10-725: Optimization **Fall 2013**

Lecture 14 Duality uses and correspondences: Oct 10 2013

Lecturer: Ryan Tibshirani *Scribes: Jay-Yoon Lee*

This lecture's notes illustrate the duality uses and correspondence.

14.1 Topics Covered

- Uses of duality
- Conjugate functions
- Conjugates and dual problems

14.2 Uses of Duality

If we have a primal feasible point and we have a pair u, v that are dual feasible,

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

is called the duality gap between x and u, v. Since

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

,

a zero duality gap implies optimality. Also, this inequality on duality gap can be used as a stopping criterion in algorithms.

Under strong duality, given u^*, v^*, the dual optimal, the $x^* = \arg\min_x L(x, u^*, v^*)$ and often times this $\min_x L(x, u^*, v^*)$ can be expressed in closed form. When dual is easier to solve, we will exploit this fact and get the primal solution from the dual.

For example, the multivariate minimization problem of the Primal problem: (from B & V p.249)

$$\min_x \sum_{i=1}^{n} f_i(x_i) \qquad s.t.\ \ a^T x = b$$

can become univariate minimization problem Dual problem:

$$L(x,v) = \sum_{i=1}^{n} f_i(x_i) + v(b - \sum_{i=1}^{n} a_i Xi)$$

$$\begin{aligned} g(v) &= \min_x L(x,v) \\ &= \sum_{i=1}^{n} \min_{x_i}(f_i(x_i) - va_iX_i) + vb \\ &= \sum_{i=1}^{n} -f_i^*(a_iv) + vb \end{aligned}$$

Therefore the dual problem is

$$\begin{aligned} &\max_v - \sum_{i=1}^{n} f_i^*(a_iv) + bv \\ or \qquad &\min_v \sum_{i=1}^{n} f_i^*(a_iv) - bv \end{aligned}$$

Note that the dual problem became element wise problem and also note that $f_i^*(a_iv) = -\min_{x_i}(f_i(x_i) - va_iX_i)$ is a conjugate function (we will learn this concept later this lecture) which is convex in terms of v and bv is also convex function of v. Thus, the problem became a convex minimization problem with sclar variable v - much easier to solve than primal. For $v^* = \max_v g(v)$, x^* must minimize $L(x, v*) = \sum_{i=1}^{n} \min_{x_i}(f_i(x_i) - v^*a_iX_i)$ over x. So we can simply solve $\nabla f_i(x_i) = v^*a_i$ for each i, for differentiable f.

14.2.1 Dual norm

$$\begin{aligned} &norm && ||x|| \quad e.g. \; || \; ||_p, \; || \; ||_{nuc} \\ &dualnorm && ||y||_* = \max_{||x|| \le 1} x^T y \end{aligned}$$

- Typical examples:
  $(|| \; ||_p)_* = || \; ||_q \quad where \; \frac{1}{p} + \frac{1}{q} = 1$
  $(|| \; ||_n uc)_* = || \; ||_o p$
- Useful fact (Dual of the dual norm):
  $|| \; ||_{**} = || \; ||$
- Inequality from this definition (like Cauchy-Schwarz):

$$\forall x, y \qquad |x^T y| \le ||x|| \, ||y||_*$$

Proof: Take $z = \frac{x}{||x||}, \; ||z|| \le 1 \; ||y||_* = \max_{||x|| \le 1} x^T y \ge z^t y = \frac{x^T y}{||x||} \Rightarrow x^T y \le ||x|| ||y||_*$ ■

14.3 Conjugate functions

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

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

Note that f^* is always convex since it is the point wise maximum of convex functions in y.

- Fenchel's inequality: for any x, y,
$$f(x) + f^*(y) \geq x^T y$$
- Hence conjugate of conjugate $f^{**} \leq f$
- If f is closed and convex, then $f^{**} = f$
- 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 $f^*(w, z) = f_1^*(w) + f_2^*(z)$

14.3.1 Examples of conjugate functions

- Simple quadratic: $f(x) = \frac{1}{2} x^T Q x$ where $Q \succ 0$
$$f^*(y) = \frac{1}{2} y^T Q^- 1 y$$
Note that Fenchel's inequality gives
$$\frac{1}{2} x^T Q x + \frac{1}{2} y^T Q^- 1 y \geq x^T y$$
- Indicator function: $f(x) = I_C(x)$
$$f^*(y) = I_C^*(y) = \max_{x \in C} y^T x$$
called **support function** of C.
- Norm: $f(x) = ||x||$,
$$f * (y) = I_{\{z : ||z||_* \leq 1\}}(y)$$
Proof: Recall that $f^* * = f$ for a closed & convex f and that $f(x) = ||x||$ is such function. Thus, from the conjugate function of indicator function:
$$(\max_{x \in C} x^T y)^* = I_C(x)$$
Using this fact,
$$(||x||)^*(y) = (\max_{x \in \{z : ||z||_* \leq 1\}} x^T z)^*(y) = I_{\{z : ||z||_* \leq 1\}}(y)$$

■

14.4 Conjugates and dual problems

14.4.1 Lasso dual

Recall the lasso problem:

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

Since there's no constraints, its dual function is simply a constant (f^*), however rewriting the primal with $z = X\beta$:

$$\min_{\beta \in \mathbb{R}^p, z \in \mathbb{R}^n} \frac{1}{2}||y - z||_2^2 + \lambda||\beta||_1 \qquad s.t. \;\; z = X\beta$$

Deriving the dual,

$$\begin{aligned} L(\beta, z, u) &= \frac{1}{2}||y - z||_2^2 + \lambda||\beta||_1 + u^T(z - X\beta) \\ g(u) &= \min_{\beta \in \mathbb{R}^p, z \in \mathbb{R}^n} L(\beta, z, u) \\ &= \frac{1}{2}||y||_2^2 - \frac{1}{2}||y - u||_2^2 - I_{\{v:||v||_* \leq 1\}}(X^T u/\lambda) \end{aligned} \tag{14.1}$$

Note that the last Indicator term is from the conjugate function of L1-norm, that is

$$\begin{aligned} & \min_{\beta \in \mathbb{R}^p} \lambda(||\beta||_1 - (X^T u)^T \beta/\lambda) \\ = & -\max_{\beta \in \mathbb{R}^p} \lambda((X^T u)^T \beta/\lambda ||\beta||_1 -) \\ = & -\lambda * I_{\{v:||v||_* \leq 1\}}(X^T u/\lambda) \end{aligned}$$

Therefore rewriting eq.(14.1), the **lasso dual** problem is:

$$\begin{aligned} & \max_{u \in \mathbb{R}^n} \frac{1}{2}||y||_2^2 - \frac{1}{2}||y - u||_2^2 - I_{\{v:||v||_* \leq 1\}}(X^T u/\lambda) \\ or \qquad & \min_{u \in \mathbb{R}^n} \frac{1}{2}||y - u||_2^2 - I_{\{v:||v||_* \leq 1\}}(X^T u/\lambda) \end{aligned} \tag{14.2}$$

However, note that optimal solution of eq.(14.2) is not equal to the original problem since we modified the form of $g(u)$ although the strong duality holds here by Slater's condition. Also note that although the dual problem is not much easier to solve, in case of $X \in \mathbb{R}^{nxp}$ with $n << p$, then we gain by solving dual problem in space of $u \in \mathbb{R}^n$.

14.4.2 Conjugates and dual problems

The conjugate function appearing in Lasso dual problem is not a coincidence. By inspecting definition of conjugate function:

$$-f^*(u) = \min_{x \in \mathbb{R}^n} f(x) - u^T x$$

We can easily notice that $-f^*(u)$ is dual function of $f(x)$ with $x \geq 0$. For example, consider:

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} f(x) - u^T x \\ \Longleftrightarrow \qquad & \min_{x \in \mathbb{R}^n, z \in \mathbb{R}^n} f(x) + g(z) + u^T(z - x) = -f^*(u) - g * (-u) \end{aligned}$$

hence, using the definition of conjugate function, the dual problem simply becomes

$$\max_{u \in \mathbb{R}^n} -f^*(u) - g^*(-u)$$

References

[Boyd] S. Boyd and L. Vandenberghe, “Convex Optimization’ *Cambridge University Press*