Math 273A Notes:
Chapter 3

Ernest K. Ryu

December 5, 2025

1 Optimization duality

Maximin-minimax inequality In many introductory texts of convex optimization, one starts with a primal optimization problem and finds a corresponding dual problem. Here, we take a slightly different viewpoint. We view the primal and dual problems as the two halves of a larger saddle point problem.

Let $\mathbf{L}\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$. We say $\mathbf{L}(x, y)$ is convex-concave if $\mathbf{L}$ is convex in x when y is fixed and concave in y when x is fixed. We say $(x^\star, y^\star)$ is a saddle point of $\mathbf{L}$ if

$$\mathbf{L}(x^\star, y) \le \mathbf{L}(x^\star, y^\star) \le \mathbf{L}(x, y^\star) \qquad \forall\, x \in \mathbb{R}^n,\, y \in \mathbb{R}^m.$$

We call

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \sup_{y\in\mathbb{R}^m} \mathbf{L}(x, y)$$

the *primal problem* generated by $\mathbf{L}$ and write $p^\star = \inf_x \sup_y \mathbf{L}(x, y)$ for the primal optimal value. We call

$$\underset{y\in\mathbb{R}^m}{\text{maximize}} \quad \inf_{x\in\mathbb{R}^n} \mathbf{L}(x, y)$$

the *dual problem* generated by $\mathbf{L}$ and write $d^\star = \sup_y \inf_x \mathbf{L}(x, y)$ for the dual optimal value. In most engineering settings, one starts with an optimization problem, not a convex-concave saddle function. With this view of duality, the trick is to find a convex-concave saddle function that generates the primal problem of interest.

Example. Let f be a CCP function on $\mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Consider the Lagrangian

$$\mathbf{L}(x, y) = f(x) + \langle y, Ax - b\rangle, \tag{1}$$

which generates the primal problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax = b \end{array} \tag{2}$$

and dual problem

$$\underset{y\in\mathbb{R}^m}{\text{maximize}} \quad -f^*(-A^\intercal y) - b^\intercal y. \tag{3}$$

The dual variable y is also called the Lagrange multipliers.

Example. Consider the Lagrangian

$$\mathbf{L}(x,y) = f(x) + \langle y, Ax\rangle - g^*(y), \tag{4}$$

which generates the primal problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + g^{**}(Ax) \tag{5}$$

and dual problem

$$\underset{y\in\mathbb{R}^m}{\text{maximize}} \quad -f^*(-A^\intercal y) - g^*(y). \tag{6}$$

This primal-dual problem pair is sometimes called the Fenchel–Rockafellar dual.

An *augmented Lagrangian* is a saddle function that has additional terms while sharing the same saddle points as its unaugmented counterpart.

Example. Consider the Lagrangian

$$\mathbf{L}(x,u) = f(x) + \langle u, Ax - b\rangle$$

with the associated primal problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax = b. \end{array}$$

We will often use the augmented Lagrangian

$$\mathbf{L}_\rho(x,u) = f(x) + \langle u, Ax - b\rangle + \frac{\rho}{2}\|Ax-b\|^2 \tag{7}$$

with $\rho > 0$. It is straightforward to show that (x,u) is a saddle point of $\mathbf{L}$ if and only if it is a saddle point of $\mathbf{L}_\rho$ for any $\rho > 0$.

Example: Dual of the LASSO Problem Consider the optimization problem

$$\min_{x\in\mathbb{R}^n} \; \frac{1}{2}\|Ax-b\|_2^2 + \lambda\|x\|_1, \tag{8}$$

with $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $\lambda > 0$. We derive its Fenchel dual.

Step 1. Write the objective in the form

$$f(Ax) + g(x),$$

where

$$f(z) = \frac{1}{2}\|z-b\|^2, \qquad g(x) = \lambda\|x\|_1.$$

The Fenchel–Rockafellar dual is

$$\max_{y\in\mathbb{R}^m} \left\{-f^*(y) - g^*(-A^\top y)\right\}.$$

Step 2. Conjugate the quadratic term For $f(z) = \frac{1}{2}\|z-b\|^2$, one computes

$$f^*(y) = \sup_z \left(y^\top z - \frac{1}{2}\|z-b\|^2\right) = y^\top b + \frac{1}{2}\|y\|^2.$$

Step 3. Conjugate the ℓ_1-term. Since the conjugate of $\lambda\|x\|_1$ is the indicator of the ℓ_∞-ball of radius λ,

$$g^*(s) = \delta_{\{\|s\|_\infty \le \lambda\}}(s),$$

we obtain

$$g^*(-A^\top y) = \begin{cases} 0, & \|A^\top y\|_\infty \le \lambda, \\ +\infty, & \text{otherwise.} \end{cases}$$

Step 4. Substituting the conjugates into the dual expression gives

$$\max_y \left(-y^\top b - \frac{1}{2}\|y\|^2\right) \quad \text{s.t.} \quad \|A^\top y\|_\infty \le \lambda.$$

Equivalently, completing the square,

$$\max_{\|A^\top y\|_\infty \le \lambda} \left(-\frac{1}{2}\|y+b\|^2 + \frac{1}{2}\|b\|^2\right),$$

where the constant term $\frac{1}{2}\|b\|^2$ may be omitted.

$$\boxed{\begin{aligned} \text{Dual:} \quad & \max_{y\in\mathbb{R}^m} -\frac{1}{2}\|y\|^2 - b^\top y \\ & \text{subject to } \|A^\top y\|_\infty \le \lambda. \end{aligned}} \tag{9}$$

Weak duality States $d^\star \le p^\star$, always holds. To prove this, note that for any x, u we have

$$\begin{aligned} \inf_x \mathbf{L}(x,u) &\le \mathbf{L}(x,u) \\ \sup_u \inf_x \mathbf{L}(x,u) &\le \sup_u \mathbf{L}(x,u) \\ d^\star = \sup_u \inf_x \mathbf{L}(x,u) &\le \inf_x \sup_u \mathbf{L}(x,u) = p^\star. \end{aligned}$$

Lemma 1 (Maximin-minimax inequality)**.** *Let $L\colon X \times Y \to \mathbb{R}$ be an arbitrary function. Then,*

$$\sup_{y\in Y} \inf_{x\in X} L(x,y) \le \inf_{x\in X} \sup_{y\in Y} L(x,y).$$

Proof. This follows from

$$\begin{aligned} L(x,y) &\le \sup_{y\in Y} L(x,y), \qquad \forall x \in X,\ y\in Y \\ \inf_{x\in X} L(x,y) &\le \inf_{x\in X}\sup_{y\in Y} L(x,y), \qquad \forall y \in Y \\ \sup_{y\in Y}\inf_{x\in X} L(x,y) &\le \inf_{x\in X}\sup_{y\in Y} L(x,y). \end{aligned}$$

□

General weak duality Let $L\colon X\times Y\to\mathbb{R}$ be an arbitrary function. Define $f\colon X\to\mathbb{R}\cup\{\infty\}$ and $g\colon Y\to\mathbb{R}\cup\{-\infty\}$ as

$$f(x)=\sup_{y\in Y} L(x,y) \qquad g(y)=\inf_{x\in X} L(x,y)$$

We call

$$\underset{x\in X}{\text{minimize}} \quad f(x) \tag{P}$$

the primal problem with optimal value $p_\star \in [-\infty,\infty]$

$$\underset{y\in Y}{\text{maximize}} \quad g(y) \tag{D}$$

the dual problem with optimal value $d_\star \in [-\infty,\infty]$.

Theorem 1 (General weak duality)**.** *For the primal and dual optimization problems defined above, we have*

$$d_\star = \sup_{y\in Y} g(y) \le \inf_{x\in X} f(x) = p_\star.$$

Proof. Immediate consequence of the maximin-minimax inequality. □

Primal-dual pair via Lagrangian L

$$\begin{array}{ccc} f(x)=\sup_{y\in Y} L(x,y) & & g(y)=\inf_{x\in X} L(x,y) \\ \underset{x\in X}{\text{minimize}} \quad f(x) & \overset{\text{dual}}{\longleftrightarrow} & \underset{y\in Y}{\text{maximize}} \quad g(y) \end{array}$$

We call L a *Lagrangian*. (Terminology comes from method of Lagrange multipliers.) Pick any L, and we get a primal-dual pair of problems. If we pick L such that the primal problem becomes our problem of interest, then we have a useful corresponding dual problem.

Strong duality. States $d^\star = p^\star$, holds often but not always in convex optimization. Regularity conditions that ensure strong duality are sometimes called constraint qualifications.

Sion's Minimax Theorem Sion's minimax theorem is a powerful theorem with a wide range of applications. Although it cannot be used to establish strong duality in our context (due to the lack of compactness), it gives us a sense of why strong duality is "morally" the right thing to expect.

Theorem 2 (Sion, 1958)**.** *Let X be a convex subset of a linear topological space, and let Y be a convex subset of a linear topological space. Assume X or Y (or both) is compact. Let $F\colon X \times Y \to \mathbb{R}$ satisfy:*

- *For every fixed $y \in Y$, the function $x \mapsto F(x, y)$ is convex in x*
- *For every fixed $x \in X$, the function $y \mapsto F(x, y)$ is concave in y.*

Then the minimax identity holds:

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

The actual Sion's theorem generalizes the convex-concavity condition slightly. However, the compactness condition is crucial.

Total duality. States that a primal solution exists, a dual solution exists, and strong duality holds. Total duality holds if and only if $\mathbf{L}$ has a saddle point. Solving the primal and dual optimization problems is equivalent to finding a saddle point of the saddle function generating the primal and dual problems, provided that total duality holds.

Let us prove the equivalence. Assume $\mathbf{L}$ has a saddle point $(x^\star, u^\star)$. Then

$$\begin{aligned}
\mathbf{L}(x^\star, u^\star) &= \inf_x \mathbf{L}(x, u^\star) \\
&\le \sup_u \inf_x \mathbf{L}(x,u) = d^\star \\
&\le \inf_x \sup_u \mathbf{L}(x,u) = p^\star \\
&\le \sup_u \mathbf{L}(x^\star, u) = \mathbf{L}(x^\star, u^\star),
\end{aligned}$$

and equality holds throughout. Since $\inf_x \sup_u \mathbf{L}(x,u) = \sup_u \mathbf{L}(x^\star, u)$, $x^\star$ is a primal solution. Since $\inf_x \mathbf{L}(x, u^\star) = \sup_u \inf_x \mathbf{L}(x,u)$, $u^\star$ is a dual solution. Since $d^\star = \sup_u \inf_x \mathbf{L}(x,u) = \inf_x \sup_u \mathbf{L}(x,u) = p^\star$, strong duality holds.

On the other hand, assume total duality holds and $x^\star$ and $u^\star$ are primal and dual solutions. Then

$$\begin{aligned}
\inf_x \mathbf{L}(x, u^\star) &= \sup_u \inf_x \mathbf{L}(x,u) = d^\star \\
&= \inf_x \sup_u \mathbf{L}(x,u) = p^\star \\
&= \sup_u \mathbf{L}(x^\star, u).
\end{aligned}$$

Since

$$\mathbf{L}(x^\star, u^\star) \le \sup_u \mathbf{L}(x^\star, u) = \inf_x \mathbf{L}(x, u^\star) \le \mathbf{L}(x^\star, u^\star),$$

equality holds throughout and we conclude

$$\sup_u \mathbf{L}(x^\star, u) = \mathbf{L}(x^\star, u^\star) = \inf_x \mathbf{L}(x, u^\star),$$

i.e., $(x^\star, u^\star)$ is a saddle point.

2 ADMM

Let f and g be convex, $A \in \mathbb{R}^{n\times p}$, $B \in \mathbb{R}^{n\times q}$, and $c \in \mathbb{R}^n$. Consider the primal

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x) + g(y) \\ \text{subject to} & Ax + By = c \end{array}$$

and the dual problem

$$\underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -f^*(-A^\intercal u) - g^*(-B^\intercal u) - c^\intercal u$$

generated by the Lagrangian

$$\mathbf{L}(x, y, u) = f(x) + g(y) + \langle u, Ax + By - c\rangle.$$

We will use the augmented Lagrangian:

$$\mathbf{L}_\rho(x, y, u) = f(x) + g(y) + \langle u, Ax + By - c\rangle + \frac{\rho}{2}\|Ax + By - c\|^2.$$

The algorithm alternating direction method of multipliers (ADMM) is

$$\begin{aligned} x_{k+1} &\in \operatorname*{argmin}_x L_\rho(x, z_k, y_k) \\ z_{k+1} &\in \operatorname*{argmin}_z L_\rho(x_k, z, y_k) \\ y_{k+1} &= y_k + \rho(Ax_{k+1} + Bz_{k+1} - c) \end{aligned}$$

ADMM Convergence via the Summability Lemma

$$\begin{aligned} x_{k+1} &\in \operatorname*{argmin}_x \left\{ f(x) + g(z_k) + \langle y_k, Ax + Bz_k - c\rangle + \frac{\rho}{2}\|Ax + Bz_k - c\|^2 \right\} \\ z_{k+1} &\in \operatorname*{argmin}_z \left\{ f(x_{k+1}) + g(z) + \langle y_k, Ax_{k+1} + Bz - c\rangle + \frac{\rho}{2}\|Ax_{k+1} + Bz - c\|^2 \right\} \\ y_{k+1} &= y_k + \rho(Ax_{k+1} + Bz_{k+1} - c) \end{aligned}$$

Theorem 3. *Assume total duality holds, i.e., assume the unaugmented Lagrangian L_0 has a saddle point $(x^\star, z^\star, y^\star)$. Assume the iterates $\{x_k\}_k$ and $\{z_k\}_k$ are well defined (exists, but need not unique). Then,*

$$Ax_k + Bz_k - c \to 0, \qquad f(x_k) + g(z_k) \to p_\star.$$

(It is also true that $y_k \to y_\star$, where $y_\star$ is a dual solution, but we will not prove this.)

Proof. Introduce the nonnegative Lyapunov function

$$V_k = \frac{1}{2\rho}\|y_k - y^\star\|_2^2 + \frac{\rho}{2}\|B(z_k - z^\star)\|_2^2.$$

Define the primal residual

$$r_{k+1} := Ax_{k+1} + Bz_{k+1} - c,$$

We will show that

$$V_{k+1} \overset{(iii)}{\leq} V_k - \rho\|r_{k+1}\|_2^2 - \rho\|B(z_{k+1} - z_k)\|_2^2.$$

This immediately implies that $r_k \to 0$ by the summability argument. This also implies $B(z_{k+1} - z_k) \to 0$.

Next, let

$$p_k = f(x_k) + g(z_k).$$

Then we have

$$-\langle y_\star, r_{k+1}\rangle \overset{(i)}{\leq} p_{k+1} - p_\star \overset{(ii)}{\leq} -\langle y_{k+1}, r_{k+1}\rangle - \rho\langle B(z_{k+1} - z_k), B(z_{k+1} - z_\star) - r_{k+1}\rangle.$$

This shows that $p_k \to p_\star$. It remains to show inequalities (i), (ii), and (iii).

Inequality (i).

$$p_\star = f(x_\star) + g(z_\star) + \langle y_\star, \underbrace{Ax_\star + Bz_\star - c}_{=0}\rangle = L_0(x_\star, z_\star, y_\star)$$

$$\leq L_0(x_{k+1}, z_{k+1}, y_\star) = f(x_{k+1}) + g(z_{k+1}) + \langle y_\star, \underbrace{Ax_{k+1} + Bz_{k+1} - c}_{=r_{k+1}}\rangle = p_{k+1} + \langle y_\star, r_{k+1}\rangle$$

Inequality (ii).

Lemma 2. *Let $h\colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be convex and $M \in \mathbb{R}^{m\times n}$. If*

$$x_\circ \in \operatorname*{argmin}_x \left\{h(x) + \frac{1}{2}\|M(x - x_\circ)\|^2\right\}$$

then,

$$x_\circ \in \operatorname*{argmin}_x \left\{h(x)\right\}$$

Proof. For simplicity, assume h is differentiable at $h_\circ$. Then, the first condition implies that

$$0 = \nabla_x\left\{h(x) + \frac{1}{2}\|M(x - x_\circ)\|^2\right\}\Big|_{x=x_\circ} = h(x_\circ).$$

By convexity of h, stationarity of h at $x_\circ$ implies optimality of h at $x_\circ$. The general proof can be done using subgradients. $\square$

We can show that

$$\begin{aligned} x_{k+1} &= \underset{x}{\operatorname{argmin}}\{f(x) + \langle y_k, Ax\rangle + \frac{\rho}{2}\|Ax + Bz_k - c\|^2\} \\ &= \underset{x}{\operatorname{argmin}}\{f(x) + \langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax\rangle + \frac{\rho}{2}\|A(x - x_{k+1})\|^2\} \end{aligned}$$

By the lemma, this implies

$$f(x_{k+1}) + \langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax_{k+1}\rangle \leq f(x_\star) + \langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax_\star\rangle \tag{10}$$

We can also show that

$$z_{k+1} = \underset{z}{\operatorname{argmin}}\left\{g(z) + \langle y_{k+1}, Bz\rangle + \frac{\rho}{2}\|B(z - z_{k+1})\|^2\right\}$$

By the lemma, this implies

$$g(z_{k+1}) + \langle y_{k+1}, Bz_{k+1}\rangle \leq g(z_\star) + \langle y_{k+1}, Bz_\star\rangle \tag{11}$$

Adding (10) and (11), we get

$$\begin{aligned} &f(x_{k+1}) + g(z_{k+1}) + \langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax_{k+1}\rangle + \langle y_{k+1}, Bz_{k+1}\rangle \\ &\quad\leq f(x_\star) + \langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax_\star\rangle + g(z_\star) + \langle y_{k+1}, Bz_\star\rangle \end{aligned}$$

reorganizing, we get

$$\begin{aligned} p_{k+1} - p_\star &\leq -\langle y_{k+1} - \rho B(y_{k+1} - y_k), Ax_{k+1} - Ax_\star\rangle - \langle y_{k+1}, Bz_{k+1} - Bz_\star\rangle \\ &= -\langle y_{k+1}, Ax_{k+1} - Ax_\star + Bz_{k+1} - Bz_\star\rangle + \langle \rho B(y_{k+1} - y_k), Ax_{k+1} - Ax_\star\rangle \\ &= -\langle y_{k+1}, \underbrace{Ax_{k+1} + Bz_{k+1} - c}_{=r_{k+1}}\rangle - \rho\langle B(y_{k+1} - y_k), B(z_{k+1} - z_\star) - r_{k+1}\rangle \end{aligned}$$

where in the last step we substitute $Ax_\star + Bz_\star = c$, $Ax_{k+1} = r_{k+1} - Bz_{k+1} + c$, and $Ax_\star = -Bz_\star + c$. We now have inequality (ii).

Inequality (iii). Consider

$$2 \cdot (\text{Ineq (i)} + \text{Ineq (ii)})$$

which gives us

$$-2\langle y_\star, r_{k+1}\rangle \leq -2\langle y_{k+1}, r_{k+1}\rangle - 2\rho\langle B(z_{k+1} - z_k), B(z_{k+1} - z_\star) - r_{k+1}\rangle$$

which can be reorganized to

$$2\langle y_{k+1} - y_\star, r_{k+1}\rangle + 2\rho\langle B(z_{k+1} - z_k), B(z_{k+1} - z_\star)\rangle - 2\rho\langle B(z_{k+1} - z_k), r_{k+1}\rangle \leq 0$$

Using $y_{k+1} = y_k + \rho r_{k+1}$, we can show

$$2\langle y_{k+1} - y_\star, r_{k+1}\rangle = \frac{1}{\rho}\left(\|y_{k+1} - y_\star\|^2 - \|y_k - y_\star\|^2\right) + \rho\|r_{k+1}\|^2.$$

Then, we can show

$$\begin{aligned}&\rho\|r_{k+1}\|^2 + 2\rho\langle B(z_{k+1}-z_k), B(z_{k+1}-z_\star)\rangle - 2\rho\langle B(z_{k+1}-z_k), r_{k+1}\rangle\\&= \rho\|r_{k+1} - B(z_{k+1}-z_k)\|^2 + \rho\big(\|B(z_{k+1}-z_\star)\|^2 - \|B(z_k - z_\star)\|^2\big)\end{aligned}$$

Therefore,

$$\begin{aligned}&\frac{1}{\rho}\big(\|y_{k+1}-y_\star\|^2 - \|y_k - y_\star\|^2\big) + \rho\big(\|B(z_{k+1}-z_\star)\|^2 - \|B(z_k - z_\star)\|^2\big)\\&+ \rho\|r_{k+1} - B(z_{k+1}-z_k)\|^2 \le 0,\end{aligned}$$

i.e.,

$$V_{k+1} \le V_k - \rho\|r_{k+1} - B(z_{k+1}-z_k)\|^2 \overset{(*)}{\le} V_k - \rho\|r_{k+1}\|_2^2 - \rho\|B(z_{k+1}-z_k)\|_2^2.$$

It remains to show the final inequality (*), i.e., whether

$$\begin{aligned}&-\rho\|r_{k+1} - B(z_{k+1}-z_k)\|^2 + \rho\|r_{k+1}\|_2^2 + \rho\|B(z_{k+1}-z_k)\|_2^2\\&= 2\rho\langle r_{k+1}, B(z_{k+1}-z_k)\rangle = 2\langle y_{k+1}-y_k, B(z_{k+1}-z_k)\rangle \overset{(*)}{\le} 0.\end{aligned}$$

Recall,

$$z_{k+1} = \operatorname*{argmin}_z \left\{ g(z) + \langle y_{k+1}, Bz\rangle + \frac{\rho}{2}\|B(z - z_{k+1})\|^2 \right\}$$

So, by the lemma, we have

$$g(z_{k+1}) + \langle y_{k+1}, Bz_{k+1}\rangle \le g(z_k) + \langle y_{k+1}, Bz_k\rangle$$

Likewise,

$$z_k = \operatorname*{argmin}_z \left\{ g(z) + \langle y_k, Bz\rangle + \frac{\rho}{2}\|B(z - z_k)\|^2 \right\}$$

So, by the lemma, we have

$$g(z_k) + \langle y_k, Bz_k\rangle \le g(z_{k+1}) + \langle y_k, Bz_{k+1}\rangle$$

Adding the two inequalities gives us

$$\langle y_{k+1}, Bz_{k+1}\rangle + \langle y_k, Bz_k\rangle - \langle y_{k+1}, Bz_k\rangle - \langle y_k, Bz_{k+1}\rangle \le 0$$

which can be reorganized to

$$\langle y_{k+1} - y_k, B(z_{k+1}-z_k)\rangle \le 0$$

This proves (*) and completes the proof. □