1 Outline

In this lecture, we study

- conic duality,
- optimality conditions for convex minimization,
- Normal cones and projection.

2 Conic duality theorem

In the last lecture, we learned that a *conic program* is an optimization problem is defined with a pointed and closed convex cone K with a nonempty interior:

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax - b \in K. \end{aligned} \tag{CP}$$

Recall that when $K = \mathbb{R}^n_+$, (CP) reduces to linear program

$$\begin{aligned} \text{minimize} \quad & c^\top x \\ \text{subject to} \quad & Ax \geq b. \end{aligned} \tag{LP}$$

The *dual cone* of $K \subseteq \mathbb{R}^n$ is defined as

$$K^* = \left\{ y \in \mathbb{R}^n : \ y^\top x \geq 0 \ \ \forall x \in K \right\}.$$

The dual cone of the nonnegative orthant $\mathbb{R}^d_+$ is $\mathbb{R}^d_+$ itself.

Example 7.1. The dual cone of the positive semidefinite cone $\mathbb{S}^d_+$ is given by

$$\left\{ X \in \mathbb{R}^{d \times d} : \ \mathrm{tr}(X^\top S) = \sum_{i=1}^d \sum_{j=1}^d X_{ij} S_{ij} \geq 0 \ \ \forall S \in \mathbb{S}^d_+ \right\}.$$

In fact, the positive semidefinite cone $\mathbb{S}^d_+$ is *self-dual*, meaning that its dual cone is itself.

Example 7.2. The dual cone of the Lorentz cone (the ice-cream cone) is also self-dual.

Theorem 7.3 (See Theorem 2.3.1 in [BTN01])**.** *Let K be a pointed and closed convex cone with nonempty interior. Then its dual cone K^* is also a pointed and closed convex cone with nonempty interior. Moreover, $(K^*)^* = K$.*

Then the dual of (CP) is defined and can be derived by the following procedure.

(1) Take x such that $Ax - b \in K$ and $y \in K^*$. Then $y^\top (Ax - b) \geq 0$, and therefore,

$$y^\top Ax \geq y^\top b.$$

(2) If $y \in K^*$ further satisfies $A^\top y = c$, then

$$c^\top x = y^\top Ax \geq y^\top b = b^\top y.$$

(3) Then

$$\begin{aligned} \text{maximize} \quad & b^\top y \\ \text{subject to} \quad & A^\top y = c \\ & y \in K^* \end{aligned} \tag{dual-CP}$$

provides a lower bound on the value of (CP). Here, (dual-CP) is the dual conic program of (CP).

Example 7.4. The dual of the linear program (LP) is given by

$$\begin{aligned} \text{maximize} \quad & b^\top y \\ \text{subject to} \quad & A^\top y = c \\ & y \geq 0. \end{aligned} \tag{dual-LP}$$

Remark 7.5. Here, what is the intuition behind (dual-CP)? What we argued while deriving (dual-CP) is the following. Take a feasible solution x to (CP). Then we have $Ax - b \in K$. Now suppose that we have $y \in K^*$, where K^* is the dual cone of K, such that $A^\top y = c$. Then we saw that

$$c^\top x \geq b^\top y.$$

Note that this holds for any pair of (x, y) satisfying the condition. In particular, what this implies that $b^\top y$ provides a lower bound on the optimum value of (CP), as we may minimize $c^\top x$ over all x satisfying $Ax - b \in K$.

That being said, then what does (dual-CP) do? Well, we saw that for any y with $A^\top y = c$ and $y \in K^*$, $b^\top y$ provides a lower bound. What is the strongest possible lower bound obtained that way? (dual-CP) computes it.

Example 7.6. We consider the following semidefinite program.

$$\begin{aligned} \text{maximize} \quad & \sum_{\ell=1}^m b_\ell y_\ell \\ \text{subject to} \quad & \sum_{\ell=1}^m y_\ell A_\ell \preceq C \end{aligned}$$

To obtain its dual, we take a positive semidefinite matrix X. As the positive semidefinite cone $\mathbb{S}_+^d$ is self-dual, it follows that

$$\operatorname{tr}\left(X^\top \left(C - \sum_{\ell=1}^m y_\ell A_\ell\right)\right) = \operatorname{tr}(C^\top X) - \sum_{\ell=1}^m y_\ell \cdot \operatorname{tr}((A_\ell)^\top X) \geq 0.$$

If X satisfies

$$\operatorname{tr}((A_\ell)^\top X) = b_\ell \quad \text{for } \ell = 1, \ldots, m,$$

then

$$\mathrm{tr}(C^\top X) \geq \sum_{\ell=1}^m y_\ell \cdot \mathrm{tr}((A_\ell)^\top X) = \sum_{\ell=1}^m b_\ell y_\ell.$$

This means that

$$\begin{aligned} \text{minimize} \quad & \mathrm{tr}(C^\top X) \\ \text{subject to} \quad & \mathrm{tr}((A_\ell)^\top X) = b_\ell \quad \text{for } \ell = 1, \ldots, m \\ & X \succeq 0 \end{aligned}$$

provides an upper bound on the first semidefinite program.

Example 7.7. Let us consider the following second-order cone program:

$$\begin{aligned} \text{minimize} \quad & f^\top x \\ \text{subject to} \quad & \|Ax - b\|_2 \leq c^\top x - d \end{aligned}$$

where A has m rows and b has m components. To obtain its dual, we take $(y, t) \in \mathbb{R}^m \times \mathbb{R}$ such that

$$t \geq \|y\|_2.$$

As the Lorentz cone is self-dual, we obtain

$$t(c^\top x - d) + y^\top (Ax - b) \geq 0.$$

Then we deduce that

$$(A^\top y + tc)^\top x \geq b^\top y + td.$$

Based on this, we take the following optimization model:

$$\begin{aligned} \text{maximize} \quad & b^\top y + td \\ \text{subject to} \quad & A^\top y + tc = f \\ & t \geq \|y\|_2. \end{aligned}$$

We have shown that the value of the conic program (CP) is lower bounded by the value of its dual, given by (dual-CP). What is striking is that the two values coincide under some conditions! For the case of linear programming, this is known as the strong LP duality theorem. To be precise, the value of (LP) and that of (dual-LP) are the same if (LP) is feasible and bounded. It turns out that we may extend this strong duality result to general conic programs.

We have already observed that the weak duality for LP extends to conic programming. Formally,

Theorem 7.8. *The optimal value of* (dual-CP) *is a lower bound on that of* (CP)*.*

Before we state the strong duality theorem for conic programming, we need to define the notion of *strict feasibility.*

Definition 7.9. We say that a solution x is *strictly feasible* to (CP) if $Ax - b$ belongs to the interior of K. When that is the case, we write it as

$$Ax - b >_K 0 \quad \text{or} \quad Ax >_K 0.$$

Moreover, we say that (CP) is *strictly feasible* if it has a strictly feasible solution.

For example, x is strictly feasible to (LP) if $Ax > b$. Now we are ready to state the following duality result.

Theorem 7.10 (See Theorem 2.4.1 in [BTN01])**.** *The following statements hold for* (CP) *and* (dual-CP)*.*

1. *If* (CP) *is strictly feasible and bounded, then* (dual-CP) *is solvable and the optimal values of* (CP) *and* (dual-CP) *are the same.*
2. *If* (dual-CP) *is strictly feasible and bounded, then* (CP) *is solvable and the optimal values of* (CP) *and* (dual-CP) *are the same.*

We will learn the notion of *Lagrangian duality* later in the course, and the strict feasibility condition is analogous to the *Slater condition*.

3 Optimality conditions for convex minimization

3.1 Local optimality implies global optimality

A feasible solution x^* is *locally optimal* to the optimization problem

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & x \in C \end{array} \tag{P}$$

if there exists $R > 0$ such that

$$f(x^*) = \min\left\{f(x): \; x \in C, \; \|x - x^*\| \le R\right\}.$$

Theorem 7.11. *Any locally optimal solution to a convex optimization problem is (globally) optimal.*

Proof. Suppose for a contradiction that a locally optimal solution x^* to a convex optimization problem $\min_{x \in C} f(x)$ is not globally optimal. Then there exists $y \in C$ such that $f(y) < f(x^*)$. By the local optimality of x^*, there exists $R > 0$ such that $f(x^*) = \min\{f(x) : x \in C, \; \|x - x^*\| \le R\}$, which implies that $\|y - x^*\| > R$. Let z be defined as

$$z = x^* + \frac{R}{\|y - x^*\|}(y - x^*) = \left(1 - \frac{R}{\|y - x^*\|}\right) x^* + \frac{R}{\|y - x^*\|} y.$$

Since z is a convex combination of x^* and y, it follows that $z \in C$ and

$$f(z) \le \frac{R}{\|y - x^*\|} f(y) + \left(1 - \frac{R}{\|y - x^*\|}\right) f(x^*) < f(x^*)$$

However, we have $\|z - x^*\| = R$, contradicting the assumption that $f(x^*) = \min\{f(x) : x \in C, \; \|x - x^*\| \le R\}$. $\square$

For nonconvex problems, a locally optimal solution is not necessarily an optimal solution, illustrated in Figure 7.1.

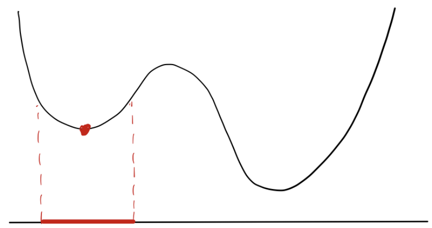

Figure 7.1: Local optimal solution that is not optimal

3.2 First-order optimality condition

Next we establish an optimality condition for convex optimization problems with a differentiable objective.

Theorem 7.12. *For a convex optimization problem of the form* (P) *with f differentiable, $x^* \in C$ is an optimal solution if and only if*

$$\nabla f(x^*)^\top (x - x^*) \geq 0 \quad \textit{for all } x \in C.$$

We will prove this later in the course, when we discuss the general case allowing nondifferentiable objectives. Figure 7.2 describes the optimality conditions for functions from $\mathbb{R}^2$ to $\mathbb{R}$. Basically, a solution x^* is optimal if we cannot move further from x^* in C in the direction of decreasing f. If $\nabla f(x^*) = 0$, then x^* is optimal.

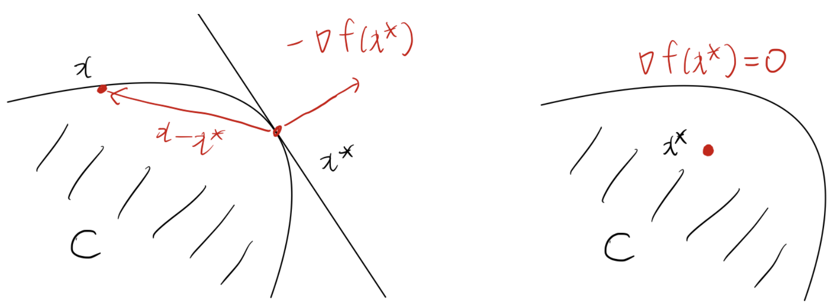

Figure 7.2: Optimality of bi-variate convex functions

By Theorem 7.12, a sufficient condition for optimality is that $\nabla f(x^*) = 0$. This, in fact, is a necessary and sufficient condition for the unconstrained case.

Theorem 7.13. *$x^* \in \mathbb{R}^d$ is optimal to $\min_{x \in \mathbb{R}^d} f(x)$ if and only if*

$$\nabla f(x^*) = 0.$$

Proof. ($\Leftarrow$) If $\nabla f(x^*) = 0$, then it trivially holds that $\nabla f(x^*)^\top (x - x^*) \geq 0$ for $x \in \mathbb{R}^d$. Then x^* is optimal due to Theorem 7.12.

($\Rightarrow$) Let $x = x^* - \alpha \nabla f(x^*)$. Then by Theorem 7.12, we have

$$\nabla f(x^*)^\top (x - x^*) = -\alpha \|\nabla f(x^*)\|_2^2 \geq 0.$$

This in turn implies that $\|\nabla f(x^*)\|_2 = 0$ and thus $\nabla f(x^*) = 0$. $\square$

Figure 7.3 describes the optimality conditions for functions from $\mathbb{R}$ to $\mathbb{R}$.

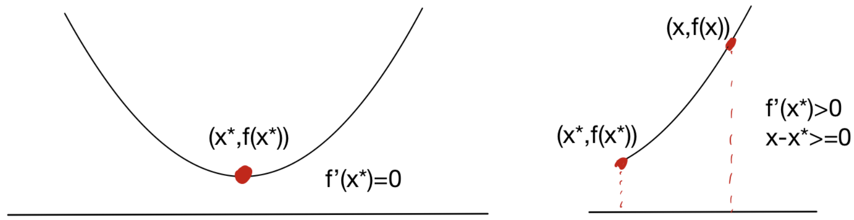

Figure 7.3: Optimality of univariate convex functions

Example 7.14. Consider the following equality-constrained problem.

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b \end{array}$$

where f is convex and A, b are matrices of appropriate dimensions. Then a solution x^* is optimal if and only if $\nabla f(x^*)^\top (x - x^*) \geq 0$ for all x such that $Ax = b$. Note that the latter condition is equivalent to $\nabla f(x^*)^\top v = 0$ for all v in the null space of A. Since the orthogonal complement of $\text{null}(A)$ is the column space of $A^\top$, we have $\nabla f(x^*) = A^\top u$ for some u.

The *normal cone* of C at $x \in C$ is defined as

$$N_C(x) = \{g \in \mathbb{R}^d : \ g^\top (y - x) \leq 0 \text{ for all } y \in C\}.$$

Figure 7.4 shows some examples.

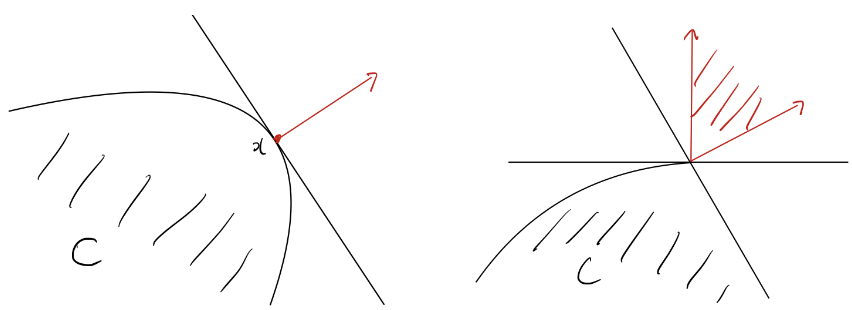

Figure 7.4: Optimality of univariate convex functions

Then the optimality condition in Theorem 7.12 is equivalent to

$$-\nabla f(x^*) \in N_C(x^*) \quad \leftrightarrow \quad 0 \in \nabla f(x^*) + N_C(x^*).$$

Later in the course, we will give a direct proof for this equivalent condition.

3.3 Projection

We consider the problem of projecting a point p onto a convex set C, that is to find a point $x \in C$ minimizing the distance to p.

$$\begin{array}{ll} \text{minimize} & \|x - p\|_2^2 \\ \text{subject to} & x \in C \end{array}$$

Let $\mathrm{Proj}_C(p)$ denote the projection of p to C[1]. By definition, $\mathrm{Proj}_C(p)$ is an optimal solution to the optimization problem. Note that the gradient of $\|x-p\|_2^2$ is

$$2(x-p).$$

As $\mathrm{Proj}_C(p)$ is the optimal solution to the above optimization problem, it follows from Theorem 7.12 that

$$2(\mathrm{Proj}_C(p)-p)^\top(x-\mathrm{Proj}_C(p)) \geq 0 \quad \text{for all } x \in C.$$

Equivalently,

$$\langle \mathrm{Proj}_C(p)-p,\ \mathrm{Proj}_C(p)-x\rangle \leq 0 \quad \text{for all } x \in C.$$

Next let us consider two points u, v and their projections onto C, given by $\mathrm{Proj}_C(u)$ and $\mathrm{Proj}_C(v)$, respectively. Then we have

$$\begin{aligned}
\langle \mathrm{Proj}_C(u)-u,\ \mathrm{Proj}_C(u)-\mathrm{Proj}_C(v)\rangle \leq 0,\\
\langle \mathrm{Proj}_C(v)-v,\ \mathrm{Proj}_C(v)-\mathrm{Proj}_C(u)\rangle \leq 0.
\end{aligned}$$

Adding these two inequalities, we obtain

$$\|\mathrm{Proj}_C(u)-\mathrm{Proj}_C(v)\|_2^2 - \langle u-v,\ \mathrm{Proj}_C(u)-\mathrm{Proj}_C(v)\rangle \leq 0.$$

Then it follows from the Cauchy-Schwarz inequality that

$$\|\mathrm{Proj}_C(u)-\mathrm{Proj}_C(v)\|_2 \leq \|u-v\|_2.$$

References

[BTN01] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on Modern Convex Optimization.* Society for Industrial and Applied Mathematics, 2001. 7.3, 7.10

[1] In fact, there exists a unique optimal solution to the above optimization problem. Why?