10-725/36-725: Convex Optimization **Fall 2015**

Lecture 11: October 8

Lecturer: Ryan Tibshirani *Scribes: Tian Tong*

Note: *LaTeX template courtesy of UC Berkeley EECS dept.*

11.1 Primal and dual problems

11.1.1 Lagrangian

Consider a general optimization problem (called as primal problem)

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \le 0, i = 1, \cdots, m \\ & \ell_j(x) = 0, j = 1, \cdots, r. \end{aligned} \tag{11.1}$$

We define its Lagrangian as

$$L(x, u, v) = f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j \ell_j(x).$$

Lagrange multipliers $u \in \mathbb{R}^m, v \in \mathbb{R}^r$.

Lemma 11.1 *At each feasible x, $f(x) = \sup_{u \ge 0, v} L(x, u, v)$, and the supremum is taken iff $u \ge 0$ satisfying $u_i h_i(x) = 0, i = 1, \cdots, m$.*

Proof: At each feasible x, we have $h_i(x) \le 0$ and $\ell(x) = 0$, thus $L(x, u, v) = f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j \ell_j(x) \le f(x)$. The last inequality becomes equality iff $u_i h_i(x) = 0, i = 1, \cdots, m$. ■

Proposition 11.2 *The optimal value of the primal problem, named as $f^\star$, satisfies:*

$$f^\star = \inf_x \sup_{u \ge 0, v} L(x, u, v).$$

Proof: First considering feasible x (marked as $x \in C$), we have $f^\star = \inf_{x \in C} f(x) = \inf_{x \in C} \sup_{u \ge 0, v} L(x, u, v)$. Second considering non-feasible x, since $\sup_{u \ge 0, v} L(x, u, v) = \infty$ for any $x \notin C$, $\inf_{x \notin C} \sup_{u \ge 0, v} L(x, u, v) = \infty$. In total, $f^\star = \inf_x \sup_{u \ge 0, v} L(x, u, v)$. ■

11.1.2 Lagrange dual function

Given a Lagrangian, we define its Lagrange dual function as

$$g(u, v) = \inf_x L(x, u, v).$$

It is worth mentioning that the infimum here does not require x to be taken in the feasible set.

11.1.3 Lagrange dual problem

Given primal problem (11.1), we define its Lagrange dual problem as

$$\begin{aligned} \max_{u,v} \quad & g(u,v) \\ \text{subject to} \quad & u \geq 0. \end{aligned} \tag{11.2}$$

Proposition 11.3 *The optimal value of the dual problem, named as $g^\star$, satisfies:*

$$g^\star = \sup_{u\geq 0, v} \inf_x L(x,u,v).$$

Proof: From the definitions, we have $g^\star = \sup_{u\geq 0,v} g(u,v) = \sup_{u\geq 0,v} \inf_x L(x,u,v)$. ■

Although the primal problem is not required to be convex, the dual problem is always convex.

Proposition 11.4 *The dual problem is a convex optimization problem.*

Proof: By definition, $g(u,v) = \inf_x f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j \ell_j(x)$ can be viewed as pointwise infimum of affine functions of u and v, thus is concave. $u \geq 0$ is affine constraints. Hence dual problem is a concave maximization problem, which is a convex optimization problem. ■

11.2 Weak and strong duality

11.2.1 Weak duality

The Lagrangian dual problem yields a lower bound for the primal problem. It always holds true that $f^\star \geq g^\star$, called as weak duality.

Proof: We have stated that

$$f^\star = \inf_x \sup_{u\geq 0} L(x,u,v) \qquad\qquad g^\star = \sup_{u\geq 0,v} \inf_x L(x,u,v).$$

The minimax inequality shows that $f^\star \geq g^\star$. ■

We can interpret the weak duality as a mixed strategies game. Consider a game with two players J and R: if J chooses the primal variable x, while R chooses the dual variables $u \geq 0, v$, then J must pay R amount $L(x,u,v)$. J decides over the primal variables, and seeks to minimize $L(x,u,v)$; R decides over the dual variables u, v, and seeks to maximize his income $L(x,u,v)$. We assume that one of the players goes first, the game is played only once, and both players have full information on the other's choice, once their decision is made.

Under our assumptions, $f^\star$ is R's optimal income if J plays first, while $g^\star$ is his optimal income if R plays first. We claim that playing first is at a disadvantage. Therefore, we expect the income of R to be higher in the case when J plays first than if J plays second, i.e., $f^\star \geq g^\star$.

11.2.2 Strong duality

In some problems, we actually have $f^\star = g^\star$, which is called strong duality. In fact, for convex optimization problems, we nearly always have strong duality, only in addition to some slight conditions. A most common condition is the Slater's condition.

Theorem 11.5 (Slater's theorem) *If the primal is a convex problem, and there exists at least one strictly feasible $\tilde{x} \in \mathbb{R}^n$, satisfying the Slater's condition, meaning that*

$$\exists \tilde{x}, h_i(\tilde{x}) < 0, i = 1, \ldots, m, \ell_j(\tilde{x}) = 0, j = 1, \ldots r.$$

Then strong duality holds.

We will propose a geometric proof to the Slater's theorem in next section.

A conception having close relationship with strong duality is the duality gap: given primal feasible x and dual feasible u, v, the quantity

$$f(x) - g(u, v)$$

is called the duality gap. From the weak duality, we have $f(x) - g(u, v) \geq f^\star - g^\star \geq 0$. Furthermore, we declare a sufficient and necessary condition for duality gap equal to 0.

Proposition 11.6 *With $x, (u, v)$, the duality gap equals to 0, iff x is the primal optimal solution, (u, v) is the dual optimal solution, and the strong duality holds.*

Proof: From definitions and the weak duality, we have $f(x) \geq p^\star \geq g^\star \geq g(u, v)$. The duality gap equals to 0, iff the three inequalities become equalities, respectively, x is the primal optimal solution, (u, v) is the dual optimal solution, and the strong duality holds. ■

The KKT conditions can be induced from this proposition, which will be discussed in detail in next lecture.

11.2.3 Geometric interpretation of duality

This part is not included in the lecture, but I hope to propose a brief present, for it is a pretty neat theory and very beneficial for understanding the duality.

Given primal problem (11.1), we define its epigraph as

$$\mathcal{A} = \{(p, q, t) \mid \exists x \in \mathbb{R}^n : h_i(x) \leq p_i, i = 1, \cdots, m; \ell_j(x) = q_j, j = 1, \cdots, r; f(x) \leq t\}.$$

The geometric interpretation of several key values are listed as

- $f^\star = \inf\{t \mid (0, 0, t) \in \mathcal{A}\}$ is the lowest intersection of the the t-axis and $\mathcal{A}$;
- $g(u, v) = \inf\{(u, v, 1)^T (p, q, t) \mid (p, q, t) \in \mathcal{A}\}$ is the intersection of the t-axis and a supporting hyperplane to $\mathcal{A}$ with normal vector $(u, v, 1)$. This is sometimes referred to as a nonvertical supporting hyperplane, because the last component of the normal vector is nonzero (it is actually 1).
- $g^\star$ is the highest intersection of the t-axis and all nonvertical supporting hyperplane of $\mathcal{A}$. Notice that $u \geq 0$ holds true for each nonvertical supporting hyperplane of $\mathcal{A}$.

From the geometric interpretation of $f^\star$ and $g^\star$, we actually have an equivalent geometric statement of strong duality:

Proposition 11.7 *The strong duality holds, iff there exists a nonvertical supporting hyperplane of $\mathcal{A}$ passing through $(0, 0, f^\star)$.*

Proof: From weak duality $f^\star \geq g^\star$, the intersection of the t-axis and a nonvertical supporting hyperplane cannot exceed $(0, 0, f^\star)$. The strong duality holds, i.e, $f^\star = g^\star$, iff $(0, 0, f^\star)$ is just the highest intersection, meaning that there exists a nonvertical supporting hyperplane of $\mathcal{A}$ passing through $(0, 0, f^\star)$. ■

For a general non-convex optimization problem, $\mathcal{A}$ is usually non-convex, thus there may not exist a supporting hyperplane at $(0, 0, f^\star)$. We give an example where the strong duality does not hold.

Example: Consider a non-convex optimization problem

$$\begin{aligned} \min_x \quad & x^4 - 50x^2 + 100x \\ \text{subject to} \quad & x \geq -2.5. \end{aligned} \tag{11.3}$$

Its epigraph $\mathcal{A} = \{(p, t) \mid \exists x, -x - 2.5 \leq p; x^4 - 50x^2 + 100x \leq t\}$ is shown as the yellow region in Fig. 11.1, as well as the primal optimal value $f^\star$, as the lowest intersection of the t-axis and $\mathcal{A}$, and dual optimal value $g^\star$, as the highest intersection of the t-axis and all nonvertical supporting hyperplane of $\mathcal{A}$. In this case, there does not exist a supporting hyperplane of $\mathcal{A}$ passing through $(0, f^\star)$, thus the strong duality does not hold.

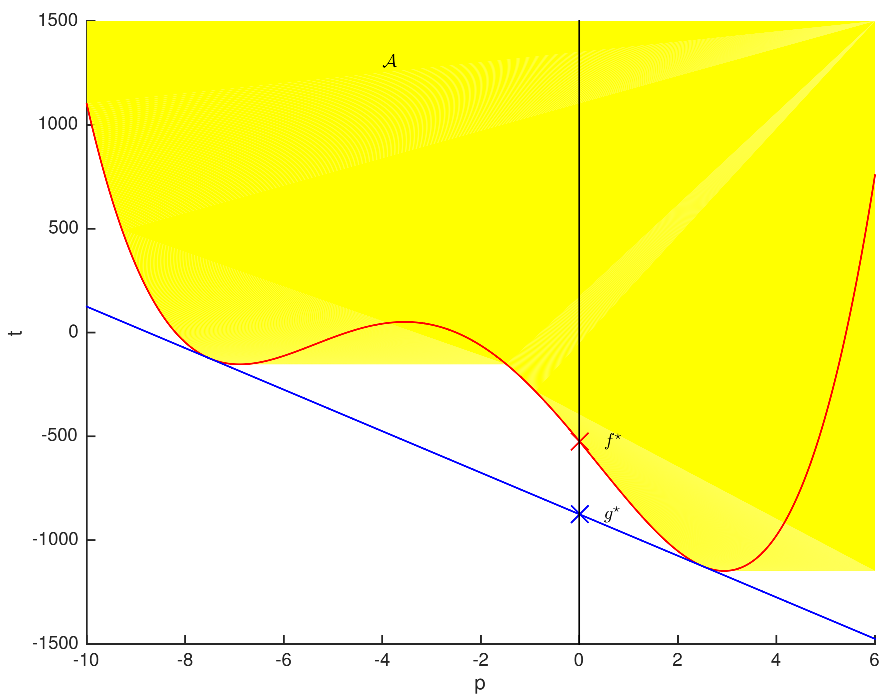

Figure 11.1: Illustration of a counterexample of strong duality

Different from general problems, if the optimization problem is convex, $\mathcal{A}$ is actually promised to be convex.

Proposition 11.8 *For a convex optimization problem, its epigraph $\mathcal{A}$ is a convex set. There must exist a supporting hyperplane of $\mathcal{A}$ passing through $(0, 0, f^\star)$.*

Proof: Take two points (p_1, q_1, t_1) and (p_2, q_2, t_2) in the epigraph $\mathcal{A}$: $\exists x_1$, s.t. $h(x_1) \le p_1, \ell(x_1) = q_1, f(x_1) \le t_1$; $\exists x_2$, s.t. $h(x_2) \le p_2, \ell(x_2) = q_2, f(x_2) \le t_2$. For any $\theta \in [0,1]$, we have

$$\begin{aligned}
h(\theta x_1 + (1-\theta)x_2) &\le \theta h(x_1) + (1-\theta)h(x_2) \le \theta p_1 + (1-\theta)p_2 \\
\ell(\theta x_1 + (1-\theta)x_2) &= \theta \ell(x_1) + (1-\theta)\ell(x_2) = \theta q_1 + (1-\theta)q_2 \\
f(\theta x_1 + (1-\theta)x_2) &\le \theta f(x_1) + (1-\theta)f(x_2) \le \theta t_1 + (1-\theta)t_2.
\end{aligned}$$

$\theta(p_1, q_1, t_1) + (1-\theta)(p_2, q_2, t_2) \in \mathcal{A}$, thus $\mathcal{A}$ is convex. From the hyperplane separation theorem, there exists a supporting hyperplane at every boundary points of a convex set. Since $(0, 0, f^\star)$ is on the boundary of $\mathcal{A}$, there exists a supporting hyperplane passing through it. ∎

Combining the Proposition 11.7 and 11.8, we derive a corollary:

Corollary 11.9 *For a convex optimization problem, the only case where strong duality does not hold is that the supporting hyperplane of $\mathcal{A}$ passing through $(0, 0, f^\star)$ is vertical.*

We propose an example of a convex optimization problem where the strong duality does not hold.

Example: Consider a convex optimization problem

$$\begin{aligned}
\min_{x,y} \quad & e^{-x} \\
\text{subject to} \quad & \frac{x^2}{y} I(y > 0) \le 0.
\end{aligned} \tag{11.4}$$

Here the indicator function $I(y > 0) = \begin{cases} 0, y > 0 \\ \infty, y \le 0 \end{cases}$. The epigraph for this problem is $\mathcal{A} = \{(p, t) \mid \exists x, y > 0, x^2/y \le p; e^{-x} \le t\} = \mathbb{R}^2_{++} \bigcup (\{0\} \times [1, \infty])$, which can be checked as a convex set. The primal optimal value $f^\star$, as the lowest intersection of the t-axis and $\mathcal{A}$, is 1; the dual optimal value $g^\star$, as the highest intersection of the t-axis and all nonvertical supporting hyperplane of $\mathcal{A}$, is 0. In this case, there only exists a vertical supporting hyperplane of $\mathcal{A}$ passing through $(0, 1)$, thus the strong duality does not hold.

At the end of the section, we propose a geometric proof to the Slater's theory. We shall prove that under the Slater's condition, the supporting hyperplane passing through $(0, 0, f^\star)$ must be nonvertical. It is sufficient to show that the left and right sides of each p-axis and q-axis are not empty, for in such case $\mathcal{A}$ cannot be entirely contained in one of the two closed half-spaces bounded by a vertical hyperplane.

Proof:[Slater's theorem] From the discussion above, we only need to show that the left and right sides of each p-axis and q-axis are not empty.

Slater's condition provides $h(\tilde{x}) < 0$, which directly shows that the left side of each p-axis is not empty. Since $\mathcal{A} = \mathcal{A} + (\mathbb{R}^m_+ \times \{0\}^r \times \mathbb{R}_+)$, the right side of any p in $\mathcal{A}$ is still contained in $\mathcal{A}$, thus the right side of each p-axis is not empty.

For the equality constraints $Ax = b$, we assume the rows of matrix A are linearly independent, which equivalently means A is full column rank, otherwise we shall first eliminate the dependent rows from our problem. Since Slater's condition provides a $\tilde{x}$ s.t. $A\tilde{x} = b$, we actually have $A(\tilde{x} + \mathbb{R}^n) - b = \mathbb{R}^r$. The epigraph $\mathcal{A}$ contains a point for every q, thus both sides of each q-axis are not empty. ∎

Based on this proof, we can also propose a weaker version of the Slater's condition.

Corollary 11.10 *The Slater's condition can be weaken to only requiring strict inequalities over functions h_i that are not affine.*

Proof: For each affine functions h_i, if the strict inequality is not satisfied, i.e., $h_i(\tilde{x}) = 0$, we can combine it with those linear equality constraints, and similarly prove that both sides of p_i-axis are not empty. ■

For a linear programming problem, since all inequality constraints are affine, we can deduct that strong duality holds if it is feasible. Applying the same logic to its dual problem, strong duality holds if the dual problem is feasible.

Corollary 11.11 *Strong duality holds for LPs, except when both primal and dual problems are infeasible, in which $f^\star = \infty$ and $g^\star = -\infty$.*

11.3 Applications

11.3.1 Dual of quadratic program

Consider a quadratic program with $Q \succeq 0$

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} \quad & Ax = b, x \geq 0. \end{aligned} \tag{11.5}$$

Its Lagrangian is

$$L(x,u,v) = \frac{1}{2}x^TQx + c^Tx - u^Tx + v^T(Ax - b).$$

The differential of Lagrangian is

$$dL(x,u,v) = (Qx + c - u + A^Tv)^Tdx.$$

To obtain a zero gradient, we must have $Qx + c - u + A^Tv = 0$, i.e., $c - u + A^Tv \in \mathrm{Col}(Q)$, in such case we have $x = -Q^\dagger(c - u + A^Tv)$, where $Q^\dagger$ is the generalized inverse of Q. The Lagrange dual function is

$$g(u,v) = \begin{cases} -\frac{1}{2}(c-u+A^Tv)^TQ^\dagger(c-u+A^Tv) - b^Tv & \text{if } c - u + A^Tv \in \mathrm{Col}(Q) \\ -\infty & \text{otherwise} \end{cases}$$

Its Lagrange dual problem is

$$\begin{aligned} \max_{u,v} \quad & -\frac{1}{2}(c-u+A^Tv)^TQ^\dagger(c-u+A^Tv) - b^Tv \\ \text{subject to} \quad & (I - QQ^\dagger)(c - u + A^Tv) = 0, u \geq 0. \end{aligned} \tag{11.6}$$

11.3.2 Dual of support vector machine

Give $y \in \{-1,1\}^n, X \in \mathbb{R}^{n\times p}$, rows of X as $x_1, \cdots, x_n$, the support vector machine problem is

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{subject to} \quad & \xi_i \geq 0, i = 1, \cdots, n \\ & y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i, i = 1, \cdots, n. \end{aligned} \tag{11.7}$$

Define $\tilde{X} = \text{diag}(y)X$, we can rewrite it in matrix forms

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C1^T\xi \\ \text{subject to} \quad & \xi \geq 0 \\ & \tilde{X}\beta + \beta_0 y \geq 1 - \xi. \end{aligned} \tag{11.8}$$

Introducing dual variables $v, w \geq 0$, we form the Lagrangian:

$$L(\beta, \beta_0, \xi, v, w) = \frac{1}{2}\|\beta\|_2^2 + C1^T\xi - v^T\xi + w^T(1 - \xi - \tilde{X}\beta - \beta_0 y).$$

Its differential is

$$dL(\beta, \beta_0, \xi, v, w) = (\beta - \tilde{X}^T w)^T d\beta - w^T y d\beta_0 + (C1 - v - w)^T d\xi$$

To obtain a zero gradient, we must have $\beta = \tilde{X}^T w$, $w^T y = 0$, and $w = C1 - v$, which gives the dual function as

$$g(v, w) = \begin{cases} -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w & \text{if } w = C1 - v, w^T y = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Its Lagrange dual problem is

$$\begin{aligned} \max_{v,w} \quad & -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w \\ \text{subject to} \quad & w = C1 - v, w^T y = 0 \\ & w \geq 0, v \geq 0. \end{aligned} \tag{11.9}$$

We can eliminate the slack variable v, resulting into

$$\begin{aligned} \max_{w} \quad & -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w \\ \text{subject to} \quad & 0 \leq w \leq C1, w^T y = 0. \end{aligned} \tag{11.10}$$