CS 2601 Linear and Convex Optimization

15. Lagrange duality in general problems

Bo Jiang

John Hopcroft Center for Computer Science
Shanghai Jiao Tong University

Fall 2022

Outline

- Dual function and dual problem
- Weak and strong duality
- Slater's condition
- KKT conditions revisited

Lagrange dual function

Consider the general optimization problem (not necessarily convex),

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & h_i(\boldsymbol{x}) = 0,\ i = 1, 2, \ldots, k \\ & g_j(\boldsymbol{x}) \le 0,\ j = 1, 2, \ldots, m \end{aligned} \tag{P}$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x})$$

The (Lagrange) dual function is

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in D} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in D} \left(f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right)$$

where $D = \operatorname{dom} f \cap (\bigcap_{i=1}^{k} \operatorname{dom} h_i) \cap (\bigcap_{j=1}^{m} \operatorname{dom} g_j)$ is the domain of the problem. We will downplay the role of D and focus on the case $D = \mathbb{R}^n$.

Example

Given $\boldsymbol{A} \in \mathbb{R}^{k \times n}$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \mathbf{0} \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \boldsymbol{x}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{\mu}^T \boldsymbol{x}$$

Since $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is convex in $\boldsymbol{x}$, its minimum satisifies

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = 2\boldsymbol{x} + \boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{\mu} = \mathbf{0} \implies \boldsymbol{x} = \frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda})$$

The dual function is

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}\left(\frac{1}{2}(\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}), \boldsymbol{\lambda}, \boldsymbol{\mu}\right) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda}$$

Example

Given $\boldsymbol{A} \in \mathbb{R}^{k \times n}$,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = \boldsymbol{x}^T \boldsymbol{x} + \boldsymbol{\lambda}^T (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b})$$

Since $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})$ is convex in $\boldsymbol{x}$, its minimum satisifies

$$\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = 2\boldsymbol{x} + \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{0} \implies \boldsymbol{x} = -\frac{1}{2} \boldsymbol{A}^T \boldsymbol{\lambda}$$

The dual function is

$$\phi(\boldsymbol{\lambda}) = \mathcal{L}\left(-\frac{1}{2}\boldsymbol{A}^T \boldsymbol{\lambda}, \boldsymbol{\lambda}\right) = -\frac{1}{4} \boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{A}^T \boldsymbol{\lambda} - \boldsymbol{b}^T \boldsymbol{\lambda} = -\frac{1}{4} \|\boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda}$$

Lower bound for optimal value

For any $\boldsymbol{\lambda}$ and any $\boldsymbol{\mu} \geq \mathbf{0}$, the optimal value f^* of (P) is bounded by

$$f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Proof. Let $X = \{\boldsymbol{x} : h_i(\boldsymbol{x}) = 0, \forall i; g_j(\boldsymbol{x}) \leq 0, \forall j\}$ be the feasible set.

- If $X = \emptyset$, then $f^* = +\infty$, trivially true.
- If $X \neq \emptyset$, for $\boldsymbol{\mu} \geq \mathbf{0}$ and $\boldsymbol{x} \in X$,

$$f(\boldsymbol{x}) \geq f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i \underbrace{h_i(\boldsymbol{x})}_{=0} + \sum_{j=1}^{m} \underbrace{\mu_j g_j(\boldsymbol{x})}_{\leq 0} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$$

Minimizing over $\boldsymbol{x}$,

$$f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}} f(\boldsymbol{x}) \geq \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Concavity of dual function

The dual function is always concave, whether the primal problem (P) is convex or not.

Proof. Note $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is affine in $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. Thus $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is the pointwise infimum of a family of affine functions indexed by $\boldsymbol{x}$, and hence concave. (Recall the pointwise supremum of convex functions is convex).

$$\begin{aligned}
\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x} \in D} \left(f(\boldsymbol{x}) + \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) + \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right) \\
&= - \underbrace{\sup_{\boldsymbol{x} \in D} \left(-f(\boldsymbol{x}) - \sum_{i=1}^{k} \lambda_i h_i(\boldsymbol{x}) - \sum_{j=1}^{m} \mu_j g_j(\boldsymbol{x}) \right)}_{\text{pointwise supremum of convex (affine) functions in } (\boldsymbol{\lambda}, \boldsymbol{\mu})}
\end{aligned}$$

Example. $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{4}\|\boldsymbol{\mu} - \boldsymbol{A}^T\boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T\boldsymbol{\lambda}$ is concave.

Lagrange dual problem

To find the best lower bound given by the dual function

$$f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

solve the (Lagrange) dual problem associated with the primal problem (P),

$$\begin{array}{ll} \max\limits_{\boldsymbol{\lambda},\boldsymbol{\mu}} & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{s.t.} & \boldsymbol{\mu} \geq \mathbf{0} \end{array} \tag{D}$$

The dual problem (D) is always convex, whether or not (P) is convex.

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \mathbf{0}$ and $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$.

Note. The domain of a convex function f is $\mathrm{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) < +\infty\}$, while the domain of a concave function f is $\mathrm{dom} f = \{\boldsymbol{x} : f(\boldsymbol{x}) > -\infty\}$. Thus the condition $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) > -\infty$ just means $(\boldsymbol{\lambda}, \boldsymbol{\mu}) \in \mathrm{dom}\, \phi$.

Example

Recall the dual problem of the following LP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{c}^T\boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \geq \boldsymbol{h} \end{aligned}$$

is

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda},\boldsymbol{\mu}) = \begin{cases} \boldsymbol{\lambda}^T\boldsymbol{b} + \boldsymbol{\mu}^T\boldsymbol{h}, & \text{if } \boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ -\infty, & \text{otherwise} \end{cases} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

$(\boldsymbol{\lambda},\boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \boldsymbol{0}$ and $\boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c}$, which just means it is feasible for the dual LP,

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \psi(\boldsymbol{\lambda},\boldsymbol{\mu}) = \boldsymbol{\lambda}^T\boldsymbol{b} + \boldsymbol{\mu}^T\boldsymbol{h} \\ \text{s.t.} \quad & \boldsymbol{A}^T\boldsymbol{\lambda} + \boldsymbol{G}^T\boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

Example

Recall the dual problem of the following problem

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) = \boldsymbol{x}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{x} \geq \boldsymbol{0} \end{aligned}$$

is

$$\begin{aligned} \max_{\boldsymbol{\lambda}, \boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) = -\frac{1}{4} \|\boldsymbol{\mu} - \boldsymbol{A}^T \boldsymbol{\lambda}\|^2 - \boldsymbol{b}^T \boldsymbol{\lambda} \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

$(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual feasible if $\boldsymbol{\mu} \geq \boldsymbol{0}$, as there is no implicit constraint in ϕ.

Outline

- Dual function and dual problem
- Weak and strong duality
- Slater's condition
- KKT conditions revisited

Weak and strong duality

Denote by f^* and ϕ^* the primal and dual optimal values, i.e.

$$f^* = \inf_{\boldsymbol{x} \in X} f(\boldsymbol{x}), \qquad \phi^* = \sup_{\boldsymbol{\lambda}, \boldsymbol{\mu} : \boldsymbol{\mu} \geq \boldsymbol{0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

Weak duality: $f^* \geq \phi^*$

- always holds.

  Proof. Recall $f^* \geq \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ for any $\boldsymbol{\lambda}$ and any $\boldsymbol{\mu} \geq \boldsymbol{0}$. Weak duality follows by maximizing over $\boldsymbol{\lambda}$ and $\boldsymbol{\mu} \geq \boldsymbol{0}$.
- $f^* - \phi^*$ is called the (optimal) duality gap of the problem.

Strong duality: $f^* = \phi^*$

- does not hold in general.
- typically holds for convex problems under various conditions known as constraint qualifications, e.g. Slater's condition.
- may also hold for nonconvex problems.
- can solve the dual problem instead if it is easier than the primal.

Duality gap

Given primal feasible $\boldsymbol{x}$ and dual feasible $(\boldsymbol{\lambda}, \boldsymbol{\mu})$, the difference

$$f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$$

is called the duality gap associated with $\boldsymbol{x}$ and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

Note

$$\phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \phi^* \leq f^* \leq f(\boldsymbol{x})$$

If the duality gap is zero, i.e. $f(\boldsymbol{x}) = \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$, then all inequalities become equalities, so $\boldsymbol{x}$ is primal optimal, and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is dual optimal.

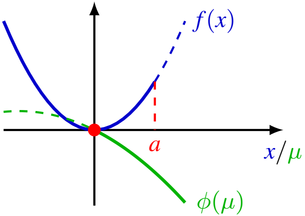
If the gap $f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \epsilon$, then the dual solution $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ serves as a proof or certificate that $\boldsymbol{x}$ is ϵ-suboptimal,

$$f(\boldsymbol{x}) - f^* \leq f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \epsilon$$

When strong duality holds, this can serve as a stopping criterion in an iterative algorithm, i.e. stop when $f(\boldsymbol{x}) - \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \epsilon$ for some $(\boldsymbol{\lambda}, \boldsymbol{\mu})$.

Example

$$\min_{x\in\mathbb{R}} \quad f(x) = x^2$$
$$\text{s.t.} \quad x \le a$$



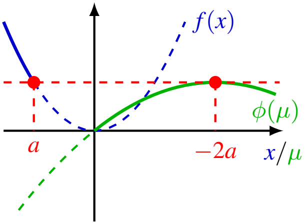
Case 1. $a \ge 0$

The dual function is

$$\phi(\mu) = \inf_x [x^2 + \mu(x-a)] = -\frac{\mu^2}{4} - a\mu$$

The dual problem is

$$\max_{\mu\in\mathbb{R}} \quad \phi(\mu) = -\frac{\mu^2}{4} - a\mu$$
$$\text{s.t.} \quad \mu \ge 0$$



Case 2. $a \le 0$

The primal and dual optimal values are

1. If $a \ge 0, f^* = f(0) = \phi^* = \phi(0) = 0$
2. If $a \le 0, f^* = f(a) = \phi^* = \phi(-2a) = a^2$

Strong duality holds in both cases.

Example

Consider

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & f(x) = x^3 \\ \text{s.t.} \quad & x \geq 0 \end{aligned}$$

The optimal value is $f^* = f(0) = 0$.

The dual function is

$$\phi(\mu) = \inf_x [x^3 - \mu x] = -\infty$$

so the dual optimal value is

$$\phi^* = \sup_{\mu \geq 0} \phi(\mu) = -\infty$$

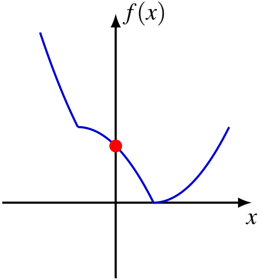
The duality gap is infinite. In particular, strong duality does not hold.

Example

Consider

$$\min_{x\in\mathbb{R}} \quad f(x) = \begin{cases} -x^2 - x + \frac{3}{4}, & |x| \le \frac{1}{2} \\ x^2 - x + \frac{1}{4}, & |x| \ge \frac{1}{2} \end{cases}$$

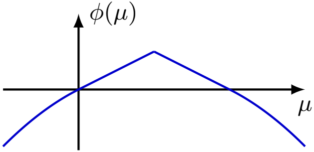
$$\text{s.t.} \quad x \le 0$$



The primal optimal value is $f^* = f(0) = \frac{3}{4}$.

The dual function is

$$\phi(\mu) = \inf_x [f(x)+\mu x] = \begin{cases} \frac{1-|\mu-1|}{2}, & |\mu - 1| \le 1 \\ \frac{1-(\mu-1)^2}{4}, & |\mu - 1| \ge 1 \end{cases}$$



The dual optimal value is $\phi^* = \phi(1) = \frac{1}{2}$.

The duality gap is $f^* - \phi^* = \frac{1}{4}$.

Example (cont'd)

To compute the dual function, note

$$\mathcal{L}(x,\mu) = f(x) + \mu x = \begin{cases} -x^2 + (\mu-1)x + \frac{3}{4}, & |x| \le \frac{1}{2} \\ x^2 + (\mu-1)x + \frac{1}{4}, & |x| > \frac{1}{2} \end{cases}$$

Since $y = -x^2 + (\mu-1)x + \frac{3}{4}$ is a parabola opening down,

$$\phi_1(\mu) = \inf_{|x|\le\frac{1}{2}} \mathcal{L}(x,\mu) = \min\left\{\mathcal{L}(\frac{1}{2},\mu), \mathcal{L}(-\frac{1}{2},\mu)\right\} = \frac{1-|\mu-1|}{2}$$

Since $y = x^2 + (\mu-1)x + \frac{1}{4}$ is a parabola opening up,

$$\phi_2(\mu) = \inf_{|x|\ge\frac{1}{2}} \mathcal{L}(x,\mu) = \begin{cases} \mathcal{L}(\frac{1-\mu}{2},\mu) = \frac{1-(\mu-1)^2}{4}, & |\mu-1| \ge 1 \\ \min\left\{\mathcal{L}(\frac{1}{2},\mu), \mathcal{L}(-\frac{1}{2},\mu)\right\} = \frac{1-|\mu-1|}{2}, & |\mu-1| \le 1 \end{cases}$$

Thus

$$\phi(\mu) = \min\{\phi_1(\mu), \phi_2(\mu)\} = \phi_2(\mu)$$

Example (cont'd)

By definition of dual function,

$$\phi(\mu) = \inf_x [f(x) + \mu x] \leq f(x) + \mu x$$

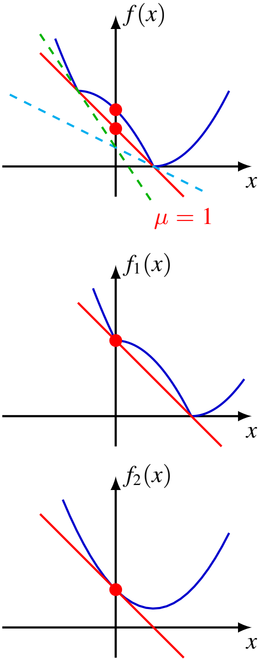
Rearranging,

$$\ell(x) \triangleq -\mu x + \phi(\mu) \leq f(x)$$

Note $\ell(x)$ is a line with slope $-\mu$ and intercept $\phi(\mu)$ that supports $\text{epi} f$.

The dual optimal value ϕ^* is the largest intercept of such lines. We can see pictorially there is a gap.

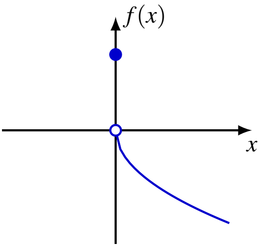
This also give us intuition about why strong duality may hold for nonconvex problem, and why it usually holds for convex problems.



Example

$$\min_{x \in \mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

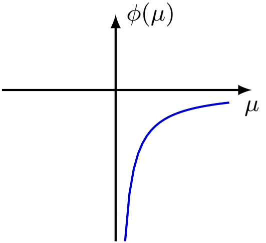
s.t. $\quad x \leq 0$



The primal optimal value is $f^* = f(0) = 1$.

The dual function is

$$\phi(\mu) = \inf_x [f(x) + \mu x] = \begin{cases} -\frac{1}{4\mu}, & \mu > 0 \\ -\infty, & \mu \leq 0 \end{cases}$$



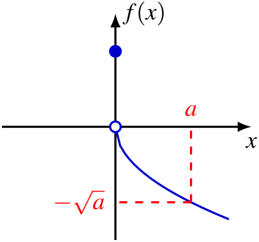
The dual optimal value is $\phi^* = 0$, which is not attainable.

This is a convex problem with nonzero duality gap $f^* - \phi^* = 1$, a nontypical case.

Example

$$\min_{x\in\mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

$$\text{s.t.} \quad x \le a$$

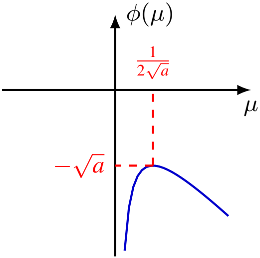


where $a > 0$.

The primal optimal value is $f^* = f(a) = -\sqrt{a}$.

The dual function is

$$\phi(\mu) = \inf_x [f(x) + \mu(x - a)] = \begin{cases} -\frac{1}{4\mu} - a\mu, & \mu > 0 \\ -\infty, & \mu \le 0 \end{cases}$$



The dual optimal value is $\phi^* = \phi(\frac{1}{2\sqrt{a}}) = -\sqrt{a}$

Strong duality holds in this case.

Outline

- Dual function and dual problem
- Weak and strong duality
- Slater's condition
- KKT conditions revisited

Slater's condition for convex problems

Consider a convex problem,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & g_j(\boldsymbol{x}) \leq 0,\ j = 1, 2, \ldots, m \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \end{aligned} \tag{CP}$$

with domain $D = \operatorname{dom} f \cap \left(\bigcap_{j=1}^{m} \operatorname{dom} g_j\right)$.

Slater's condition. The above problem is strictly feasible, i.e.

$$\exists \boldsymbol{x} \in \operatorname{int} D^{1} \quad \text{s. t.} \quad g_j(\boldsymbol{x}) < 0 \text{ for } j = 1, 2, \ldots, m, \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

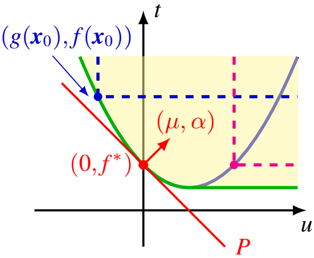
Refined Slater's condition. If some g_j are affine, the requirement $g_j(\boldsymbol{x}) < 0$ can be relaxed to feasibility $g_j(\boldsymbol{x}) \leq 0$ for those g_j.

---

[1] $\operatorname{int} D$ stands for the interior of D. $\boldsymbol{x} \in \operatorname{int} D$ if there exists $\delta > 0$ s.t. $B(\boldsymbol{x}, \delta) \subset D$. Again we focus on the case $D = \mathbb{R}^n$, so the requirement $\boldsymbol{x} \in \operatorname{int} D$ is always satisfied.

Slater's Theorem

Slater's Theorem. Strong duality holds for (CP) under (refined) Slater's condition. Furthermore, if $\phi^* > -\infty$, it is attained by some $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$.



Proof idea. Consider the case with only one inequality constraint g. Let

$$C = \{(u,t) : \exists \boldsymbol{x} \in D \text{ s.t. } g(\boldsymbol{x}) \le u, f(\boldsymbol{x}) \le t\}$$

C is convex and has a supporting hyperplane P at $(0, f^*) \in \partial C$,

$$\mu u + \alpha t \ge \alpha f^*, \quad \forall (u,t) \in C$$

Letting $u, t \to +\infty$ shows $\mu \ge 0$, $\alpha \ge 0$. By Slater's condition, $\exists \boldsymbol{x}_0 \in D$ s.t. $g(\boldsymbol{x}_0) < 0$, so P is non-vertical, i.e. $\alpha \ne 0$. Since $(g(\boldsymbol{x}), f(\boldsymbol{x})) \in C$,

$$f^* \le \mu^* g(\boldsymbol{x}) + f(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \mu^*), \text{ where } \mu^* = \frac{\mu}{\alpha} \ge 0$$

Minimizing over $\boldsymbol{x} \in D$, $f^* \le \phi(\mu^*) \le \phi^*$. Weak duality then implies $f^* = \phi(\mu^*) = \phi^*$. The condition $\boldsymbol{x}_0 \in \operatorname{int} D$ will be used to deal with $\boldsymbol{h}$.

Proof

If $f^* = -\infty$, then $\phi^* = f^* = -\infty$ by weak duality. Assume $f^* > -\infty$. Since (CP) is strictly feasible by Slater's condition, $f^* < +\infty$. Also assume $\boldsymbol{A} \in \mathbb{R}^{k \times n}$ has $\operatorname{rank} \boldsymbol{A} = k$, by removing redundant constraints. Now let

$$C = \{(\boldsymbol{u}, \boldsymbol{v}, t) : \exists \boldsymbol{x} \in D \text{ s.t. } \boldsymbol{g}(\boldsymbol{x}) \leq \boldsymbol{u}, \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{v}, f(\boldsymbol{x}) \leq t\}$$

1. C is convex. The proof is similar to that on slide 7 of §4 part 2.
2. $C \neq \emptyset$ and $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$.
   - 2.1 Note $(\mathbf{0}, \mathbf{0}, t) \in C$ iff $\exists \boldsymbol{x} \in D$ s.t. $\boldsymbol{g}(\boldsymbol{x}) \leq \mathbf{0}$, $\boldsymbol{h}(\boldsymbol{x}) = \mathbf{0}$, $f(\boldsymbol{x}) \leq t$, i.e. iff $\exists \boldsymbol{x} \in X$ s.t. $f(\boldsymbol{x}) \leq t$. In particular, $(\mathbf{0}, \mathbf{0}, f(\boldsymbol{x})) \in C$ for $\boldsymbol{x} \in X$.
   - 2.2 Since $f^* = \inf_{x \in X} f(\boldsymbol{x})$, there exists a sequence $\{\boldsymbol{x}_i\} \subset X$ s.t. $f(\boldsymbol{x}_i) \to f^*$. Since $C \ni (\mathbf{0}, \mathbf{0}, f(\boldsymbol{x}_i)) \to (\mathbf{0}, \mathbf{0}, f^*)$, we have $(\mathbf{0}, \mathbf{0}, f^*) \in \overline{C}$.
   - 2.3 $(\mathbf{0}, \mathbf{0}, t) \notin C$ for any $t < f^*$. Thus $(\mathbf{0}, \mathbf{0}, f^*) \notin \operatorname{int} C$ and $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$.
3. There exists a supporting hyperplane at $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$, i.e. there exists $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) \neq \mathbf{0}$ s.t. for all $\forall (\boldsymbol{u}, \boldsymbol{v}, t) \in C$,

$$(\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) \cdot (\boldsymbol{u}, \boldsymbol{v}, t) = \boldsymbol{\mu}^T \boldsymbol{u} + \boldsymbol{\lambda}^T \boldsymbol{v} + \alpha t \geq (\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) \cdot (\mathbf{0}, \mathbf{0}, f^*) = \alpha f^*$$

Proof (cont'd)

4. Since $\boldsymbol{u}, t$ can be arbitrarily large for $(\boldsymbol{u}, \boldsymbol{v}, t) \in C$, letting $\boldsymbol{u}, t \to \infty$ yields $\boldsymbol{\mu} \geq \mathbf{0}$, $\alpha \geq 0$.
5. Since $(\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x}), f(\boldsymbol{x})) \in C$,

$$\boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) + \alpha f(\boldsymbol{x}) \geq \alpha f^*, \quad \forall \boldsymbol{x} \in D$$

6. If $\alpha \neq 0$, then

$$f^* \leq f(\boldsymbol{x}) + (\boldsymbol{\lambda}^*)^T \boldsymbol{h}(\boldsymbol{x}) + (\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

   where $\boldsymbol{\mu}^* = \boldsymbol{\mu}/\alpha \geq \mathbf{0}$ and $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}/\alpha$. Minimizing over $\boldsymbol{x}$,

$$f^* \leq \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \phi^*$$

   Weak duality then implies

$$f^* = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \phi^*$$

   i.e. strong duality holds, and the dual optimal is attained.

Proof (cont'd)

7. Now we show Slater's condition implies $\alpha \neq 0$. Suppose $\alpha = 0$.

   7.1 By 5,
   $$\boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}) + \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in D$$

   7.2 Let $\boldsymbol{x}_0$ satisfy Slater's condition, i.e. $\boldsymbol{x}_0 \in \operatorname{int} D$, $\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}$, and $\boldsymbol{g}(\boldsymbol{x}_0) < \boldsymbol{0}$. Since $\boldsymbol{\mu} \geq \boldsymbol{0}$,
   $$\underbrace{\boldsymbol{\mu}^T \boldsymbol{g}(\boldsymbol{x}_0)}_{\leq 0} + \boldsymbol{\lambda}^T \underbrace{\boldsymbol{h}(\boldsymbol{x}_0)}_{=\boldsymbol{0}} \geq 0 \implies \boldsymbol{\mu} = \boldsymbol{0}$$

   7.3 By 7.1 and 7.2,
   $$\boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}) \geq 0, \quad \forall \boldsymbol{x} \in D$$

   7.4 Since $\boldsymbol{x}_0 \in \operatorname{int} D$, there exists $\delta > 0$ s.t. $\boldsymbol{x}_0 + \boldsymbol{z} \in D$ for all $\boldsymbol{z} \in B(\boldsymbol{0}, \delta)$. Recalling $\boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}$ and $\boldsymbol{h}(\boldsymbol{x}_0) = \boldsymbol{0}$, we have
   $$\boldsymbol{\lambda}^T \boldsymbol{A}\boldsymbol{z} = \boldsymbol{\lambda}^T \boldsymbol{h}(\boldsymbol{x}_0 + \boldsymbol{z}) \geq 0, \quad \forall \boldsymbol{z} \in B(\boldsymbol{0}, \delta) \implies \boldsymbol{A}^T \boldsymbol{\lambda} = \boldsymbol{0}$$

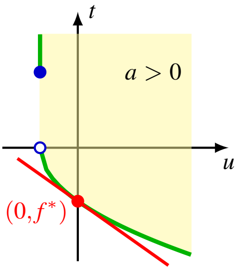
   7.5 Since $\boldsymbol{A}$ has full row rank, $\boldsymbol{\lambda} = \boldsymbol{0}$. Thus $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) = \boldsymbol{0}$, contradicting $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \alpha) \neq \boldsymbol{0}$ given by 3.

Example

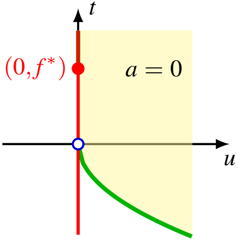
$$\min_{x\in\mathbb{R}} \quad f(x) = \begin{cases} -\sqrt{x}, & x > 0 \\ 1 & x = 0 \\ +\infty, & x < 0 \end{cases}$$

$$\text{s.t.} \quad x \le a$$

is a convex problem with domain $D = [0, \infty)$.
Note $\operatorname{int} D = (0, \infty)$.



- If $a > 0$, Slater's condition is satisfied, e.g. $\frac{a}{2} \in \operatorname{int} D$ and $\frac{a}{2} < a$, so strong duality must hold.



- If $a = 0$, no point in $\operatorname{int} D$ is feasible. Slater's Theorem is not applicable[2] , and it turns out that strong duality does not hold.

---

[2] Slater's condition is only a sufficient condition for strong duality. It is not necessary.

Example: Strong duality for LP

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & \boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} \geq \boldsymbol{h} \end{aligned}$$

dual LP

$$\begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \boldsymbol{b}^T \boldsymbol{\lambda} + \boldsymbol{h}^T \boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{\lambda} + \boldsymbol{G}^T \boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

equivalence

equivalence

$$\begin{aligned} \max_{\boldsymbol{x},\boldsymbol{y}} \quad & -\boldsymbol{c}^T \boldsymbol{x} \\ \text{s.t.} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \\ & \boldsymbol{G}\boldsymbol{x} + \boldsymbol{y} = \boldsymbol{h} \\ & \boldsymbol{y} \geq \boldsymbol{0} \end{aligned}$$

dual LP

$$\begin{aligned} \min_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & -\boldsymbol{b}^T \boldsymbol{\lambda} - \boldsymbol{h}^T \boldsymbol{\mu} \\ \text{s.t.} \quad & \boldsymbol{A}^T \boldsymbol{\lambda} + \boldsymbol{G}^T \boldsymbol{\mu} = \boldsymbol{c} \\ & \boldsymbol{\mu} \geq \boldsymbol{0} \end{aligned}$$

- Essentially, dual of dual is primal.
- By refined Slater's condition, strong duality holds if either the primal or the dual is feasible.
- When either f^* or ϕ^* is finite, then $f^* = \phi^*$ and they are both attained.

Example: Strong duality for LP (cont'd)

There are four possibilities

1. Primal feasible, dual feasible, $-\infty < \phi^* = f^* < +\infty$
2. Primal feasible, dual infeasible, $f^* = \phi^* = -\infty$

$$\begin{array}{ll} \min & x_1 - 2x_2 \\ \text{s.t.} & x_1 - x_2 = -1 \\ & x_1, x_2 \geq 0 \end{array} \qquad \begin{array}{ll} \max & \lambda \\ \text{s.t.} & \lambda + \mu_1 = 1 \\ & -\lambda + \mu_2 = -2 \\ & \mu_1, \mu_2 \geq 0 \end{array}$$

3. Primal infeasible, dual feasible, $f^* = \phi^* = +\infty$
4. Primal infeasible, dual infeasible, $f^* = +\infty, \phi^* = -\infty$

$$\begin{array}{ll} \min & x_1 - 2x_2 \\ \text{s.t.} & x_1 - x_2 \leq 1 \\ & -x_1 + x_2 \leq -2 \end{array} \qquad \begin{array}{ll} \max & -\mu_1 + 2\mu_2 \\ \text{s.t.} & -\mu_1 + \mu_2 = 1 \\ & \mu_1 - \mu_2 = -2 \\ & \mu_1, \mu_2 \geq 0 \end{array}$$

Note. No duality gap in Case 2 and Case 3, but $f^* - \phi^*$ is undefined.

Example: Dual formulation of SVM

Recall the primal formulation of SVM,

$$
\begin{aligned}
\min_{\boldsymbol{w}, b, \boldsymbol{\xi}} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\mathbf{1}^T\boldsymbol{\xi} \\
\text{s. t.} \quad & y_i(\boldsymbol{x}_i^T\boldsymbol{w} + b) \geq 1 - \xi_i, \quad i = 1, 2, \ldots, n \\
& \boldsymbol{\xi} \geq \mathbf{0}
\end{aligned}
$$

where $C > 0$ is a hyperparameter, and $\mathbf{1}$ is the vector of all 1's.

- convex problem with affine constraints.
- always feasible. Indeed, given any $\boldsymbol{w}, b$,

$$\xi_i = [1 - y_i(\boldsymbol{w}^T\boldsymbol{x}_i + b)]^+, \quad i = 1, 2, \ldots, n$$

  yields a feasible solution $(\boldsymbol{w}, b, \boldsymbol{\xi})$, where $(x)^+ = \max\{x, 0\}$.
- strong duality holds by refined Slater's condition
- can solve the dual problem instead, which turns out to be useful!

Example: Dual formulation of SVM (cont'd)

The Lagrangian is

$$\mathcal{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\mu}, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\mathbf{1}^T\boldsymbol{\xi} + \sum_{i=1}^{n} \mu_i[1 - \xi_i - y_i(\boldsymbol{x}_i^T\boldsymbol{w} + b)] - \boldsymbol{\alpha}^T\boldsymbol{\xi}$$

$$= \frac{1}{2}\|\boldsymbol{w}\|_2^2 - \left(\sum_{i=1}^{n} y_i\mu_i\boldsymbol{x}_i\right)^T \boldsymbol{w} - \boldsymbol{\mu}^T\boldsymbol{y}b + (C\mathbf{1} - \boldsymbol{\mu} - \boldsymbol{\alpha})^T\boldsymbol{\xi} + \mathbf{1}^T\boldsymbol{\mu}$$

Minimizing over $\boldsymbol{w}, b, \boldsymbol{\xi}$ yields the dual function ($\boldsymbol{w} = \sum_{i=1}^{n} y_i\mu_i\boldsymbol{x}_i$),

$$\phi(\boldsymbol{\mu}, \boldsymbol{\alpha}) = \begin{cases} \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j, & \text{if } \boldsymbol{\mu}^T\boldsymbol{y} = 0, C\mathbf{1} - \boldsymbol{\mu} - \boldsymbol{\alpha} = \mathbf{0} \\ -\infty, & \text{otherwise} \end{cases}$$

The dual problem is

$$\begin{aligned} \max_{\boldsymbol{\mu}, \boldsymbol{\alpha}} \quad & \phi(\boldsymbol{\mu}, \boldsymbol{\alpha}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\alpha} \geq \mathbf{0} \end{aligned}$$

Example: Dual formulation of SVM (cont'd)

Making the constraints explicit, we obtain the equivalent problem,

$$
\begin{aligned}
\max_{\boldsymbol{\mu},\boldsymbol{\alpha}} \quad & \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \\
\text{s.t.} \quad & \boldsymbol{\mu}^T\boldsymbol{y} = 0 \\
& \boldsymbol{\mu} + \boldsymbol{\alpha} = C\mathbf{1} \\
& \boldsymbol{\mu} \geq \mathbf{0},\ \boldsymbol{\alpha} \geq \mathbf{0}
\end{aligned}
$$

Eliminating $\boldsymbol{\alpha}$, we obtain the following dual formulation of SVM,

$$
\begin{aligned}
\max_{\boldsymbol{\mu}} \quad & \mathbf{1}^T\boldsymbol{\mu} - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\mu_i\mu_j y_i y_j \boldsymbol{x}_i^T\boldsymbol{x}_j \\
\text{s.t.} \quad & \boldsymbol{\mu}^T\boldsymbol{y} = 0 \\
& \mathbf{0} \leq \boldsymbol{\mu} \leq C\mathbf{1}
\end{aligned}
$$

Can be solved efficiently by an algorithm called Sequential Minimal Optimization (SMO). Also amenable to further generalization using the kernel trick that replaces $\boldsymbol{x}_i^T\boldsymbol{x}_j$ by a kernel (function) $K(\boldsymbol{x}_i, \boldsymbol{x}_j)$.

Outline

- Dual function and dual problem
- Weak and strong duality
- Slater's condition
- KKT conditions revisited

KKT conditions for convex problems

Consider a differentiable convex problem and its dual,

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & f(\boldsymbol{x}) \\ \text{s.t.} \quad & \boldsymbol{g}(\boldsymbol{x}) \le \boldsymbol{0} \\ & \boldsymbol{h}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} = \boldsymbol{0} \end{aligned} \qquad \begin{aligned} \max_{\boldsymbol{\lambda},\boldsymbol{\mu}} \quad & \phi(\boldsymbol{\lambda},\boldsymbol{\mu}) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x},\boldsymbol{\lambda},\boldsymbol{\mu}) \\ \text{s.t.} \quad & \boldsymbol{\mu} \ge \boldsymbol{0} \end{aligned}$$

Theorem. KKT conditions hold at $\boldsymbol{x}^*$ with Lagrange multipliers $\boldsymbol{\lambda}^*$, $\boldsymbol{\mu}^*$,

1. (primal feasibility) $\boldsymbol{h}(\boldsymbol{x}^*) = \boldsymbol{0}$, $\boldsymbol{g}(\boldsymbol{x}^*) \le \boldsymbol{0}$
2. (dual feasibility) $\boldsymbol{\mu}^* \ge \boldsymbol{0}$
3. (stationarity) $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \boldsymbol{0}$
4. (complementary slackness) $\mu_j^* g_j(\boldsymbol{x}^*) = 0, j = 1, 2, \ldots, m$

if and only if all the following conditions hold,

1. strong duality holds, i.e. $f^* = \phi^*$
2. $\boldsymbol{x}^*$ is a primal optimal solution, i.e. $f^* = f(\boldsymbol{x}^*)$
3. $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is a dual optimal solution, i.e. $\phi^* = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$

Proof of necessity

Assume KKT holds at $\boldsymbol{x}^*$ with Lagrange multipliers $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$.

- Since $\boldsymbol{\mu}^* \geq \mathbf{0}$, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}) + (\boldsymbol{\lambda}^*)^T \boldsymbol{h}(\boldsymbol{x}) + (\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x})$ is convex in $\boldsymbol{x}$.
- The stationarity condition $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$ implies $\boldsymbol{x}^*$ is a global minimum of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, i.e.

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

- By primal feasibility and complementary slackness,

$$\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*) + (\boldsymbol{\lambda}^*)^T \underbrace{\boldsymbol{h}(\boldsymbol{x}^*)}_{=\mathbf{0}} + \underbrace{(\boldsymbol{\mu}^*)^T \boldsymbol{g}(\boldsymbol{x}^*)}_{=\mathbf{0}} = f(\boldsymbol{x}^*)$$

  so

$$f(\boldsymbol{x}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$$

- By the discussion on slide 12, $\boldsymbol{x}^*$ is primal optimal, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual optimal and strong duality holds.

Proof of sufficiency

Assume strong duality holds, $\boldsymbol{x}^*$ is primal optimal, and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is dual optimal. We only need to show the stationarity condition and the complementary slackness condition.

$$
\begin{aligned}
f^* &= \phi^* \qquad \text{(strong duality)} \\
&= \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \qquad \text{(dual optimality of } (\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)\text{)} \\
&= \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \qquad \text{(definition of dual function)} \\
&\leq \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \qquad \text{(definition of infimum)} \\
&= f(\boldsymbol{x}^*) + (\boldsymbol{\lambda}^*)^T \underbrace{\boldsymbol{h}(\boldsymbol{x}^*)}_{=\mathbf{0}} + (\underbrace{\boldsymbol{\mu}^*}_{\geq\mathbf{0}})^T \underbrace{\boldsymbol{g}(\boldsymbol{x}^*)}_{\leq\mathbf{0}} \\
&\leq f(\boldsymbol{x}^*) \qquad \text{(primal and dual feasibility of } \boldsymbol{x}^*, \boldsymbol{\mu}^*\text{)} \\
&= f^* \qquad \text{(primal optimality of } \boldsymbol{x}^*\text{)}
\end{aligned}
$$

So both inequality holds with equality. The first implies $\boldsymbol{x}^*$ is a minimum of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, so $\nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$. The second implies $(\boldsymbol{\mu}^*)^T\boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0}$, so $\mu_j g_j(\boldsymbol{x}^*) = 0$ for $j = 1, 2, \ldots, m$.