($f_i(x)$ convex, $h_i(x)$ affine):

$$\begin{aligned} f_i(\tilde{x}) &\le 0 & i&=1\dots m \\ h_i(\tilde{x}) &= 0 & i&=1\dots p \\ \tilde{\lambda}_i &\geq 0 & i&=1\dots m \\ \tilde{\lambda}_i \cdot f_i(x) &= 0 & i&=1\dots m \end{aligned}$$

$$\nabla f_0(\tilde{x}) + \sum \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x}) = 0$$

then $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ are primal & dual optimal with zero duality gap.

This sufficiency condition is easy to show:

- The first 2 conditions imply $\tilde{x}$ feasible.
- The last condition implies $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$.

$$\rightarrow \; g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$
$$= f_0(\tilde{x}) + \sum \tilde{\lambda}_i f_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x})$$
$$= f_0(\tilde{x})$$

$\rightarrow$ Strong duality holds for $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ (with zero duality gap)

Convexity is used in the second argument: $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$

This is based on the fact that

$$\tilde{\lambda}_i \geq 0 \rightarrow \mathcal{L}(x, \tilde{\lambda}, \tilde{\nu}) \text{ is convex in } x$$

(non-negative weighted sum, and $h_i(x)$ are affine).

then the last KKT condition which states the gradient of $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$ vanishes at $x = \tilde{x}$ implies $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$

This conclusion may not hold if the functions f_i are not convex or h_i are not affine.

Lecture 15:

Example: Water filling in achieving the communication capacity.

$$\min \; -\sum_{i=1}^{n} \log(\alpha_i + x_i)$$
$$\text{s.t.} \quad x \succeq 0, \quad 1^T x = 1,$$

where $\alpha_i > 0$.

x_i : transmit power on channel i

$1/\alpha_i$: noise power on channel i.

$$\mathcal{L}(x, \lambda, \nu) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^T x + \nu(1^T x - 1)$$

The KKT conditions reduce to

$$x^* \succeq 0, \quad 1^T x^* = 1, \quad \lambda_i^* x^* = 0, \quad \lambda^* \succeq 0$$
$$-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0; \quad i = 1 \ldots n$$

This sufficiency condition is easy to show:

- The first 2 conditions imply $\tilde{x}$ feasible.
- The last condition implies $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$.

$$\rightarrow g(\tilde{\lambda}, \tilde{\nu}) = L(\tilde{x}, \tilde{\lambda}, \tilde{\nu})$$

$$= f_0(\tilde{x}) + \sum \tilde{\lambda}_i f_i(\tilde{x}) + \sum \tilde{\nu}_i h_i(\tilde{x})$$

$$= f_0(\tilde{x})$$

$\rightarrow$ Strong duality holds for $\tilde{x}$, $(\tilde{\lambda}, \tilde{\nu})$ (with zero duality gap)

Convexity is used in the second argument: $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$
This is based on the fact that

$$\tilde{\lambda}_i \geq 0 \rightarrow \mathcal{L}(x, \tilde{\lambda}, \tilde{\nu}) \text{ is convex in } x$$

(non-negative weighted sum, and $h_i(x)$ are affine).

then the last KKT condition which states the gradient of $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$ vanishes at $x = \tilde{x}$ implies $\tilde{x}$ minimizes $\mathcal{L}(x, \tilde{\lambda}, \tilde{\nu})$
This conclusion may not hold if the functions f_i are not convex or h_i are not affine.

Lecture 15:

Example: Water filling in achieving the communication capacity.

$$\min \; -\sum_{i=1}^{n} \log(\alpha_i + x_i)$$

$$\text{s.t.} \quad x \succeq 0, \quad 1^T x = 1,$$

where $\alpha_i > 0$.

x_i : transmit power on channel i

$1/\alpha_i$: noise power on channel i.

$$\mathcal{L}(x, \lambda, \nu) = -\sum_{i=1}^{n} \log(\alpha_i + x_i) - \lambda^T x + \nu(1^T x - 1)$$

The KKT conditions reduce to

$$x^* \succeq 0, \quad 1^T x^* = 1, \quad \lambda_i^* x^* = 0, \quad \lambda^* \succeq 0$$

$$-\frac{1}{\alpha_i + x_i^*} - \lambda_i^* + \nu^* = 0; \quad i = 1 \ldots n$$

Since λ_i^* is a slack variable, we can eliminate it, for which we can write

$$x_i^*\left(\frac{-1}{\alpha_i+x_i^*}+\nu^*\right)=0$$

$$x^*\succeq 0\ ,\quad 1^T x^*=1$$

$$\nu^*\geq 1/(\alpha_i+x_i^*)$$

Solving this set of conditions leads to

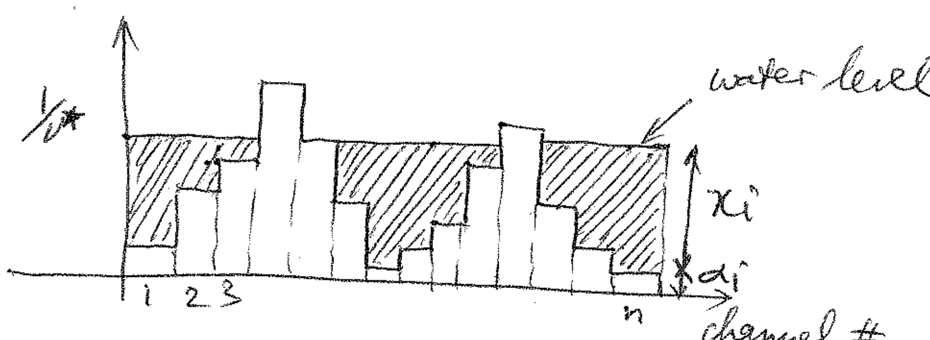
$$x_i^*=\begin{cases}1/\nu^*-\alpha_i & \nu^*<\frac{1}{\alpha_i}\\ 0 & \nu^*\geq\frac{1}{\alpha_i}.\end{cases}$$

or simply

$$x_i^*=\left(\frac{1}{\nu^*}-\alpha_i\right)^+\quad\text{(notation)}.$$

To find ν^*, substitute this expression into $1^T x=1$ to get

$$\sum_{i=1}^{n}\max\left\{0,\frac{1}{\nu^*}-\alpha_i\right\}=1$$



t) Solving the primal problem via the dual:

Sometimes the dual problem is easier to solve (for example, it has some special structures, or admits analytical solutions).

Suppose that we have strong duality and know an optimal point (λ^*,ν^*). Then we can solve the following problem to find the optimal primal values: (Suppose its solution is unique)

$$\min f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

If the solution of the above problem is primal feasible, then it must be the primal optimal point. If that solution is not primal feasible then no primal optimal point exists, i.e., the primal optimum is not attained.

4. Perturbation & sensitivity analysis.

+) The perturbed problem:

$$\begin{aligned} &\min f_0(x) \\ &\text{s.t. } f_i(x) \le u_i, \quad i = 1 \dots m \\ &\qquad h_i(x) = v_i, \quad i = 1 \dots p. \end{aligned}$$

- This problem is the same as the original one if $u_i = 0, v_i = 0$
- If $u_i > 0$, we relax the ith inequality constraint
  $u_i < 0$, we tighten "

- Define $p^*(u, v)$ as the optimal value of the perturbed prob

$$p^*(u,v) = \inf \{ f_0(x) \mid \exists x \in D : \; f_i(x) \le u_i, \; i = 1 \dots m; \; h_i(x) = v_i, \; i = 1 \dots p \}.$$

Note that $p^*(0,0) = p^*$, the optimal value of the primal prob

+) Comparison:

	primal	dual
unperturbed	$\min f_0(x)$ s.t. $f_i(x) \le 0$, $h_i(x) = 0$	$\max g(\lambda, \nu)$ s.t. $\lambda \succeq 0$
perturbed	$\min f_0(x)$ s.t. $f_i(x) \le u_i$, $h_i(x) = v_i$	$\max g(\lambda,\nu) - \lambda^T u - \nu^T v$ s.t. $\lambda \succeq 0$

Assume that strong duality holds, and the dual optimal is attained.

Let (λ^*, ν^*) be the dual optimal point, then

$$\forall\, u, v: \quad p^*(u,v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v.$$

To see this, suppose $\tilde{x}$ is a feasible point of the perturbed problem: By strong duality, we have

$$\begin{aligned}
p^*(0,0) &= g(\lambda^*, \nu^*) \\
&= \inf_x \left(f_0(x) + \sum_1^m \lambda_i^* f_i(x) + \sum_1^p \nu_i^* h_i(x) \right) \\
&\leq f_0(\tilde{x}) + \sum_1^m \underbrace{\lambda_i^*}_{\geq 0} \underbrace{f_i(\tilde{x})}_{\leq u_i} + \sum_1^p \nu_i^* \underbrace{h_i(x)}_{= v_i} \\
&\leq f_0(\tilde{x}) + \lambda^{*T} u + \nu^{*T} v \qquad \forall \tilde{x} \text{ feasible.}
\end{aligned}$$

Hence for all $\tilde{x}$ feasible for the perturbed problem, then

$$f_0(\tilde{x}) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v \qquad \forall \tilde{x} \in \mathcal{F}(u,v)$$

$$\rightarrow \quad p^*(0,0) - \lambda^{*T} u - \nu^{*T} v \leq \inf_{\tilde{x} \in \mathcal{F}(u,v)} f_0(\tilde{x}) = p^*(u,v).$$

+) Sensitivity interpretation: When strong duality holds, we can infer various sensitivity interpretations:

- λ_i^* is the sensitivity factor of the i^{th} inequality constraint.

  If λ_i^* is large, then tightening the i^{th} constraint ($u_i < 0$) will greatly increase the optimal value $p^*(u,v)$.

  If λ_i^* is small and we loosen the i^{th} constraint ($u_i > 0$) then the optimal value will not decrease much

If $\lambda_i^* = 0$, the problem is insensitive to the i^{th} inequality constraints (see the local analysis below).

- Similarly for ν_i^*:
  If $\nu_i^* > 0$ and is large $\to p^*$ very sensitive to v_i ($v_i < 0$)
  $\nu_i^* < 0$ " $\to$ " ($v_i > 0$).

+) Local sensitivity analysis: In addition, if $p^*(u, v)$ is differentiable at $(0,0)$ then:

$$\lambda_i^* = \left.\frac{\partial p^*(u,v)}{\partial u_i}\right|_{u=0, v=0}$$

$$\nu_i^* = \left.\frac{\partial p^*(u,v)}{\partial v_i}\right|_{u=0, v=0}$$

This is because:

$$\frac{\partial p^*(0,0)}{\partial u_i} = \lim_{t\to 0} \frac{p^*(t.e_i, 0) - p^*(0,0)}{t}, \quad e_i = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix} \leftarrow i^{th}$$

e_i is a vector with the i^{th} element as 1 and the rest 0.

Since we have $p^*(u, v) \geq p^*(0,0) - \lambda^{*T} u - \nu^{*T} v$

$$\to p^*(t.e_i, 0) - p^*(0,0) \geq -\lambda_i^* t$$

Thus

$$\frac{p^*(te_i, 0) - p^*(0,0)}{t} \begin{cases} \geq -\lambda_i^* & \text{if } t > 0 \\ \leq -\lambda_i^* & \text{if } t < 0. \end{cases}$$

Taking the limit as $t \to 0$ yields

$$\frac{\partial p^*(0,0)}{\partial u_i} = -\lambda_i^*.$$

+ Thus λ_i^* gives the local sensitivity of p^* to the ith inequality constraint.

- If $\lambda_i^* = 0$, $f_i(x^*) < 0$ → the constraint is inactive and can be tighten or loosen by a small amount without affecting the optimal value p^*.

- If $\lambda_i^* > 0$, $f_i(x^*) = 0$ → the constraint is active, then λ_i^* tells how sensitive is the optimal value to this constraint if it is loosen or tighten a little bit. The interpretation is symmetric:

  Tightening by $u_i < 0$, $|u_i|$ small → increase in p^* by $\approx -\lambda_i^* u_i$

  Loosening by $u_i > 0$, $|u_i|$ small → decrease p^* by $\approx +\lambda_i^* u_i$

+) <u>Shadow pricing interpretation:</u>

Let $f_0(x)$ be the cost ($-f_0(x)$ the profit)
$f_i(x)$ be constraints on resources

→ λ_i^* tells approximately how much more profit the firm can make for a small increase in availability of resource i.

Then λ_i^* is the equilibrium price for resource i. The firm can buy more resource i if the market price is lower than λ_i^*, and can sell resource i if the price is higher than λ_i^*. Based on that, the firm can make profit.

+) <u>Theorems of alternatives:</u>

Here we apply Lagrangian duality theory to determine the feasibility of a system of inequalities and equalities.

$$\begin{cases} f_i(x) \le 0, & i = 1 \ldots m \\ h_i(x) = 0, & i = 1 \ldots p \end{cases}$$

- This problem can be cast as

$$\begin{aligned} \min \quad & 0 \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1 \ldots m \\ & h_i(x) = 0, \quad i = 1 \ldots p \end{aligned}$$

Assume the domain D is non_empty.

If $p^* = 0 \rightarrow$ set of inequalities & equalities are feasible
$p = \infty \rightarrow$ " infeasible

- Dual function:

$$g(\lambda, \nu) = \inf_{x \in D} \left(\sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right)$$

and the dual problem is

$$\begin{aligned} \max \quad & g(\lambda, \nu) \\ \text{s.t.} \quad & \lambda \succeq 0. \end{aligned}$$

- Note that for $\alpha > 0 \rightarrow g(\alpha\lambda, \alpha\nu) = \alpha g(\lambda, \nu)$. Thus the optimal dual value is:

$$d^* = \begin{cases} \infty & \text{if } \lambda \succeq 0, \ g(\lambda, \nu) > 0 \text{ is feasible} \\ 0 & \text{if } \lambda \succeq 0, \ g(\lambda, \nu) > 0 \text{ is infeasible} \end{cases}$$

- Since $d^* \le p^*$, we can conclude that
if $\lambda \succeq 0$, $g(\lambda, \nu) > 0$ is feasible ($d^* = \infty$)
then $f_i(x) \le 0$, $h_i(x) = 0$ is infeasible ($p^* = \infty$).

Alternatively, if $f_i(x) \le 0$, $h_i(x) = 0$ is feasible ($p^* = 0$)
then $\lambda \succeq 0$, $g(\lambda, \nu) = 0$ is infeasible ($d^* = 0$).

Thus the two sets of inequalities and equalities:

$$f_i(x) \leq 0, \quad h_i(x) = 0$$

and $\lambda \succeq 0$, $g(\lambda, \nu) > 0$

are called weak alternatives. That is, at most one of the two set is feasible.

Weak alternatives hold for all problems (Convex not required)

+) Weak alternatives with strict inequalities:

$$f_i(x) < 0, \quad h_i(x) = 0$$

and $\lambda \succeq 0$, $\lambda \neq 0$, $g(\lambda, \nu) \geq 0$

are weak alternatives.

This is easy to show: Suppose $\exists \tilde{x}$: $f_i(\tilde{x}) < 0$, $h_i(\tilde{x}) = 0$, then $\forall \lambda \succeq 0$, $\lambda \neq 0$:

$$\sum_i \lambda_i f_i(\tilde{x}) + \sum \nu_i h_i(\tilde{x}) < 0$$

thus

$$g(\lambda, \nu) = \inf_x \left(\sum \lambda_i f_i(x) + \sum \nu_i h_i(x) \right) \leq \sum \lambda_i f_i(\tilde{x}) + \sum \nu_i h_i(\tilde{x}) < 0.$$

+) Strong alternatives: Strong alternatives mean exactly one system is feasible.

Strong alternatives hold for convex feasibility problems (f_i are convex and h_i are affine), provided some qualification holds.

Two systems: $f_i(x) < 0$, $i = 1 \dots m$, $Ax = b$

and $\lambda \succeq 0$, $\lambda \neq 0$, $g(\lambda, \nu) \geq 0$

are strong alternatives if $\exists x$ in the interior of D such that $Ax = b$.

The proof is based on strong duality & Slater's condition.

Similarly, two systems

$f_i(x) \leq 0$, $i = 1 \dots m$, $Ax = b$

and $\lambda \succeq 0$, $g(\lambda, \nu) > 0$

are strong alternatives if $\exists x$ in the interior of D such that $Ax = b$ and the value p^* is attained.

Example: Two systems

$Ax \preceq 0$, $c^T x < 0$, $A \in \mathbb{R}^{m \times n}$, $c \in \mathbb{R}$

and

$A^T y + c = 0$, $y \succeq 0$

are strong alternatives. (Farka's lemma)

This can be seen by analyzing the dual of the LP:

$$\begin{array}{ll} \min c^T x & \\ \text{s.t. } Ax \preceq 0 & \end{array} \quad \xrightarrow{\text{dual}} \quad \begin{array}{l} \max 0 \\ \text{s.t. } A^T y + c = 0 \\ \quad\;\; y \succeq 0 \end{array}$$

$$\mathcal{L}(x, \lambda) = c^T x + \lambda^T (Ax) = (A^T \lambda + c)^T x$$

$$\rightarrow g(\lambda) = \inf_x \mathcal{L}(x, \lambda) = \begin{cases} 0 & \text{if } A^T \lambda + c = 0 \\ -\infty & \text{o/w.} \end{cases}$$

Since $x = 0$ is feasible for the primal, strong duality holds and $p^* = d^*$. Thus strong alternatives hold.

+) Semidefinite program:

$$\min \ c^T x$$
$$\text{s.t.} \ \sum_{i=1}^{n} x_i F_i + G \preceq 0 \quad , F_i, G \in S^k.$$

The Lagrangian variable of this SDP is a matrix $Z \in S_+^k$

$$\mathcal{L}(x, Z) = c^T x + tr\Big(\Big(\sum_{i=1}^{n} x_i F_i + G\Big) Z\Big)$$
$$= x_1\big(c_1 + tr(F_1 Z)\big) + x_2\big(c_2 + tr(F_2 Z)\big) + \dots$$
$$+ x_n\big(c_n + tr(F_n Z)\big) + tr(GZ).$$

$\mathcal{L}(x, Z)$ is affine in x

$$\rightarrow g(Z) = \begin{cases} tr(GZ) & \text{if } c_i + tr(F_i Z) = 0 \ \forall i = 1 \dots n \\ -\infty & \text{otherwise} \end{cases}$$

The dual problem can then be expressed as

$$\min \ tr(GZ)$$
$$\text{s.t.} \ tr(F_i Z) + c_i = 0 \ , \ i = 1 \dots n$$
$$Z \succeq 0.$$

Strong duality holds when the primal problem is strictly feasible, that is, $\exists x$ such that

$$x_1 F_1 + x_2 F_2 + \dots + x_n F_n + G \prec 0.$$