Class 5: Separation theorems and strong duality

In this class, we still consider the following primal problem:

$$\begin{array}{ll} \min & f(x) \\ \text{s.t.} & x \in X, \end{array} \tag{NLP}$$

where

$$X := \{x \in \mathbb{R}^n \mid g(x) \leq 0, h(x) = 0\}, \tag{1}$$

with $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^m$ and $h\colon \mathbb{R}^n \to \mathbb{R}^p$. Moreover, its dual problem is given by

$$\begin{array}{ll} \max & \omega(\lambda, \mu) \\ \text{s.t.} & (\lambda, \mu) \in D, \end{array} \tag{2}$$

where

$$D := \left\{(\lambda, \mu) \in \mathbb{R}^m_+ \times \mathbb{R}^p \mid \omega(\lambda, \mu) > -\infty\right\}.$$

and $\omega\colon D \to \mathbb{R}$ is defined by

$$\omega(\lambda, \mu) := \inf_{x \in \mathbb{R}^n} L(x, \lambda, \mu).$$

We already know that the weak duality theorem holds for general problems of the type (NLP). Now, we will prove that the strong duality is satisfied when the problem is convex, under the Slater CQ. Before that, we need some other tools. Let $B(x, r)$ be the ball centered in $x \in \mathbb{R}^n$ with radius $r > 0$. Recall first that for a set $S \subseteq \mathbb{R}^n$ and a point $x \in \mathbb{R}^n$, if there exists $r > 0$ such that $B(x, r) \subseteq S$, then x is in the interior of S. We denote by $\operatorname{int} S$ the set of all points that is in the interior of S. Moreover, $\operatorname{cl} S$ denotes the closure of S, that is, the minimum closed set that contains S. Furthermore, $\operatorname{bd} S := \operatorname{cl} S \setminus \operatorname{int} S$ is called boundary of S.

Theorem 1. *Let $S \subset \mathbb{R}^n$ be a nonempty, convex and closed set. Then, $\bar{x} \in \mathbb{R}^n$ is the unique minimizer of problem*

$$\begin{array}{ll} \min & \|y - x\| \\ \text{s.t.} & y \in S \end{array}$$

if and only if

$$\langle x - \bar{x}, y - \bar{x} \rangle \leq 0 \quad \text{for all } y \in S.$$

This point $\bar{x}$ is called projection *of x onto S.*

In the following lemma, for a convex set S, we show that a point that is not in the closure of S can be strictly separated from S.

Lemma 1. (Minkowski's lemma) *Let $S \subset \mathbb{R}^n$ be a nonempty convex set and $x \notin \operatorname{cl} S$. Then, there exists $a \in \mathbb{R}^n \setminus \{0\}$ such that*

$$\langle a, x \rangle < \langle a, y \rangle \quad \text{for all } y \in S.$$

Proof. Since S is nonempty and convex, $\mathrm{cl}\, S$ is nonempty, convex and closed. From Theorem 1, there exists a unique $\bar{x} \in \mathbb{R}^n$ that is the projection of x onto $\mathrm{cl}\, S$. Once again from Theorem 1, we have

$$\langle \bar{x} - x, y - \bar{x} \rangle \geq 0 \quad \text{for all } y \in \mathrm{cl}\, S.$$

Define $a := \bar{x} - x$, which is not zero because $x \notin \mathrm{cl}\, S$. From the above inequality, for all $y \in S$, we obtain

$$\begin{aligned} \langle a, y \rangle &\geq \langle \bar{x} - x, \bar{x} \rangle \\ &= \|\bar{x}\|^2 - \langle x, \bar{x} \rangle \\ &= \|x - \bar{x}\|^2 + \langle \bar{x} - x, x \rangle \\ &> \langle a, x \rangle, \end{aligned}$$

where the last inequality holds because $a \neq 0$, and the result follows. □

The next result is similar to Minkowski's lemma, but the point to consider is in the boundary of the set. In this case, we can also "separate" that point from the set, but the separation is not strict.

Theorem 2. *Let $S \subset \mathbb{R}^n$ be a nonempty convex set and $x \in \mathrm{bd}\, S$. Then there exist $a \in \mathbb{R}^n \setminus \{0\}$ such that*

$$\langle a, x \rangle \leq \langle a, y \rangle \quad \text{for all } y \in S.$$

Proof. Since $x \in \mathrm{bd}\, S$, there exists a sequence $\{x^k\}$ such that $x^k \to x$ and $x^k \notin \mathrm{cl}\, S$ for all k. From Lemma 1, for all k, there exists $a^k \in \mathbb{R}^n \setminus \{0\}$ such that

$$\langle a^k, x \rangle < \langle a^k, y \rangle \quad \text{for all } y \in S.$$

Without loss of generality, assume that $\{a^k / \|a^k\|\} \to a$ for some $a \in \mathbb{R}^n \setminus \{0\}$. Dividing the above inequality by $\|a^k\|$ and taking the limit $k \to \infty$, gives the desired result. □

The next lemma will be used in the proof of the strong duality.

Lemma 2. *For problem (NLP), assume that f and g are convex and h is affine, with $h(x) := Ax - b$ for some $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$. Then, the following set is nonempty and convex:*

$$U := \left\{ (w, y, z) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R} \;\middle|\; \exists\, x \in \mathbb{R}^n \text{ s.t. } Ax - b = w,\ g(x) \leq y,\ f(x) \leq z \right\}. \qquad (3)$$

Proof. Since $(Ax - b, g(x), f(x)) \in U$ for any $x \in \mathbb{R}^n$, we have $U \neq \emptyset$. Now, let $(w^1, y^1, z^1) \in U$, $(w^2, y^2, z^2) \in U$ and $\alpha \in [0, 1]$. Then, there exist x^1 and x^2 such that $Ax^1 - b = w^1$, $g(x^1) \leq y^1$, $f(x^1) \leq z^1$, $Ax^2 - b = w^2$, $g(x^2) \leq y^2$ and $f(x^2) \leq z^2$. Define $x := \alpha x^1 + (1 - \alpha) x^2$. Then, we have

$$\begin{aligned} Ax - b &= A(\alpha x^1 + (1 - \alpha) x^2) - b \\ &= \alpha (Ax^1 - b) + (1 - \alpha)(Ax^2 - b) \\ &= \alpha w^1 + (1 - \alpha) w^2. \end{aligned}$$

Moreover, since f is convex, we obtain

$$\begin{aligned} f(x) &\leq \alpha f(x^1) + (1-\alpha) f(x^2) \\ &\leq \alpha z^1 + (1-\alpha) z^2. \end{aligned}$$

Finally, for all i, from the convexity of g_i, we get

$$\begin{aligned} g_i(x) &\leq \alpha g_i(x^1) + (1-\alpha) g_i(x^2) \\ &\leq \alpha y^1 + (1-\alpha) y^2. \end{aligned}$$

The above relations show that $\alpha(w^1, y^1, z^1) + (1-\alpha)(w^2, y^2, z^2) \in U$, which means that U is convex. $\square$

Theorem 3. (Strong duality) *For problem (NLP), assume that f and g are convex, and h is affine. Suppose also that the Slater CQ holds for (NLP), i.e., there exists $\hat{x} \in \mathbb{R}^n$ such that $h(\hat{x}) = 0$ and $g_i(\hat{x}) < 0$ for all $i = 1, \dots, m$. If the optimal value of the primal problem is finite, i.e., $\inf_{x \in X} f(x) > -\infty$, then the dual problem (2) has an optimal solution and the duality gap is zero.*

Proof. Since h is affine, we can write $h(x) = Ax - b$ for some $A \in \mathbb{R}^{p \times n}$ and $b \in \mathbb{R}^p$. We can assume, without loss of generality, that A has linearly independent rows, because otherwise we can remove the constraints without changing the feasible set, until all the rows become linearly independent.
We first observe that $(0, 0, \bar{f}) \in \operatorname{bd} U$, where $\bar{f} := \inf_{x \in X} f(x)$ and U is defined in (3). In fact, if $(0, 0, \bar{f}) \in \operatorname{int} U$, for some $\varepsilon > 0$, we have $(0, 0, \bar{f} - \varepsilon) \in \operatorname{int} U$. But this means that there exits some $x \in X$ such that $f(x) \leq \bar{f} - \varepsilon < \bar{f}$, which contradicts the definition of $\bar{f}$. Thus, $(0, 0, \bar{f}) \in \operatorname{bd} U$ holds. From Lemma 2, the set U is also convex. Thus, we can use Theorem 2, which says that there exist $(\bar{\lambda}, \bar{\mu}, \gamma) \in \mathbb{R}^p \times \mathbb{R}^m \times \mathbb{R} \setminus \{0\}$ such that

$$\gamma \bar{f} \leq \langle \bar{\lambda}, w \rangle + \langle \bar{\mu}, y \rangle + \gamma z \quad \text{for all } (w, y, z) \in U. \tag{4}$$

Now, let us first prove that $\bar{\mu}, \gamma \geq 0$.

(i) If $(w, y, z) \in U$, then clearly, $(w, y, z + t) \in U$ for all $t > 0$. Thus, from (4), we obtain

$$\gamma \bar{f} \leq \langle \bar{\lambda}, w \rangle + \langle \bar{\mu}, y \rangle + \gamma (z + t) \quad \text{for all } t > 0.$$

If $\gamma < 0$, then we can take t sufficiently large, so that the above inequality does not hold. Therefore, we must have $\gamma \geq 0$.

(ii) Assume that $\bar{\mu}_i < 0$ for some $i \in \{1, \dots, m\}$. If $(w, y, z) \in U$, then $(w, y(t), z) \in U$ for all $t > 0$, where

$$y_j(t) := \begin{cases} y_j + t, & \text{if } j = i, \\ y_j, & \text{if } j \in \{1, \dots, m\} \setminus \{i\}. \end{cases}$$

From (4), we have

$$\gamma \bar{f} \leq \langle \bar{\lambda}, w \rangle + \bar{\mu}_i (y_i + t) + \sum_{j \in \{1, \dots, m\} \setminus \{i\}} \bar{\mu}_j y_j + \gamma z \quad \text{for all } t > 0.$$

Once again, when t is sufficiently large, the above inequality does not hold. Thus, we must have $\bar{\mu} \geq 0$.

Now, assume that $\gamma = 0$. Using the fact that $((Ax - b), g(x), f(x)) \in U$ for any $x \in \mathbb{R}^n$, from (4), we obtain

$$\begin{aligned} 0 &\leq \langle \bar{\lambda}, Ax - b \rangle + \langle \bar{\mu}, g(x) \rangle \\ &= \langle \bar{\lambda}, A(x - \hat{x}) \rangle + \langle \bar{\mu}, g(x) \rangle, \end{aligned}$$

where the last equality holds because $h(\hat{x}) = A\hat{x} - b = 0$ (Slater CQ). In particular, when $x = \hat{x}$, we have $\langle \bar{\mu}, g(\hat{x}) \rangle \geq 0$. Since $\bar{\mu} \geq 0$ from the above discussion (ii), and $g(\hat{x}) < 0$ from Slater CQ, we conclude that $\bar{\mu} = 0$. Returning to the above inequality, we obtain $\langle A^\top \bar{\lambda}, x - \hat{x} \rangle = \langle \bar{\lambda}, A(x - \hat{x}) \rangle \geq 0$ for all x, which implies $A^\top \bar{\lambda} = 0$. Since the rows of A are linearly independent, we get $\bar{\lambda} = 0$. Because $(\bar{\lambda}, \bar{\mu}, \gamma) \neq 0$, we have a contradiction. Therefore, from discussion (i) above, we have $\gamma > 0$. Considering now (4) at $((Ax - b), g(x), f(x)) \in U$ and dividing such an inequality by $\gamma > 0$, we obtain

$$\bar{f} \leq \left\langle \frac{\bar{\lambda}}{\gamma}, Ax - b \right\rangle + \left\langle \frac{\bar{\mu}}{\gamma}, g(x) \right\rangle + f(x) \quad \text{for all } x \in \mathbb{R}^n.$$

The above inequality, together with the weak duality theorem [Class 4, Theorem 1], gives

$$\begin{aligned} \sup_{(\lambda,\mu) \in D} \omega(\lambda, \mu) \leq \bar{f} &\leq \inf_{x \in \mathbb{R}^n} \left[\left\langle \frac{\bar{\lambda}}{\gamma}, Ax - b \right\rangle + \left\langle \frac{\bar{\mu}}{\gamma}, g(x) \right\rangle + f(x) \right] \\ &= \inf_{x \in \mathbb{R}^n} L(x, \bar{\lambda}/\gamma, \bar{\mu}/\gamma) \\ &= \omega(\bar{\lambda}/\gamma, \bar{\mu}/\gamma) \\ &\leq \sup_{(\lambda,\mu) \in D} \omega(\lambda, \mu). \end{aligned}$$

We conclude that the above inequalities hold as equalities. Then, $(\bar{\lambda}/\gamma, \bar{\mu}/\gamma)$ is a solution of the dual problem and the duality gap is zero. □

Operations Research, Advanced (Graduate School of Informatics, Kyoto University)
1st part by Ellen H. Fukuda (e-mail: ellen(at)i.kyoto-u.ac.jp, where (at) = @)