10-725/36-725: Convex Optimization **Fall 2018**

Lecture 13: Duality: Uses and Correspondences

Lecturer: Ryan Tibshirani *Scribes: Ben Eysenbach, Ramon Villa-Cox*

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

Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications. They may be distributed outside this class only with the permission of the Instructor.*

13.1 Review: KKT conditions

KKT conditions necessary for optimality under strong duality, and always sufficient:

1. stationarity: gradient of Lagrangian is 0 (or similar for subgradients)
2. complementary slackness
3. primal feasibility
4. dual feasibility

We looked at two uses of duality:

1. The duality gap can be used to determine optimality, and can be used as a stopping criterion in algorithms.

$$\begin{aligned} f(x) - f^* &\leq f(x) - g(u,v) \\ g^* - g(u,v) &\leq f(x) - g(u,v) \end{aligned}$$

2. Under strong duality, given optimal dual variables, we can simply solve the Lagrangian for the primal variables, fixing these optimal dual variables.

$$\min_x \mathcal{L}(x, u^*, v^*)$$

Often, solutions of this unconstrained problem give an explicit characterization of primal solutions from dual solutions. If the solution is unique, it must be the primal solution, which is helpful when the dual is easier to solve than the primal.

Example: B & V (pg 245)

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

where each $f_i : \mathbb{R} \to \mathbb{R}$ is smooth, strictly convex. The Dual function can be written using the *conjugate function.*

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

This allows us to decompose the problem into n separate (i.e., easy) problems. Once we have v^*, solving for x^* is easy:

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

Since each f_i is strictly convex, we can simply compute the derivative and set equal to 0 for each i:

$$\nabla f_i(x_i) = a_i v^*$$

13.2 Dual norms

Given a norm $\|x\|$, we define the *dual norm* as:

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

Some useful facts about dual norms:

- ℓ_p norm dual: $(\|x\|_p)_* = \|x\|_q$, where $\frac{1}{p} + \frac{1}{q} = 1$. For example, $p = 2$ and $q = 2$ are dual pairs, as are $p = 1$ and $q = \infty$
- Trace norm dual: $(\|X\|_{\text{tr}})_* = \|X\|_{\text{op}} = \sigma_1(X)$
- The dual norm of the dual norm is the original norm!

Lemma 13.1 *A "generalized Holder" inequality gives us the following:*

$$|z^T x| \leq \|z\| \|x\|_*$$

Proof: Let $y = \frac{z}{\|z\|}$, then

$$\begin{aligned} |y^T x| &\leq \max_{\|w\| \leq 1} w^T x = \|x\|_* \\ \frac{|z^T x|}{\|z\|} &\leq \|x\|_* \end{aligned}$$

■

13.3 Conjugate functions

Definition: A *closed function* is one where all sublevel sets are closed. That is, the set $\{x : f(x) \leq t\}$ is closed for all t.

Definition: For a function $f : \mathbb{R}^n \to \mathbb{R}$, define its *conjugate*[1] $f^* : \mathbb{R}^n \to \mathbb{R}$ as

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

This should be interpreted as the maximum gap between the linear function $y^T x$ and $f(x)$. The conjugate is always convex (pointwise max over convex functions). Conjugate functions satisfy various properties:

- Fenchel's inequality:
$$f(x) + f^*(y) \geq x^T y \quad \forall x, y$$
Proof: By straight use of the definition of the conjugate,
$$f^*(y) = \max_z \; y^T z - f(z) \geq y^T x - f(x) \quad \forall x, y$$
$$f^*(y) + f(x) \geq y^T x \quad \forall x, y$$
■
- For any f, $f^{**}(y) < f(y)$ for all y.
- If f is also closed and convex, then $f^{**} = f$.
- If f is closed and convex, then for any x, y:
$$\begin{aligned} x \in \partial f^*(x) &\iff y \in \partial f(x) \\ &\iff f(x) + f^*(y) = x^T y \end{aligned}$$
For example, if f and f^* are differentiable, then $(\nabla f)^{-1} = \nabla f^*$. Said another way,
$$x = \nabla f^*(y) \iff y = \nabla f(x)$$
- If $f(u, v) = f_1(u) + f_2(v)$, then
$$f^*(w, z) = f_1^*(w) + f_2^*(z)$$

Examples:

- Simple quadratic: $f(x) = \frac{1}{2} x^T Q x$, where $Q \succ 0$ (function striclty concave). Then, $y = Q^{-1} x$
$$f^*(y) = \frac{1}{2} y^T Q^{-1} y$$
- Indicator Function: $f(x) = I_C(x)$. Then,
$$f^*(y) = I_C^*(y) = \max_{x \in C} y^T x \quad (I_C^* \text{ is called the } \textit{support function})$$
- Norm: $f(x) = \|x\|$, then
$$f^*(y) = I_{\{z : \|z\|_* \leq 1\}}(y)$$
This is a simple application of the conjugate of the indicator function (dual of the norm is of the form of conjugate of indicator)

[1] Physics people call this the *Legendre transform*for differentiable f.

Lasso dual

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

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

Its dual is simply the solution, f^* (a constant). To get something more interesting, we will transform the primal into an equivalent problem:

$$\min_{\beta, z} \frac{1}{2}\|y - x\|_2^2 + \lambda\|\beta\|_1 \quad \text{s.t.} \quad z = X\beta$$

The dual function of the transformed primal is

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

For the last line notice that, for the left minimization term, we have that $z^* = y - u$. Replacing in the argument and completing the square (adding $\|y\|_2^2$) we get the desired expression. For the right minimization term, we used the fact that

$$\min_{\beta} \ \lambda\|\beta\|_1 - u^T X\beta = -\lambda \max_{\beta} \ \|\beta\|_1 - \frac{u^T X}{\lambda}\beta$$

Which is the conjugate of $\|\cdot\|_1$ evaluated at $\frac{u^T X}{\lambda}$. We also used the fact that $\lambda I_C(x) = I_C(x)$ for all C and x (the function is either 0 or ∞, so scaling doesn't do anything). Putting it together we get the third line. Thus, the lasso dual is

$$\begin{aligned} &\max_{u} \frac{1}{2}\left(\|y\|_2^2 - \|y - u\|_2^2\right) \quad \text{s.t.} \quad \|X^T u\|_\infty \leq \lambda \\ &\iff \min_{u} \|y - u\|_2^2 \quad \text{s.t.} \quad \|X^T u\|_\infty \leq \lambda \end{aligned}$$

Slater's condition holds (as primal is unconstrained) so strong duality also holds. Thus, given the dual solution u, we can compute the lasso solution as $X\beta = y - u$.
WARNING: The optimal value of the last problem, given the transformation done, is not the same as the optimal value of the LASSO objective. So we need to be careful when computing a stopping rule based on the duality gap.

Conjugates and dual problems

Note that the definition of the conjugate looks a lot like a Lagrangian:

$$-f^*(u) = \min_{x} f(x) - u^T x$$

If the primal problem is $\max_x f(x) + g(z)$ subject to $x = z$, then the dual problem is

$$\max_{u} -f^*(u) - g^*(u)$$

We can use the dual to remove linear transforms. In particular, consider the following problem:

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

The dual (by adding a constraint that $Ax = z$) is:

$$\max_u f(A^T u) \quad \text{s.t.} \quad \|u\|_* \leq 1$$

13.4 Dual tricks and subtleties

WARNING: We often simplify the dual, but forget that the value of the simplified dual may be different from the original dual.

WARNING: The dual function we obtain depends on how we introduce dummy variables with equality constraints. Usually there are many ways to do this and even though they are equivalent (if strong duality holds), some might have more desirable properties than others (so we need to consider all cases).