10-725/36-725: Convex Optimization **Fall 2016**

Lecture 13: Duality Uses and Correspondences

Lecturer: Ryan Tibshirani *Scribes: Yichong Xu, Yanyu Liang, Yuanning Li*

13.1 Last Class

13.1.1 KKT conditions

For the problem

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0, i = 1, ..., m \\ & l_j(x) = 0, j = 1, ..., r \end{aligned}$$

the KKT conditions are

- Stationary: $0 \in \partial f(x) + \sum_{i=1}^m u_i \partial h_i(x) + \sum_{j=1}^r v_j \partial l_j(x)$,
- Complementary slackness: $u_i \cdot h_i(x) = 0, \ \forall i$,
- Primal feasibility: $h_i(x) \leq 0, h_j(x) = 0, \ \forall i, j$,
- Dual feasibility: $u_i \geq 0, \ \forall i$.

The KKT conditions are always sufficient, and are necessary for optimality under strong duality.

13.1.2 Uses of duality

For a primal feasible x and a dual feasible u, v, $f(x) - g(u, v)$ is called the duality gap between x and u, v. Since $f(x) \geq g(u, v)$ we have

$$f(x) - f(x^*) \leq f(x) - g(u, v).$$

So a zero duality gap implies optimality. Also the duality gap can be used as a stopping criterion in algorithms.

Under strong duality, if we are given dual optimal u^*, v^*, any primal solution minimizes $L(x, u^*, v^*)$ over all x, because of the stationary condition. This can be used to characterize or compute primal solutions. Explicitly, given a dual solution u^*, v^*, any primal solution x^* solves

$$\min_x f(x) + \sum_{i=1}^m u_i^* \partial h_i(x) + \sum_{j=1}^r v_j^* \partial l_j(x).$$

Solutions of this unconstrained problem can often be expressed explicitly, giving an explicit characterization of primal solutions from dual solutions.
Example (B & V page 249). Consider the problem

$$\begin{aligned} \min_x \quad & \sum_{i=1}^n f_i(x_i) \\ \text{subject to} \quad & a^T x = b \end{aligned}$$

where each $f_i : \mathbb{R} \to \mathbb{R}$ is smooth and strictly convex. The dual function is

$$\begin{aligned} g(v) &= \min_x \sum_{i=1}^n f_i(x_i) + v(b - a^T x) \\ &= bv + \sum_{i=1}^n \min_{x_i} f_i(x_i) - a_i v x_i \\ &= bv - \sum_{i=1}^n f_i^*(a_i v), \end{aligned}$$

where f_i^* is the conjugate of f_i, which we will define later. The dual problem is thus

$$\max_v \ bv - \sum_{i=1}^n f_i^*(a_i v)$$

or

$$\min_v \ \sum_{i=1}^n f_i^*(a_i v) - bv.$$

This is a convex minimization problam with a scalar variable - it is much easier to solve than the primal problem. Given v^*, the primal solution x^* solves

$$\sum_{i=1}^n \min_{x_i} f_i(x_i) + a_i v x_i.$$

Since each f_i is strictly convex, the problem $\min_{x_i} f_i(x_i) + a_i v x_i$ has a unique solution, which can be computed by solving $\nabla f_i(x_i) = a_i v^*$ for each i.

13.2 Dual norms

Let $\|x\|$ be an arbitrary norm. For example:

- l_p norm: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, for $p \geq 1$.
- Trace norm: $\|X\|_{tr} = \sum_{i=1}^r \sigma_i(X)$.

Define its dual norm $\|x\|_*$ as

$$\|x\|_* = \max_{\|z\| \leq 1} z^T x.$$

The definition gives us the inequality $|z^T x| \leq \|z\| \|x\|^*$, similar to Cauchy-Schwartz inequality. For the examples:

- l_p norm dual: $(\|x\|_p)^* = \|x\|_q$, where $1/p + 1/q = 1$.
- Trace norm: $(\|X\|_{tr})_* = \|X\|_{op} = \sigma_1(X)$.

We can show that

Theorem 13.1 *The dual of dual norm is the original norm: i.e.,*

$$\|x\|_{**} = \|x\|.$$

Proof: Consider the problem

$$\min_y \|y\| \text{ subject to } y = x,$$

whose optimal value is $\|x\|$. Its Lagrangian is

$$L(y, u) = \|y\| + u^T(x - y) = \|y\| - y^T u + x^T u.$$

From the definition of $\|\cdot\|_*$, if $\|u\|_* > 1$, let z be the maximizer of $\max_{\|z\|\leq 1} z^T x$, i.e., $z^T x = \|u\|^*$. Note that $\|z\| = 1$. Let $y = tz$ for $t > 0$. Thus

$$\|y\| - y^T u = t(\|z\| - z^T u) = t(1 - \|u\|_*)$$

which $\to -\infty$ as $t \to \infty$. So in such case $\min_y \|y\| - y^T u = -\infty$.

If $\|u\|^* \leq 1$, we have $\|y\| - y^T u \geq \|y\| - \|y\|\|u\|_* \geq 0$. This can be realized by setting $y = 0$. So $\min_y \|y\| - y^T u = 0$.

Therefore the Lagrange dual problem is

$$\max_u u^T x \text{ subject to } \|u\|_* \leq 1,$$

whose optimal value is the dual of $\|\cdot\|_*$, i.e., $\|x\|_{**}$. By strong duality we have

$$\|x\|_{**} = \|x\|.$$

■

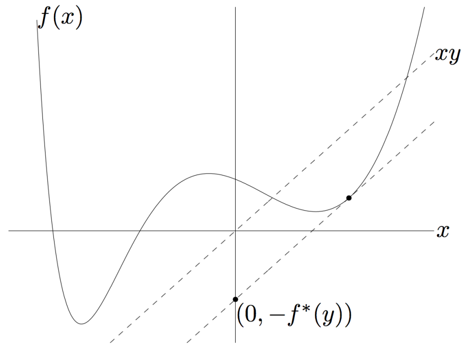
13.3 Conjugate function

13.3.1 Definition

Given a function $f : \mathbb{R}^n \to \mathbb{R}$, define its conjugate $f^* : \mathbb{R}^n \to \mathbb{R}$,

$$f^*(y) = \max_x y^T x - f(x)$$

Since $y^T x - f(x)$ is convex in y for any fixed x, f^* is always convex as it is a pointwise maximum of convex (affine) functions in y (f need not be convex). $f^*(y)$ is the maximum gap between linear function $y^T x$ and $f(x)$ and Figure 13.1 shows how it looks like when f is a scalar function. For differentiable f, conjugation is called the Legendre transform.



Given y, the upper dashed line represents the function $g(x) = yx$ and solid line represents $f(x)$. $f^*(y)$ is the biggest gap between g and f where g is above f. The lower dashed line is drawn to find such biggest gap and the absolute value of intercept corresponds to the value of biggest gap

Figure 13.1: From [1] pp. 91

13.3.2 Properties

- Fenchel's inequality: for any x, y,

$$f(x) + f^*(y) \geq x^T y$$

Proof: $f^*(y) = \max_z z^T y - f(z) \geq x^T y - f(x)$ ■

- Hence conjugate of conjugate f^{**} satisfies $f^{**} \leq f$.
Proof: $f^{**}(x) = \max_z z^T x - f^*(z) \leq \max_z f(x) = f(x)$ ('$\leq$' comes from Fenchel's inequality) ■

- If f is closed and convex, then $f^{**} = f$
Proof: $f(x)$ shares the same value with $\min_y f(y)$, subject to $y = x$. The dual is $\max_u u^T x - f^*(u) = f^{**}(x)$. Since strong duality holds, the equality follows. ■

- If f is closed and convex, then for any x, y,

$$\begin{aligned} x \in \partial f^*(y) &\iff y \in \partial f(x) \\ &\iff f(x) + f^*(y) = x^T y \end{aligned}$$

- If $f(u, v) = f_1(u) + f_2(v)$, then

$$f^*(w, z) = f_1^*(w) + f_2^*(z)$$

Proof: $f^*(w, z) = \max_{u,v} (u^T, v^T)(w, z)^T - f(u, v) = \max_{u,v} u^T w + v^T z - f_1(u) - f_2(v) = \max_u \{u^T w - f_1(u)\} + \max_v \{v^T z - f_2(v)\} = f_1^*(w) + f_2^*(z)$ ■

13.3.3 Examples

- Simple quadratic: let $f(x) = \frac{1}{2} x^T Q x$, where $Q \succ 0$. Then $y^T x - \frac{1}{2} x^T Q x$ is strictly concave in y and is maximized at $y = Q^{-1} x$, so

$$f^*(y) = \frac{1}{2} y^T Q^{-1} y$$

- Indicator function: if $f(x) = I_C(x)$, then its conjugate is

$$f^*(y) = \max_{x \in C} y^T x := I_C^*(y)$$

called the support function of C.

- Norm: if $f(x) = \|x\|$, then its conjugate is

$$f^*(y) = I_{\{z:\|z\|_* \leq 1\}}(y)$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

13.3.4 Example: lasso dual

Given $y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}$, recall the lasso problem:

$$\min_\beta \frac{1}{2}\|y - X\beta\|_2^2 + \lambda\|\beta\|_1$$

Its dual function is just a constant (equal to $f^\star$). Therefore we transform the primal to

$$\begin{aligned} \min_{\beta,z} & \frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1 \\ \text{subject to} \quad & z = X\beta \end{aligned}$$

So dual function is now

$$\begin{aligned} g(u) &= \min_{\beta,z} \frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1 + u^T(z - X\beta) \\ &= -\frac{1}{2}\|u\|_2^2 + y^T u - I_{\{v:\|v\|_\infty \leq 1\}}(X^T u/\lambda) \end{aligned}$$

Therefore the lasso dual problem is

$$\begin{aligned} \max_u & -\frac{1}{2}\|u\|_2^2 + y^T u \\ \text{subject to} \quad & \|X^T u\|_\infty \leq \lambda \end{aligned}$$

or equivalently

$$\begin{aligned} \min_u & \|y - u\|_2^2 \\ \text{subject to} \quad & \|X^T u\|_\infty \leq \lambda \end{aligned}$$

Check: Slater's condition holds, and hence so does strong duality. But note: the optimal value of the last problem is not the optimal lasso objective value.

Further, note the given the dual solution u, any lasso solution β satisfies

$$X\beta = y - u$$

This is from KKT stationarity condition for z (i.e. $z - y + \beta = 0$). So the lasso fit is just the dual residual (see Figure 13.2).

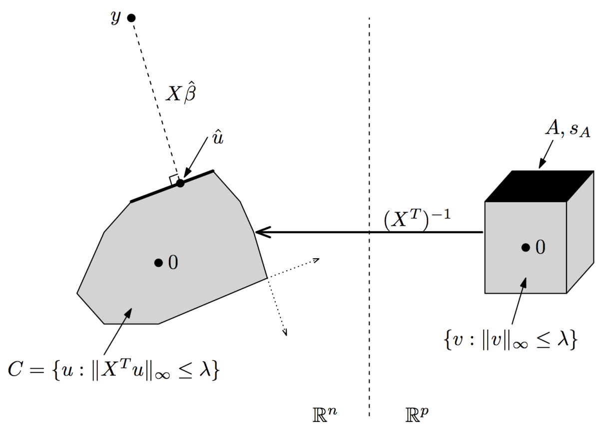


Figure 13.2: The lasso solution and its dual solution

13.3.5 Conjugates and dual problems

Conjugates appear frequently in derivation of dual problems, via

$$-f^*(u) = \min_x f(x) - u^T x$$

in minimization of the Lagrangian. E.g., consider

$$\min_x f(x) + g(x)$$

Equivalently: $\min_{x,z} f(x) + g(z)$ subject to $x = z$. Lagrange dual function is:

$$\begin{aligned}
g(u) &= \min_{x,z} f(x) + g(z) + u^T(z - x) \\
&= \min_{x,z} f(x) - u^T x + g(z) - (-u)^T z \\
&= \min_x \{f(x) - u^T x\} + \min_z \{g(z) - (-u)^T z\} \\
&= -\max\{u^T x - f(x)\} - \max_z \{(-u)^T z - g(z)\} \\
&= -f^*(u) - g^*(-u)
\end{aligned}$$

Examples of this last calculation:

- Indicator function: the dual of

$$\min_x f(x) + I_C(x)$$

  is

$$\max_u -f^*(u) - I_C^*(-u)$$

  where I_C^* is the support function of C.

- Norms: the dual of

$$\min_x f(x) + \|x\|$$

is

$$\max_u -f^*(u) - I_{\{z:\|z\|_*\le 1\}}(-u)$$

or equivalently

$$\max_u -f^*(u) \quad \text{subject to } \|u\|_* \le 1$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

13.3.6 Shifting linear transformations

Dual formulations can help us by "shifting" a linear transformation between one part of the objective and another. Let's consider

$$\min_x f(x) + g(Ax)$$

Equivalently: $\min_{x,z} f(x) + g(z)$ subject to $Ax = z$. Like before:

$$\begin{aligned} g(u) &= \min_{x,z} f(x) + g(z) + u^T(z - Ax) \\ &= -\max_x (A^T u)^T x - f(x) - \max_z (-u)^T z - g(z) \\ &= -f^*(A^T u) - g^*(-u) \end{aligned}$$

Then dual is:

$$\max_u -f^*(A^T u) - g^*(-u)$$

Example: for a norm and its dual norm, $\|\cdot\|$, $\|\cdot\|_*$, the problems

$$\min_x f(x) + \|Ax\|$$

and

$$\max_u -f^*(A^T u) \quad \text{subject to } \|u\|_* \le 1$$

are primal and dual paris.

13.4 Dual cones

13.4.1 Definition

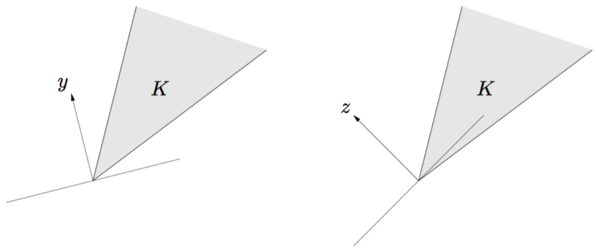
Recall that set $K \subseteq \mathbb{R}^n$ is a cone if $\forall x \in K, t \ge 0$, we have $tx \in K$.

The dual cone of K is defined as

$$K^* = \{y : y^T x \ge 0 \text{ for all } x \in K\}$$

Important properties:

- K^* is closed and convex.
- $K_1 \subseteq K_2 \Rightarrow K_2^* \subseteq K_1^*$
- K^{**} is the closure of the convex hull of K. (Hence if K is convex and closed, $K^{**} = K$)



Left. The halfspace with inward normal y contains the cone K, so $y \in K^*$. *Right.* The halfspace with inward normal z does not contain K, so $z \notin K$.

Figure 13.3: From B & V [1] pp. 52

13.4.2 Examples

- Linear subspace: the dual cone of a linear subspace V is $V^\perp$, its orthogonal complement. E.g. $(\text{row}(A))^* = (A)$.

- Norm cone: the dual cone of the norm cone

$$K = \{(x,t) \in \mathbb{R}^{n+1} : \|x\| \le t\}$$

  is the norm cone of its dual norm

$$K^* = \{(y,s) \in \mathbb{R}^{n+1} : \|y\|_* \le s\}$$

- Positive semidefinite cone: the convex cone $\mathbb{S}^n_+$ is self-dual, i.e. $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$.

$$Y \succeq 0 \quad \Longleftrightarrow \quad \text{Tr}(YX) \ge 0 \ \text{ for all } \ X \succeq 0$$

13.4.3 Dual cones and dual problems

Consider the cone constrained problem

$$\min_x f(x) \ \text{ subject to } \ Ax \in K$$

its dual problem is

$$\max_u -f^*(A^T u) - I_K^*(-u)$$

where $I_K^*(y) = \max_{z\in K} z^T y$ is the support function of K.

If K is a cone, we have $I_K^*(-u) = I_{K^*}(u)$, the this is equivalent to

$$\max_u -f^*(A^T u) \ \text{ subject to } \ u \in K^*$$

where K^* is the dual cone of K.

It is usually easier to handle cone constraints like $u \in K^*$ than constraints that the linear transform of x is in a cone, i.e. $Ax \in K$.

13.5 Double dual

Consider general minimization problem with linear constraints:

$$\begin{aligned} \min_x \quad & f(x) \\ \text{subject to} \quad & Ax \le b,\ Cx = d \end{aligned}$$

The Lagrangian is

$$L(x,u,v) = f(x) + (A^T u + C^T v)^T x - b^T u - d^T v$$

and hence the dual problem is

$$\begin{aligned} \min_x \quad & -f^*(-A^T u - C^T v) - b^T u - d^T v \\ \text{subject to} \quad & u \ge 0 \end{aligned}$$

Recall property: $f^{**} = f$ if f is closed and convex. Hence in this case, we can show that the dual of the dual is the primal.

Actually this also goes beyond linear constraints. Consider

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

If f and $h_1, ...h_m$ are closed and convex, and $l_1, ...l_r$ are affine, then the dual of the dual is the primal.

This is proved by viewing the minimization problem in terms of a bifunction. In this framework, the dual function corresponds to the conjugate of this bifunction. See Chapter 29 and 30 of Rockafellar. [2]

13.6 Dual subtleties

- We often transform the dual into an equivalent problem and still call this the dual. Under strong duality, we can use solutions of the (transformed) dual problem to characterize or compute the primal solutions.

  Warning: the optimal value of this transformed dual problem is not necessarily the optimal primal value.

- A common trick in deriving duals for unconstrained problems is to first transform the primal by adding a dummy variable and an equality constraint. e.g. The previous example of the lasso dual. Usually there is ambiguity in how to do this. Different choices can lead to different dual problems.

References

[1] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge University Press, 2004.

[2] R. Tyrrell Rockafellar. *Convex analysis.* Princeton University Press, 1970.