COMM 616: Modern Optimization with Applications in ML and OR 2024-25 Fall

Lecture 4: Duality Theory I

Instructor: Jiajin Li Scribe: Alex Chow

1 Introduction

Every optimization problem may be viewed as either from the primal or dual - *the principle of duality.* Duality develops the relationship between one optimization problem and another related optimization problem.

- Primal and dual (i.e., there are various dual formulations) enjoy different structures $\rightarrow$ computational tractability.
- Interpretation (Modeling/Economic)

$$\text{min Cost [P]} \longleftrightarrow \text{max Utility [D]}$$

What We Will Cover:

Different ways to construct the dual problems:

(i) Lagrangian Duality – Sion Minimax Theorem

(ii) A general viewpoint: Deriving the duality via studying the stability of an optimization to perturbation.

(a) Special case I: Lagrangian Duality

(b) Special case II: Fenchel-Rockafellar Duality

2 Lagrangian Duality

To begin our investigation, let us focus on the following class of optimization problems (P):

$$\begin{array}{ll} \inf & f(\boldsymbol{x}) \\ \text{subject to} & g_i(\boldsymbol{x}) \leq 0 \quad \text{for } i \in [m_1] \\ & h_j(\boldsymbol{x}) = 0 \quad \text{for } j \in [m_2] \\ & \boldsymbol{x} \in \mathcal{X}. \end{array} \tag{P}$$

Here, $f, g_1, g_2, \cdots, g_{m_1}, h_1, h_2, \cdots, h_{m_2} : \mathbb{R}^d \rightarrow \mathbb{R}$ are arbitrary functions and $\mathcal{X}$ is an arbitrary set. For the sake of brevity, we shall write the first two sets of constraints in (P) as $G(\boldsymbol{x}) \leq \boldsymbol{0}$ and $H(\boldsymbol{x}) = \boldsymbol{0}$, where $G : \mathbb{R}^d \rightarrow \mathbb{R}^{m_1}$ is given by $G(\boldsymbol{x}) = (g_1(\boldsymbol{x}), \cdots, g_{m_1}(\boldsymbol{x}))$ and $H : \mathbb{R}^d \rightarrow \mathbb{R}^{m_2}$ is given by $H(\boldsymbol{x}) = (h_1(\boldsymbol{x}), \cdots, h_{m_2}(\boldsymbol{x}))$.

One way of constructing a dual of (P) is to reformulate it using a penalty function approach. We observe that (P) is equivalent to

$$v_p^\star = \inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})$$

where $L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})$ is the Lagrangian function defined as:

$$L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega}) := f(\boldsymbol{x}) + \boldsymbol{\nu}^\top G(\boldsymbol{x}) + \boldsymbol{\omega}^\top H(\boldsymbol{x}).$$

This follows from the fact that for any $\boldsymbol{x} \in \mathcal{X}$:

$$\sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega}) = \begin{cases} f(\boldsymbol{x}) & \text{if } G(\boldsymbol{x}) \leq 0 \text{ and } H(\boldsymbol{x}) = 0 \\ +\infty & \text{otherwise.} \end{cases}$$

Now, it's clear that for any $\bar{\boldsymbol{x}} \in \mathcal{X}$ and $(\bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$,

$$\inf_{\boldsymbol{x} \in \mathcal{X}} L(\boldsymbol{x}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \leq L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \leq \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} L(\bar{\boldsymbol{x}}, \boldsymbol{\nu}, \boldsymbol{\omega})$$

Hence, we have:

$$\underbrace{\sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} \inf_{\boldsymbol{x} \in \mathcal{X}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})}_{\text{(Dual)}} \leq \underbrace{\inf_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})}_{\text{(Primal)}}$$

where the left hand side is defined as Lagrangian Dual and the right hand side is the primal problem.

Remark 1. *It is worth noting that since the set $\mathcal{X}$ is arbitrary, there are many different Lagrangian duals for the same primal problem. Different choices of dual problems will lead to different outcomes, both in terms of the dual optimal value and the computational efforts required to solve the dual problem. Whether it is easy to solve a particular choice of dual problem depends on the structure of the problem.*

We observe from the formulation of the dual problem, i.e.,

$$v_d^\star = \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} \inf_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) + \boldsymbol{\nu}^\top G(\boldsymbol{x}) + \boldsymbol{\omega}^\top H(\boldsymbol{x}). \tag{D}$$

Regardless of the structures of f, G, H, the dual problem is always concave because the pointwise infimum preserves concavity. Such an observation suggests the strong duality ($v_p^\star = v_d^\star$) may not hold. However, it is clear that the weak duality ($v_p^\star \geq v_d^\star$) always holds.

Example 2 (A Primal-Dual Pair with a Non-Zero Duality Gap)**.** *Consider the following primal problem:*

$$\begin{array}{rl} \min & -x \\ \textit{subject to} & x \leq 1 \\ & x \in \{0, 2\}. \end{array} \tag{1}$$

It is easy to get $v_p^\star = 0$ and $x^\star = 0$. Then, we derive its Lagrangian dual as:

$$v_d^\star = \sup_{\nu \geq 0} \inf_{x \in \{0,2\}} (-x + \nu(x-1)) = \min\{-\nu, \nu - 2\},$$

which gives

$$v_d^\star = -1 \quad \textit{and} \quad v^\star = 1.$$

This shows a non-zero duality gap.

It is natural to ask

When does the strong duality hold?

Definition 3 (Saddle Point)**.** $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \in \mathbb{R}^d \times \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$ *is a* ***saddle point*** *of the Lagrangian function associated with* (P) *if the following conditions are satisfied:*

(i) $\bar{\boldsymbol{x}} \in \mathcal{X}$

(ii) $\bar{\boldsymbol{\nu}} \geq 0$

(iii) For all $\boldsymbol{x} \in \mathcal{X}$ *and* $(\boldsymbol{\nu}, \boldsymbol{\omega}) \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$*:*

$$L(\bar{\boldsymbol{x}}, \boldsymbol{\nu}, \boldsymbol{\omega}) \leq L(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \leq L(\boldsymbol{x}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}).$$

It implies that $\bar{\boldsymbol{x}}$ *minimizes* L *over* $\mathcal{X}$ *when* $(\boldsymbol{\nu}, \boldsymbol{\omega})$ *is fixed at* $(\bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}})$*. Similarly,* $(\bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}})$ *maximizes* L *over* $\mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$ *when* $\boldsymbol{x}$ *is fixed at* $\bar{\boldsymbol{x}}$*.*

Theorem 4. *The point* $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}}) \in \mathbb{R}^d \times \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$ *is a saddle point of the Lagrangian function* L *associated with* (P) *if and only if the duality gap between* (P) *and* (D) *is zero.* $\bar{\boldsymbol{x}}$ *and* $(\bar{\boldsymbol{\nu}}, \bar{\boldsymbol{\omega}})$ *are the optimal solutions of* (P) *and* (D) *respectively.*

$$\textit{Existence of saddle point} \quad \Longleftrightarrow \quad \textit{Strong Duality}$$

The above theorem is difficult to apply, as it essentially implies that verifying strong duality requires solving both the primal and dual problems first. A natural question arises: Can the same relationship hold under different conditions? The following result, a special case of Sion's minimax theorem [5], provides one such possibility:

Theorem 5 (A Special Case of Sion's Minimax Theorem)**.** *Let* L *be the Lagrangian function associated with* (P)*. Suppose that*

(i) $\mathcal{X}$ *is a compact convex subset of* $\mathbb{R}^d$

(ii) $(\boldsymbol{\nu}, \boldsymbol{\omega}) \mapsto L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})$ *is continuous and concave on* $\mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$ *for each* $\boldsymbol{x} \in \mathcal{X}$

(iii) $\boldsymbol{x} \mapsto L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega})$ *is continuous and convex on* $\mathcal{X}$ *for each* $(\boldsymbol{\nu}, \boldsymbol{\omega}) \in \mathbb{R}_+^{m_1} \times \mathbb{R}^{m_2}$

Then, we have

$$\sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} \min_{\boldsymbol{x} \in \mathcal{X}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega}) = \min_{\boldsymbol{x} \in \mathcal{X}} \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} L(\boldsymbol{x}, \boldsymbol{\nu}, \boldsymbol{\omega}).$$

3 Perturbations and Bifunctions

When solving an optimization problem, one is often interested not only in the optimal value but also in how this value changes under small perturbations.
More formally, consider an arbitrary convex function $F : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ and a parametric optimization problem:

$$p(\boldsymbol{y}) = \inf_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}, \boldsymbol{y})$$

where $\boldsymbol{y}$ is the perturbation parameter. Here, $F(\boldsymbol{x}, \boldsymbol{y})$ can be viewed as a perturbation function around the original problem

$$p(\mathbf{0}) := \inf_{\boldsymbol{x} \in \mathcal{X}} F(\boldsymbol{x}, \mathbf{0}). \tag{Pe}$$

Remark 6. *A natural goal is to study how* $p(\boldsymbol{y})$ *changes when* $\boldsymbol{y}$ *is perturbed around* $\mathbf{0}$*. It turns out that* $\partial p(\mathbf{0})$ *coincides with the optimal solution of the dual problem (see [2, Section 4.2]). There are numerous elegant results connecting subdifferential calculus rules with duality theory.*

Now, we define the new parametric family of problems:

$$q(\boldsymbol{x}) = \sup_{\boldsymbol{y}} -F^*(\boldsymbol{x}, \boldsymbol{y}).$$

Then, we have the parametric dual at the origin, i.e.,

$$q(\mathbf{0}) = \sup_{\boldsymbol{y}} -F^*(\mathbf{0}, \boldsymbol{y}). \tag{De}$$

Next, we aim to connect the primal and dual problems using the following results on the biconjugate of convex functions.

Proposition 7. *If $f : \mathbb{R}^d \to \overline{\mathbb{R}}$ is proper, lower-semicontinuous and convex, we have $f^{**} = f$ [3, Theorem 12.2].*

Claim 8 (Inf-projection)**.** $p^*(\boldsymbol{y}) = F^*(\mathbf{0}, \boldsymbol{y})$.

Proof.

$$\begin{aligned} p^*(\boldsymbol{y}) &= \sup_{\boldsymbol{s}} \left(\boldsymbol{s}^\top \boldsymbol{y} - p(\boldsymbol{s}) \right) \\ &= \sup_{\boldsymbol{s}} \left(\boldsymbol{s}^\top \boldsymbol{y} - \inf_{\boldsymbol{x}} F(\boldsymbol{x}, \boldsymbol{s}) \right) \\ &= \sup_{\boldsymbol{s}, \boldsymbol{x}} \left(\boldsymbol{s}^\top \boldsymbol{y} - F(\boldsymbol{x}, \boldsymbol{s}) \right) \\ &= \sup_{\boldsymbol{s}, \boldsymbol{x}} \left((\boldsymbol{x}, \boldsymbol{s})^\top (\mathbf{0}, \boldsymbol{y}) - F(\boldsymbol{x}, \boldsymbol{s}) \right) \\ &= F^*(\mathbf{0}, \boldsymbol{y}). \end{aligned}$$

■

Moreover, we have

$$p^{**}(\mathbf{0}) = \sup_{\boldsymbol{y}} \left(\langle 0, \boldsymbol{y} \rangle - p^*(\boldsymbol{y}) \right) = \sup_{\boldsymbol{y}} -F^*(\mathbf{0}, \boldsymbol{y}) \leq \boldsymbol{p}(\mathbf{0}).$$

where the second equality follows from Claim 8. To make the last inequality an equality, we need to apply Proposition 7 and have:

Theorem 9 (General Duality Theorem)**.** *Suppose that $F : \mathcal{X} \times \mathcal{Y} \to \overline{\mathbb{R}}$ is proper, lower-semicontinuous, and convex. Then, if $\mathbf{0} \in \operatorname{int}(\operatorname{dom}(p))$, strong duality holds.*

We refer the details proof to [4, Theorem 11.39].

Definition 10 (Interior Point)**.** *Given an arbitrary set $\mathcal{S} \subseteq \mathbb{R}^d$, its interior is defined by:*

$$\operatorname{int}(\mathcal{S}) = \{ \boldsymbol{x} \in \mathcal{S} : B(\boldsymbol{x}, \epsilon) \subseteq \mathcal{S}, \ \textit{for some } \epsilon > 0 \}.$$

Example 11.

$$\mathcal{S} = [0, 1] \quad \textit{with} \quad \operatorname{int}(\mathcal{S}) = (0, 1).$$

3.1 Recovering the General Lagrangian Duality

We revisit (P) as

$$\begin{array}{ll} \inf & f(\boldsymbol{x}) \\ \text{subject to} & G(\boldsymbol{x}) \leq \mathbf{0} \\ & H(\boldsymbol{x}) = \mathbf{0} \\ & \boldsymbol{x} \in \mathcal{X}, \end{array} \tag{2}$$

which is equivalent to

$$\inf_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) + \theta(A(\boldsymbol{x})).$$

where

$$A(\boldsymbol{x}) = (G(\boldsymbol{x}), H(\boldsymbol{x})) \quad \text{and} \quad \theta(A(\boldsymbol{x})) = \mathbb{1}_{\{\boldsymbol{x} \in \mathbb{R}^d : G(\boldsymbol{x}) \leq \mathbf{0}\}}(\boldsymbol{x}) + \mathbb{1}_{\{\boldsymbol{x} \in \mathbb{R}^d : H(\boldsymbol{x}) = \mathbf{0}\}}(\boldsymbol{x}).$$

Now, we are ready to construct the parametric optimization problem:

$$p(\boldsymbol{y}) = \inf_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) + \theta(\mathbf{A}(\boldsymbol{x}) + \boldsymbol{y}) \quad \textit{with} \quad \boldsymbol{y} = (\boldsymbol{\nu}, \boldsymbol{\omega}).$$

Then, the dual problem admits:

$$\begin{aligned}
p^{**}(\mathbf{0}) &= \sup_{\boldsymbol{y}} -F^*(\mathbf{0}, \boldsymbol{y}) \\
&= \sup_{\boldsymbol{y}} - \sup_{\boldsymbol{x},\boldsymbol{u}} (\boldsymbol{x}, \boldsymbol{u})^\top (\mathbf{0}, \boldsymbol{y}) - F(\boldsymbol{x}, \boldsymbol{u}) \\
&= \sup_{\boldsymbol{y}} \inf_{\boldsymbol{x},\boldsymbol{u}} F(\boldsymbol{x}, \boldsymbol{u}) - (\boldsymbol{x}, \boldsymbol{u})^\top (\mathbf{0}, \boldsymbol{y}) \\
&= \sup_{\boldsymbol{y}} \inf_{\boldsymbol{x},\boldsymbol{u}} f(\boldsymbol{x}) + \theta(\mathbf{A}(\boldsymbol{x}) + \boldsymbol{u}) - \boldsymbol{u}^\top \boldsymbol{y} \\
&= \sup_{\boldsymbol{y}} \inf_{\boldsymbol{x}} f(\boldsymbol{x}) + \inf_{\boldsymbol{u}} \theta(A(\boldsymbol{x}) + \boldsymbol{u}) - \boldsymbol{u}^T \boldsymbol{y}.
\end{aligned}$$

Moreover, we know

$$\begin{aligned}
&\inf_{\boldsymbol{u}} \theta(A(\boldsymbol{x}) + \boldsymbol{u}) - \boldsymbol{u}^T \boldsymbol{y} \\
=& \inf_{\boldsymbol{u}_1} \mathbb{1}_{\{\boldsymbol{x} \in \mathbb{R}^d : G(\boldsymbol{x}) + \boldsymbol{u}_1 \leq \mathbf{0}\}}(\boldsymbol{x}) - \boldsymbol{u}_1^T \boldsymbol{\nu} + \inf_{\mu_2} \mathbb{1}_{\{\boldsymbol{x} \in \mathbb{R}^d : H(\boldsymbol{x}) + \boldsymbol{u}_2 = \mathbf{0}\}}(\boldsymbol{x}) - \boldsymbol{u}_2^T \boldsymbol{\omega} \\
=& \begin{cases} G(\boldsymbol{x})^T \boldsymbol{\nu} + H(\boldsymbol{x})^T \boldsymbol{\omega} & \text{if } \boldsymbol{\nu} \leq 0 \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}$$

Combining them together, we have

$$p^{**}(0) = \sup_{\boldsymbol{\nu} \in \mathbb{R}_+^{m_1}, \boldsymbol{\omega} \in \mathbb{R}^{m_2}} \inf_{\boldsymbol{x}} f(\boldsymbol{x}) + \nu^T G(\boldsymbol{x}) + \boldsymbol{w}^T H(\boldsymbol{x}).$$

Remark 12. *The regularity condition* $\mathbf{0} \in \operatorname{int}(\operatorname{dom}(p))$ *implies the standard slater condition — exists* $\boldsymbol{x}$ *such that* $G(\boldsymbol{x}) < 0$.

4 Further Reading

Please refer to [4, Chapter 11.H] and [1, Chapter 2.5] for further reading.

References

[1] J. Frédéric Bonnans and Alexander Shapiro. *Perturbation Analysis of Optimization Problems*. Springer Science & Business Media, 2013. 5

[2] Dmitriy Drusvyatskiy. *Convex Analysis and Nonsmooth Optimization*. 2020. 3

[3] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970. 4

[4] R. Tyrrell Rockafellar and Roger J-B Wets. *Variational Analysis*, volume 317. Springer Science & Business Media, 2009. 4, 5

[5] Maurice Sion. On general minimax theorems. 1958. 3