Lecture 19. Lagrange Duality

19.1 Lagrange dual function and Lagrange dual problem

Recall that, if we define the Lagrangian as follows

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

where $\boldsymbol{g}(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \dots, g_m(\boldsymbol{x}))^\mathsf{T}$ and $\boldsymbol{h}(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \dots, h_k(\boldsymbol{x}))^\mathsf{T}$, then the domain of $\mathcal{L}$ is given by

$$\boldsymbol{x} \in D \triangleq \operatorname{dom} f \cap \operatorname{dom} g_1 \cap \cdots \cap \operatorname{dom} g_m \cap \operatorname{dom} h_1 \cap \cdots \cap \operatorname{dom} h_k, \quad \boldsymbol{\lambda} \in \mathbb{R}^m, \quad \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0},$$

and the KKT condition can be expressed as

$$\nabla_{\boldsymbol{x},\boldsymbol{\lambda}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}, \quad \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \leq \mathbf{0}, \quad (\boldsymbol{\mu}^*)^\mathsf{T} \nabla_{\boldsymbol{\mu}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = 0$$

for some KKT multipliers $\boldsymbol{\lambda}^* \in \mathbb{R}^m$ and $\boldsymbol{\mu}^* \in \mathbb{R}^k_{\geq 0}$.

In the part of equality constrained convex optimization, we also mentioned that $\nabla \mathcal{L} = \mathbf{0}$ does not imply that $\mathcal{L}$ achieves its minimum value, since $\mathcal{L}$ is not convex in general. If Lagrange multipliers $\boldsymbol{\lambda}^*$ exists, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*)$ is a *saddle point* of $\mathcal{L}$ in a sense. For inequality constrained optimization problems, the situation becomes more complicated. Since $\nabla_{\boldsymbol{\mu}} \mathcal{L}$ may not be $\mathbf{0}$ in the KKT condition, $(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is possibly not a *stationary point*. But it is still a *minimax* point.

Note that if $\boldsymbol{x}$ is feasible, then $g_i(\boldsymbol{x}) = 0$ for all i and $h_j(\boldsymbol{x}) \leq 0$ for all j. So we have $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f(\boldsymbol{x})$ for all $\boldsymbol{\lambda} \in \mathbb{R}^m$ and $\boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}$. Moreover, we have $\mathcal{L}(\boldsymbol{x}, \mathbf{0}, \mathbf{0}) = f(\boldsymbol{x})$. Thus

$$f(\boldsymbol{x}) = \max_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

For infeasible $\boldsymbol{x}$, there exists $g_i(\boldsymbol{x}) \neq 0$ or $h_j(\boldsymbol{x}) > 0$. Then we known $\max_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \to \infty$ as $\lambda_i g_i(\boldsymbol{x})$ or $\mu_j h_j(\boldsymbol{x})$ goes to infinity. Overall, if $\boldsymbol{x}^*$ is a minimum point of $f(\boldsymbol{x})$ over the domain $D \subseteq \mathbb{R}^n$, we conclude that

$$f(\boldsymbol{x}^*) = \min_{\boldsymbol{x} \in D} \max_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

This equality holds even when we consider general (not necessarily convex) optimization problems.

Here, the order in which we maximize and minimize in this equation is essential and generally cannot be swapped. What happens if we swap the order in which we maximize and minimize? In this lecture, we study the optimization problem after swapping the order.

Definition (*Lagrange dual*)

Given an optimization problem and its Lagragian function $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$, the *Lagrange dual function* is defined by

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

and the *Lagrange dual problem* is given by

$$\begin{array}{rl} \sup & \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ \text{subject to} & \boldsymbol{\mu} \geq \mathbf{0}. \end{array}$$

Sometimes we write min, max instead of inf, sup for convenience.

We first introduce some examples.

Example 1

Primal optimization problem:

$$\begin{array}{rl} \min & x_1^2 + x_2^2 \\ \text{subject to} & x_1 + x_2 \leq -1 \end{array}$$

Lagrangian function:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}(x_1, x_2, \mu) = x_1^2 + x_2^2 + \mu(x_1 + x_2 + 1)$$

Dual function:

$$\phi(\mu) = \min_{x_1, x_2} x_1^2 + x_2^2 + \mu(x_1 + x_2 + 1) = -\frac{1}{2}\mu^2 + \mu$$

Dual problem:

$$\begin{aligned} \max \quad & \phi(\mu) = -\frac{1}{2}\mu^2 + \mu \\ \text{subject to} \quad & \mu \geq 0 \end{aligned}$$

The optimal solution is $\mu^* = 1$ with optimal value $\phi^* = \frac{1}{2}$.

Example 2

Primal optimization problem:

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{subject to} \quad & (x_1 - 1)^2 + x_2^2 \leq 1 \\ & (x_1 + 1)^2 + x_2^2 \leq 1 \end{aligned}$$

Lagrangian function:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \mathcal{L}(x_1, x_2, \mu_1, \mu_2) = x_1 + x_2 + \mu_1\big((x_1 - 1)^2 + x_2^2 - 1\big) + \mu_2\big((x_1 + 1)$$

Dual function:

$$\phi(\mu_1, \mu_2) = \min_{x_1, x_2} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\infty & \text{if } \mu_1 + \mu_2 \leq \\ \dfrac{-2(\mu_1 - \mu_2)^2 + 2(\mu_1 - \mu_2) - 1}{2(\mu_1 + \mu_2)} & \text{otherw} \end{cases}$$

Dual problem:

$$\begin{aligned} \sup \quad & \phi(\mu_1, \mu_2) \\ \text{subject to} \quad & \mu_1, \mu_2 \geq 0 \end{aligned}$$

Let $\mu_1 = \mu_2 \to \infty$, then the optimal value is $\phi^* = \sup \phi(\mu_1, \mu_2) = 0$.

Example 3

Primal optimization problem:

$$\begin{aligned} \max \quad & \boldsymbol{c}^\top \boldsymbol{x} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{b}_1 \\ & \boldsymbol{A}\boldsymbol{x}_2 \leq \boldsymbol{b}_2 \end{aligned}$$

Dual function:

$$\begin{aligned} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) &= \inf_{\boldsymbol{x}} (\boldsymbol{\lambda}^\top \boldsymbol{A}_1 + \boldsymbol{\mu}^\top \boldsymbol{A}_2 - \boldsymbol{c}^\top)\boldsymbol{x} - \boldsymbol{\lambda}^\top \boldsymbol{b}_1 - \boldsymbol{\mu}^\top \boldsymbol{b}_2 \\ &= \begin{cases} -\infty & \text{if } \boldsymbol{\lambda}^\top \boldsymbol{A}_1 + \boldsymbol{\mu}^\top \boldsymbol{A}_2 \neq \boldsymbol{c}^\top \\ -\boldsymbol{\lambda}^\top \boldsymbol{b}_1 - \boldsymbol{\mu}^\top \boldsymbol{b}_2 & \text{otherwise} \end{cases} \end{aligned}$$

Dual problem: $\sup_{\boldsymbol{\mu} \geq \mathbf{0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$, which is equivalent to

$$\begin{array}{ll} \min & \boldsymbol{\lambda}^\mathsf{T} \boldsymbol{b}_1 + \boldsymbol{\mu}^\mathsf{T} \boldsymbol{b}_2 \\ \text{subject to} & \boldsymbol{\lambda}^\mathsf{T} \boldsymbol{A}_1 + \boldsymbol{\mu}^\mathsf{T} \boldsymbol{A}_2 = \boldsymbol{c}^\mathsf{T} \\ & \boldsymbol{\mu} \geq \mathbf{0} \end{array}$$

The Lagrange dual problem of a linear program is exactly the same as the dual problem introduced before.

Example 4

Primal optimization problem:

$$\begin{array}{ll} \min\limits_{x_1 \in \mathbb{R}, x_2 > 0} & e^{-x_1} \\ \text{subject to} & \dfrac{x_1^2}{x_2} \leq 0 \end{array}$$

Domain: $D = \{(x_1, x_2)^\mathsf{T} \mid x_2 > 0\}$
Dual function:

$$\phi(\mu) = \inf_{x_2 > 0} e^{-x_1} + \mu \frac{x_1^2}{x_2} = \begin{cases} 0 & \text{if } \mu \geq 0, \text{ when } x_1 \to \infty \text{ and } \dfrac{x_2}{x_1^2} \to \infty \\ -\infty & \text{otherwise, when } x_1 \to \infty \text{ and } x_2 \to 0 \end{cases}$$

Dual problem:

$$\begin{array}{ll} \sup & \phi(\mu) \\ \text{subject to} & \mu \geq 0 \end{array}$$

An interesting fact is that dual functions are always *concave*, thus dual problems ars always convex, regardless of whether the primal problem is convex or not.

Theorem

For any (not necessarily convex) optimization problem, its Lagrange dual function is concave.

Proof

For any fixed $\boldsymbol{x}$, $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is an affine function of $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$, so

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

is a pointwise minimum of a family of affine functions, which implies that $-\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a convex function.

19.2 Weak and strong duality

What is the relationship between the optimal value of the primal and the optimal value of the dual?
Let D be the domain set, f^* be the optimal value of the primal problem, and ϕ^* be the optimal value of the dual problem. Then we have the following *weak duality theorem*.

Theorem (*Weak duality*)

$$f^* \geq \phi^* .$$

Proof

As we showed before,

$$f^* = \inf_{\boldsymbol{x} \in D} \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) .$$

For any fixed $\boldsymbol{\lambda}_0 \in \mathbb{R}^m, \boldsymbol{\mu}_0 \in \mathbb{R}^k_{\geq 0}$, we have

$$\phi(\boldsymbol{\lambda}_0, \boldsymbol{\mu}_0) = \inf_{\boldsymbol{x} \in D} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}_0, \boldsymbol{\mu}_0) \leq \inf_{\boldsymbol{x} \in D} \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f^* .$$

Thus,

$$\phi^* = \sup_{\boldsymbol{\lambda} \in \mathbb{R}^m,\, \boldsymbol{\mu} \in \mathbb{R}^k_{\geq 0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq f^* .$$

In fact, the weak duality theorem is a particular case of the following *min-max inequality*.

Theorem (*Min-max inequality*)

Let X, Y be two sets. For $F : X \times Y \to \mathbb{R}$, it holds that

$$\inf_{x \in X} \sup_{y \in Y} F(x, y) \geq \sup_{y \in Y} \inf_{x \in X} F(x, y) \,.$$

The min-max inequality has various examples, such as the limit superior and the limit inferior of a sequence. Given a sequence $\{a_n\}$, we have $\limsup a_n \geq \liminf a_n$, where

$$\limsup_{n \to \infty} a_n = \lim_{n \to \infty} \sup_{m \geq n} a_m = \inf_{n \to \infty} \sup_{m \geq n} a_m$$

and

$$\liminf_{n \to \infty} a_n = \lim_{n \to \infty} \inf_{m \geq n} a_m = \sup_{n \to \infty} \inf_{m \geq n} a_m \,.$$

Now we define the *duality gap* as $f^* - \phi^*$, which is always nonnegative. If $f^* = \phi^*$, then we say *strong duality* holds.
Unlike the duality for linear programs, strong duality does not always hold. See e.g., Example 4 above, where $f^* = 1$ and $\phi^* = 0$. A natural question is, under which condition the strong duality holds?

Our first result is that for convex optimization problems, KKT condition implies strong duality.

Theorem

For any convex optimization problem, if a feasible solution $\boldsymbol{x}^*$ has KKT multipliers $\boldsymbol{\lambda}^*$ and $\boldsymbol{\mu}^*$, then strong duality holds. In particular, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is an optimal solution to the dual problem.

Proof

For convex problems, if $\boldsymbol{x}^*$ has KKT multipliers, then $f^* = f(\boldsymbol{x}^*)$. By weak duality, we have

$$f(\boldsymbol{x}^*) = f^* \geq \phi^* = \sup_{\boldsymbol{\mu} \geq \mathbf{0}} \phi(\boldsymbol{\lambda}, \boldsymbol{\mu}) \geq \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \,.$$

So it suffices to show that $\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*)$.
Note that $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*)$ since $g_i(\boldsymbol{x}^*) = 0$ and $\mu_j^* h_j(\boldsymbol{x}^*) = 0$. We fix $\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*$ and let $\widehat{\mathcal{L}}(\boldsymbol{x}) = \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$. Then $\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \min_{\boldsymbol{x}} \widehat{\mathcal{L}}(\boldsymbol{x})$. Because

$\widehat{\mathcal{L}}(\boldsymbol{x})$ is a convex function of $\boldsymbol{x}$ and $\nabla\widehat{\mathcal{L}}(\boldsymbol{x}^*) = \nabla_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$, we conclude that $\boldsymbol{x}^* = \arg\min\ \widehat{\mathcal{L}}(\boldsymbol{x})$. Thus,

$$\phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \min_{\boldsymbol{x}}\ \widehat{\mathcal{L}}(\boldsymbol{x}) = \widehat{\mathcal{L}}(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = f(\boldsymbol{x}^*)\,.$$

Remark

From this proof, we find that for $\boldsymbol{x}^* \in \Omega, \boldsymbol{\lambda}^* \in \mathbb{R}^m, \boldsymbol{\mu}^* \in \mathbb{R}^k$, $f(\boldsymbol{x}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ if and only if $\boldsymbol{x}^*$ is an optimal solution to the primal problem, $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is an optimal solution to the dual problem, and strong duality holds, because

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

This fact holds for any optimization problem (not necessarily convex).

Conversely, if the primal and the dual have finite optimal solutions, and strong duality holds, then KKT condition is satisfied.

Theorem

For any optimization problem, if the primal has a (finite) optimal solution $\boldsymbol{x}^*$, the dual has a (finite) optimal solution $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, and the strong duality holds (or equivalently there exists $\boldsymbol{x}^* \in \Omega, \boldsymbol{\lambda}^* \in \mathbb{R}^m, \boldsymbol{\mu}^* \in \mathbb{R}^k_{\geq 0}$ such that $f(\boldsymbol{x}^*) = \phi(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$), then $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ are KKT multipliers of $\boldsymbol{x}^*$.

Proof

Note that $\boldsymbol{x}^*$ is feasible, so $g_i(\boldsymbol{x}^*) = 0$ for all i, and $h_j(\boldsymbol{x}^*) \leq 0$ for all j. Thus,

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

The first inequality is due to $\boldsymbol{\mu}^* \geq \mathbf{0}$. By strong duality,

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

It yields that

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

We conclude that KKT conditions are satisfied, since

1. $\mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \inf_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \implies \nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$;
2. $f(\boldsymbol{x}^*) = \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \implies \mu_j^* h_j(\boldsymbol{x}^*) = 0$ for all j;
3. $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is feasible for the dual $\implies \boldsymbol{\mu}^* \geq \mathbf{0}$.

Here, the existence of *finite* optimal solutions is essential. Only strong duality is not sufficient. Otherwise, see e.g., Example 2 above, where $\boldsymbol{x}^*$ does not have KKT multipliers.

Tip (*KKT conditions revisit*)

In a sense, KKT conditions are equivalent to existence of optimal solutions and strong duality. So we can rewrite KKT conditions in terms of the primal (P) and dual (D) problems:

1. **Stationary**: $\nabla_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}$;
2. **Primal feasible:** $\boldsymbol{x}^*$ is feasible for (P), namely, $g_i(\boldsymbol{x}^*) = 0$ for all i and $h_j(\boldsymbol{x}^*) \leq 0$ for all j;
3. **Dual feasible**: $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is feasible for (D), namely, $\mu_j^* \geq 0$ for all j;
4. **Complementary slackness**: $\boldsymbol{h}(\boldsymbol{x}^*)^\mathsf{T} \boldsymbol{\mu}^* = 0$, namely, $\mu_j^* h_j(\boldsymbol{x}^*) = 0$ for all j.

Sometimes we may find solving KKT conditions is difficult. Is there any simple condition to certify strong duality?

Geometric interpretation of duality

Before introducing other conditions, we first give a geometric interpretation for strong duality.

Example

Consider the following optimization

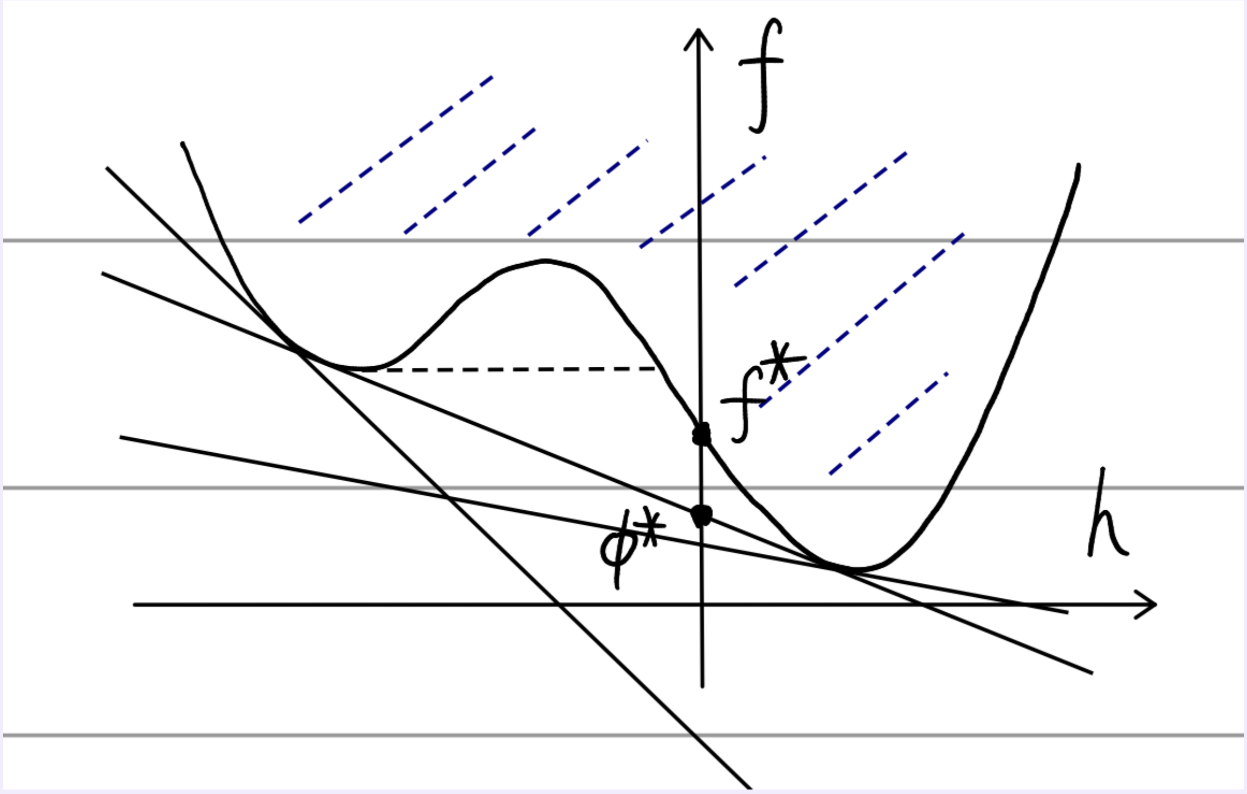
$$\begin{aligned} \min \quad & x^4 - 50x^2 + 100x \\ \text{subject to} \quad & x \geq -\frac{5}{2}\,. \end{aligned}$$

To analyze this problem, we can draw an “epigraph” of

$f(x) = x^4 - 50x^2 + 100x$ with respect to $h(x) = -\frac{5}{2} - x$. Namely, let

$$C \triangleq \{(p,t) \mid \exists\, x, h(x) \le p, f(x) \le t\}\,.$$

Then f^* is the point where C intersects f-axis.



What is ϕ^*? The Lagrangian is

$$\mathcal{L}(x,\mu) = f(x) + \mu h(x)\,.$$

Given a fixed μ and a constant φ, $f + \mu h = \varphi$ characterizes a line intersecting f-axis at $(0, \varphi)$ with slope $-\mu$. So by letting $\varphi = \phi(\mu) = \min_x\ f(x) + \mu h(x)$, we see that $\phi(\mu)$ is the *lowest* intersection between f-axis and a line which intersects the boundary of C and has slope $-\mu$. Namely, $\phi(\mu)$ is the intersection between f-axis and the *supporting lines* to C with slope $-\mu$. So ϕ^* is the *highest* intersection between f-axis and all supporting lines to C.

19.3 Slater's condition for convex optimization

Now we give the Slater's condition for convex optimizations.

Theorem (*Slater's condition*)

If there exists $x \in \mathrm{relint}(D)$ such that $g_i(x) = 0$ for all i and $h_j(x) < 0$ for all j, then strong duality holds.

Remark

Here the relint means *relative interior points*. A point $x \in D$ is a relative interior point if $\mathcal{B}(x, \varepsilon) \cap \operatorname{aff}(D) \subseteq D$, where $\operatorname{aff}(D)$ is the *affine hull* of D. For any convex set D, the relative interior is equivalent to the set of non-extreme points. We can easily see the differences between interior and relative interior. For example,

- the interior of a line segment in an at least two-dimensional space is empty, but its relative interior is the line segment without its endpoints;
- the interior of a disc in an at least three-dimensional space is empty, but its relative interior is the same disc without its circular edge.

Before proving this theorem, we first think about when strong duality does not hold. The geometric interpretation tells us f^* is the lowest intersection of C and f-axis, while ϕ^* is the highest intersection of f-axis and all supporting hyperplanes of C. So $(0, f^*)$ is a point on the boundary of C. Heuristically, if C is convex, there should be a supporting hyperplane passing through $(0, f^*)$, which gives that $\phi^* \geq f^*$.

Unfortunately, this observation is not always true, because there are vertical (orthogonal to h-axis) supporting hyperplanes. Consider the following example (Example 4 above).

Example

$$\begin{aligned} \min_{x_1 \in \mathbb{R},\, x_2 > 0} \quad & e^{-x_1} \\ \text{subject to} \quad & \frac{x_1^2}{x_2} \leq 0 \end{aligned}$$

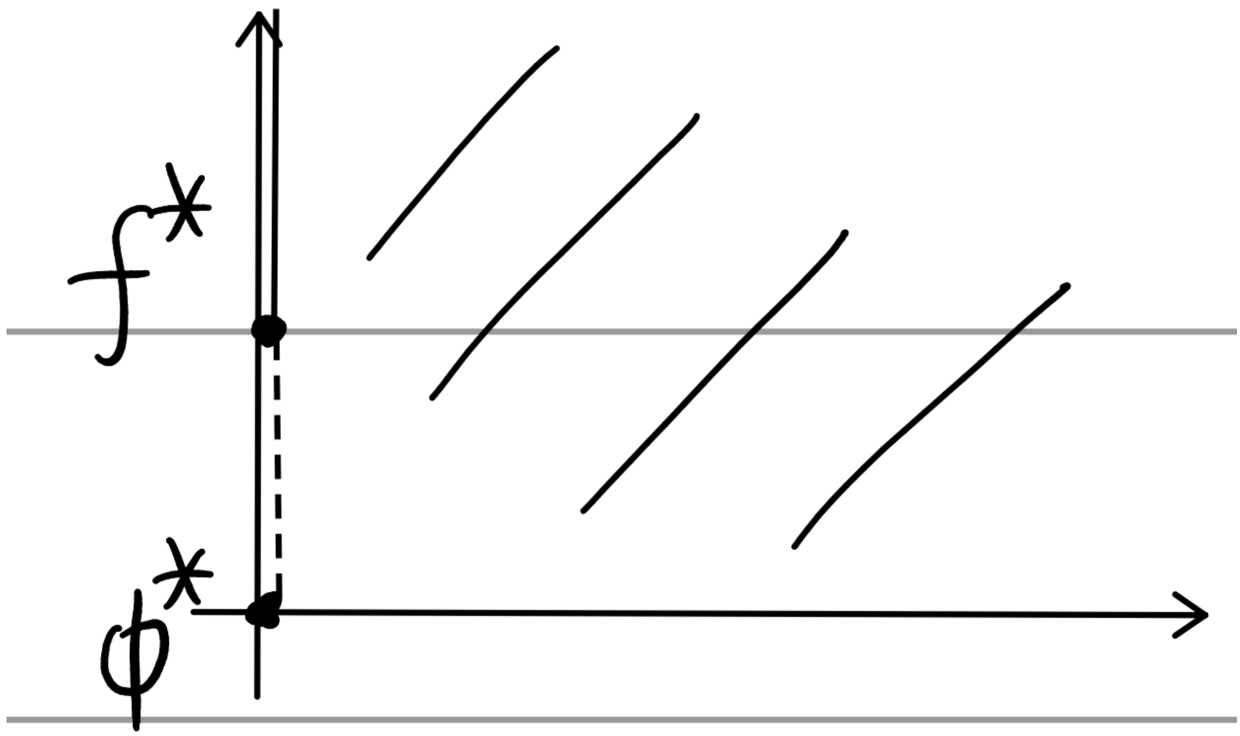
The dual function is $\phi(\mu) = 0$ if $\mu \geq 0$ and $\phi(\mu) = -\infty$ otherwise. So the optimal value to the dual problem $\sup_{\mu \geq 0} \phi(\mu)$ is $\phi^* = 0$, while the optimal value to the primal is $f^* = 1$. Strong duality does not hold.

Why does strong duality fail in this case? We can also define

$$C \triangleq \left\{ (p, t)^\mathsf{T} \mid \exists\, \boldsymbol{x} \in \mathbb{R} \times \mathbb{R}_{>0},\ h(\boldsymbol{x}) \leq p,\ f(\boldsymbol{x}) \leq t \right\}.$$

We can find that

- if $p = 0$, then $x_1^2/x_2 \leq 0$, so $f(\boldsymbol{x}) = 1$;
- if $p > 0$, then $x_1^2/x_2 \leq p$, so $f(\boldsymbol{x}) \to 0$ as $x_1 \to \infty$.
  Thus, $C = \{(0, t)^\top \mid t \geq 1\} \cup \{(p, t)^\top \mid p > 0,\, t > 0\}$. Although C is convex, strong duality does not hold, since the unique supporting hyperplane at $(0, f^*)$ is vertical.



So geometrically, the Slater's condition says that if C is convex, and C in the $h < 0$ region is not empty, then strong duality holds.

We now prove the Slater's condition. We first show that C is convex. Generally, for a convex problem

$$\begin{aligned} \min \quad & f(x) \\ \text{subject to} \quad & g_i(x) = 0, \quad 1 \leq i \leq m \\ & h_j(x) \leq 0, \quad 1 \leq j \leq k \end{aligned}$$

we define

$$C = \left\{(p_1, \cdots, p_k, q_1, \cdots, q_m, t) \in \mathbb{R}^{k+m+1} \mid \exists\, x \in D,\ h_j(x) \leq p_j,\ g_i(x) = q_i,\ f(x) \leq t\right\}.$$

Lemma

C is a convex set.

Proof

Take two points $(p_1, q_1, t_1), (p_2, q_2, t_2) \in C$, so there exists $x_1, x_2 \in D$ such that

$$h(x_1) \le p_1\,,\; q(x_1) = q_1\,,\; f(x_1) \le t_1\,,$$
$$h(x_2) \le p_2\,,\; q(x_2) = q_2\,,\; f(x_2) \le t_2\,.$$

For any $\theta \in [0,1]$, let $y = \theta x_1 + \bar{\theta} x_2$. Then by convexity of f, h_j and affinity of g_i, we have

$$\begin{aligned} &h_j(y) \le \theta h_j(x_1) + \bar{\theta} h_j(x_2) \le \theta p_1 + \bar{\theta} p_2\,,\\ &g_i(y) = \theta g(x_1) + \bar{\theta} g(x_2) = \theta q_1 + \bar{\theta} q_2\,,\\ &f(y) \le \theta f(x_1) + \bar{\theta} f(x_2) \le \theta t_1 + \bar{\theta} t_2\,. \end{aligned}$$

This leads to $\theta(p_1, q_1, t_1) + \bar{\theta}(p_2, q_2, t_2) = (\theta p_1 + \bar{\theta} p_2, \theta q_1 + \bar{\theta} q_2, \theta t_1 + \bar{\theta} t_2) \in C$.

Now we are ready to prove strong duality if Slater's condition is satisfied.

Proof of strong duality

If $f^* = -\infty$, by weak duality, $\phi^* \le f^*$, so $\phi^* = -\infty$. Thus the strong duality holds.
Now assume that $f^* > -\infty$. By Slater's condition, the feasible set $\Omega \ne \emptyset$, so $f^* < \infty$. It suffices to show that there exists a nonvertical supporting hyperplane passing through $(\mathbf{0}, \mathbf{0}, f^*)$.

- Step 1. Prove that $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$. Note that $f^* = \inf\limits_{x \in \Omega} f(x)$. So for all $\varepsilon > 0$, there exists $x \in D$ such that

  $$g_i(x) = 0 \text{ for all } i\,, \quad h_j(x) \le 0 \text{ for all } j\,, \quad f(x) < f^* + \varepsilon\,.$$

  Thus $(\mathbf{0}, \mathbf{0}, f^* + \varepsilon) \in C$. Moreover, for all $\delta > 0$, there does not exists $x \in \Omega$ such that

  $$g_i(x) = 0 \text{ for all } i\,, \quad h_j(x) \le 0 \text{ for all } j\,, \quad f(x) \le f^* - \delta\,.$$

  which gives that $(\mathbf{0}, \mathbf{0}, f^* - \delta) \notin C$. Therefore, $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$.
- Step 2. Show that there exists a nonvertical supporting hyperplane. Since $(\mathbf{0}, \mathbf{0}, f^*) \in \partial C$, there exists a supporting hyperplane passing through $(\mathbf{0}, \mathbf{0}, f^*)$, i.e., there exists $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \xi) \ne \mathbf{0}$ such that

  $$\forall\, (\boldsymbol{p}, \boldsymbol{q}, t) \in C\,, \qquad \boldsymbol{\mu}^\mathsf{T} \boldsymbol{p} + \boldsymbol{\lambda}^\mathsf{T} \boldsymbol{q} + \xi t \ge \xi f^*\,.$$

  Note that for all $t > f^*$, $(\mathbf{0}, \mathbf{0}, t)$ is in C. So we have $\xi \ge 0$. Similarly, for

all $\boldsymbol{p} \geq \mathbf{0}$, $(\boldsymbol{p}, \mathbf{0}, f^*)$ is in C by definition, so we have $\boldsymbol{\mu} \geq \mathbf{0}$.
We now claim that $\xi \neq 0$. If not, $\boldsymbol{\mu}^\top \boldsymbol{p} + \boldsymbol{\lambda}^\top \boldsymbol{q} \geq 0$. By Slater's condition, there exists $\tilde{x}$ such that

$$g_i(\tilde{x}) = 0 \text{ for all } i\,, \qquad h_j(\tilde{x}) < 0 \text{ for all } j\,.$$

Hence there exists $\tilde{t}$ such that

$$\big(h_1(\tilde{x}), \ldots, h_k(\tilde{x}), g_1(\tilde{x}), \ldots, g_m(\tilde{x}), \tilde{t}\big) \in C\,.$$

Thus $\boldsymbol{\mu}^\top \boldsymbol{h}(\tilde{x}) \geq 0$, where $\boldsymbol{h}(x) = (h_1(x), \ldots, h_k(x))$. Since $\boldsymbol{h}(\tilde{x}) < \mathbf{0}$, it implies that $\boldsymbol{\mu} = \mathbf{0}$. Now we obtain that $\boldsymbol{\lambda}^\top \boldsymbol{g}(x) \geq 0$ for all $x \in D$, where $\boldsymbol{g}(x) = (g_1(x), \ldots, g_m(x))^\top = \boldsymbol{A}x - \boldsymbol{b}$. We can show that it is impossible unless $\boldsymbol{\lambda} = \mathbf{0}$.
Suppose $\mathrm{aff}(D) = \{\boldsymbol{x} \mid \boldsymbol{U}\boldsymbol{x} = \boldsymbol{w}\}$ for some $\boldsymbol{U}$ and $\boldsymbol{w}$. Without loss of generality, we may assume the matrix $\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{U} \end{pmatrix}$ has full (row) rank.
Otherwise some constraints are redundant on the domain D. Note that $\tilde{x}$ is a relative interior point of D. If there exists $v \in \mathbb{R}^n$ and $\varepsilon > 0$ such that $\tilde{x} + \varepsilon v \in D$, then there exists $\delta > 0$ such that $\tilde{x} - \delta v \in D$. Since $\boldsymbol{g}$ is affine, and $\boldsymbol{g}(\tilde{x}) = \mathbf{0}$, we have

$$\boldsymbol{\lambda}^\top \boldsymbol{g}(\tilde{x} + \varepsilon v) = \boldsymbol{\lambda}^\top \boldsymbol{g}(\tilde{x}) + \varepsilon \boldsymbol{\lambda}^\top \boldsymbol{g}(v) = \varepsilon \boldsymbol{\lambda}^\top \boldsymbol{g}(v) \qquad \text{and} \qquad \boldsymbol{\lambda}^\top \boldsymbol{g}(\tilde{x} - \delta v) = \boldsymbol{\lambda}$$

If $\boldsymbol{\lambda}^\top \boldsymbol{g}(x) \geq 0$ for all $x \in D$, we can conclude that $\boldsymbol{\lambda}^\top \boldsymbol{g}(x) = 0$ for all $x \in D$. Namely, for all $x \in D$,

$$\boldsymbol{\lambda}^\top \boldsymbol{A}(x - \tilde{x}) = 0\,.$$

On the other hand, $\mathrm{aff}(D) = \{\boldsymbol{x} \mid \boldsymbol{U}\boldsymbol{x} = \boldsymbol{w}\}$, so $\boldsymbol{U}(x - \tilde{x}) = \mathbf{0}$ for every $x \in D$. Using the same argument for the sufficiency of Lagrange multipliers for convex optimization, we conclude that $\boldsymbol{\lambda}^\top \boldsymbol{A}$ is in the row space of $\boldsymbol{U}$. But this is impossible unless $\boldsymbol{\lambda} = \mathbf{0}$, due to the full row rank assumption of $\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{U} \end{pmatrix}$.
Overall, we obtain that if $\xi = 0$, then $\boldsymbol{\mu} = \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$, which contradicts to $(\boldsymbol{\mu}, \boldsymbol{\lambda}, \xi) \neq \mathbf{0}$. It yields that $\xi \neq 0$, and further gives that $\xi > 0$.

- Step 3. Conclude that strong duality holds. Now let $\widehat{\boldsymbol{\mu}} = \boldsymbol{\mu}/\xi$, $\widehat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}/\xi$. For all $(\boldsymbol{p}, \boldsymbol{q}, t) \in C$, since $\boldsymbol{\mu}^\top \boldsymbol{p} + \boldsymbol{\lambda}^\top \boldsymbol{q} + \xi t \geq \xi f^*$, we have

$$\widehat{\boldsymbol{\mu}}^\top \boldsymbol{p} + \widehat{\boldsymbol{\lambda}}^\top \boldsymbol{q} + t \geq f^*\,.$$

Thus,

$$\phi(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{\mu}}) = \inf_{x \in D} f(x) + \widehat{\boldsymbol{\lambda}}^\top \boldsymbol{g}(x) + \widehat{\boldsymbol{\mu}}^\top \boldsymbol{h}(x) \geq f^*,$$

because $(\boldsymbol{h}(x), \boldsymbol{g}(x), f(x)) \in C$ for all $x \in D$. By weak duality $\phi^* \leq f^*$, we conclude that $\phi^* = f^*$.

19.4 An example of solving optimization via duality

Now we show an example of solving optimization problems via duality. Consider the problem finding the projection to the polytope.

$$\begin{aligned} \min_{\boldsymbol{x} \in \mathbb{R}^n} \quad & \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ \text{subject to} \quad & \langle \boldsymbol{w}_i, \boldsymbol{x} \rangle \leq b_i, \quad \forall\, i \in [m]. \end{aligned}$$

We compute its Lagrangian function

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu}) = \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 + \sum_{i=1}^m \mu_i \left(\langle \boldsymbol{w}_i, \boldsymbol{x} \rangle - b_i\right)$$

and let $\phi(\boldsymbol{\mu}) = \inf_{\boldsymbol{x} \in \mathbb{R}^n} \mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu})$. By Slater's condition, to solve the primal optimization problem, it suffices to solve its dual problem

$$\max_{\boldsymbol{\mu} \in \mathbb{R}^m_{\geq 0}} \quad \phi(\boldsymbol{\mu}).$$

Observe that, $\mathcal{L}(x, u)$ is a strongly convex function. Then for $\boldsymbol{\mu} \geq 0$, the minimizer of $\mathcal{L}(\boldsymbol{x}, \boldsymbol{\mu})$ is attained at the point $\boldsymbol{x}(\boldsymbol{\mu}) = \boldsymbol{x}_0 - \sum_{i=1}^m \mu_i \boldsymbol{w}_i$. So it holds that

$$\phi(\boldsymbol{\mu}) = \langle \boldsymbol{\mu}, \mathbf{b} \rangle - \frac{1}{2} \boldsymbol{\mu}^\top \boldsymbol{W} \boldsymbol{\mu}$$

where

$$\begin{aligned} \mathbf{b} &\triangleq \left(\langle \boldsymbol{w}_i, \boldsymbol{x}_0 \rangle - b_i\right)_{i \in [m]} \in \mathbb{R}^m, \\ \boldsymbol{W} &\triangleq \left(\langle \boldsymbol{w}_i, \boldsymbol{w}_j \rangle\right)_{i,j \in [m]} \in \mathbb{R}^{m \times m}. \end{aligned}$$

To (approximately) maximize $\phi(\boldsymbol{\mu})$ we can apply the projected gradient descent.