Duality uses and correspondences

Barnabas Poczos & Ryan Tibshirani
Convex Optimization 10-725/36-725

Remember KKT conditions

Recall that for the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) \le 0, \;\; i = 1, \ldots m \\ & \ell_j(x) = 0, \;\; j = 1, \ldots r \end{aligned}$$

the KKT conditions are

- $0 \in \partial f(x) + \sum_{i=1}^m u_i \partial h_i(x) + \sum_{j=1}^r v_i \partial \ell_j(x)$ (stationarity)
- $u_i \cdot h_i(x) = 0$ for all i (complementary slackness)
- $h_i(x) \le 0, \; \ell_j(x) = 0$ for all i, j (primal feasibility)
- $u_i \ge 0$ for all i (dual feasibility)

These are necessary for optimality (of a primal-dual pair $x^\star$ and $u^\star, v^\star$) under strong duality, and always sufficient

Uses of duality

Two key uses of duality:

- For x primal feasible and u, v dual feasible,

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

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

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

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

- Under strong duality, given dual optimal $u^\star, v^\star$, any primal solution minimizes $L(x, u^\star, v^\star)$ over $x \in \mathbb{R}^n$ (i.e., satisfies stationarity condition). This can be used to characterize or compute primal solutions

Solving the primal via the dual

An important consequence of stationarity: under strong duality, given a dual solution $u^\star, v^\star$, any primal solution $x^\star$ solves

$$\min_{x\in\mathbb{R}^n} \; f(x) + \sum_{i=1}^m u_i^\star h_i(x) + \sum_{j=1}^r v_i^\star \ell_j(x)$$

Often, solutions of this unconstrained problem can be expressed explicitly, giving an explicit characterization of primal solutions from dual solutions

Furthermore, suppose the solution of this problem is unique; then it must be the primal solution $x^\star$

This can be very helpful when the dual is easier to solve than the primal

Consider as an example (from B & V page 249):

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

where each $f_i : \mathbb{R} \to \mathbb{R}$ is a strictly convex function. Dual function:

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

where f_i^* is the conjugate of f_i, to be defined shortly

Therefore the dual problem is

$$\max_{v \in \mathbb{R}} \; bv - \sum_{i=1}^{n} f_i^*(a_i v)$$

or equivalently

$$\min_{v \in \mathbb{R}} \; \sum_{i=1}^{n} f_i^*(a_i v) - bv$$

This is a convex minimization problem with scalar variable—much easier to solve than primal

Given $v^\star$, the primal solution $x^\star$ solves

$$\min_{x \in \mathbb{R}^n} \; \sum_{i=1}^{n} (f_i(x_i) - a_i v^\star x_i)$$

Strict convexity of each f_i implies that this has a unique solution, namely $x^\star$, which we compute by solving $\nabla f_i(x_i) = a_i v^\star$ for each i

Outline

Today:

- Dual norms
- Conjugate functions
- Dual cones
- Dual tricks and subtleties

(Note: there are many other uses of duality and relationships to duality that we could discuss, but not enough time...)

Dual norms

Let $\|x\|$ be a norm, e.g.,

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

We define its dual norm $\|x\|_*$ as

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

Gives us the inequality $|z^T x| \leq \|z\| \|x\|_*$, like Cauchy-Schwartz. Back to our examples,

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

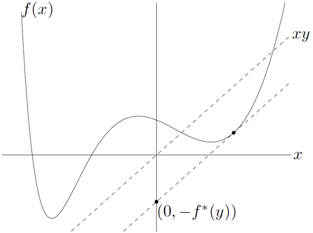
Dual norm of dual norm: can show that $\|x\|_{**} = \|x\|$

Conjugate function

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

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

Note that f^* is always convex, since it is the pointwise maximum of convex (affine) functions in y (f need not be convex)



$f^*(y)$: maximum gap between linear function $y^T x$ and $f(x)$

(From B & V page 91)

For differentiable f, conjugation is called the Legendre transform

Properties:

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

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

- Hence conjugate of conjugate f^{**} satisfies $f^{**} \leq f$
- 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 $f(u, v) = f_1(u) + f_2(v)$ (here $u \in \mathbb{R}^n$, $v \in \mathbb{R}^m$), then

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

Examples:

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

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

Note that Fenchel's inequality gives:

$$\frac{1}{2}x^T Qx + \frac{1}{2}y^T Q^{-1} y \geq x^T y$$

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

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

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 $\|\cdot\|$

Why? Note that if $\|y\|_* > 1$, then there exists $\|z\| \leq 1$ with $z^T y = \|y\|_* > 1$, so

$$(tz)^T y - \|tz\| = t(z^T y - \|z\|) \to \infty, \quad \text{as } t \to \infty$$

i.e., $f^*(y) = \infty$

On the other hand, if $\|y\|_* \leq 1$, then

$$z^T y - \|z\| \leq \|z\| \|y\|_* - \|z\| \leq 0$$

and $= 0$ when $z = 0$, so $f^*(y) = 0$

Lasso dual

Recall the lasso problem:

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

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

$$\min_{\beta \in \mathbb{R}^p,\, z \in \mathbb{R}^n} \frac{1}{2}\|y - z\|_2^2 + \lambda\|\beta\|_1 \quad \text{subject to} \quad z = X\beta$$

so dual function is now

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

Therefore the lasso dual problem is

$$\max_{u \in \mathbb{R}^n} \frac{1}{2}\Big(\|y\|_2^2 - \|y-u\|_2^2\Big) \ \text{ subject to } \ \|X^T u\|_\infty \le \lambda$$

or equivalently

$$\min_{u \in \mathbb{R}^n} \|y-u\|_2^2 \ \text{ subject to } \ \|X^T u\|_\infty \le \lambda$$

Note that strong duality holds here (Slater's condition), but the optimal value of the last problem is not necessarily the optimal lasso objective value

Further, note that given $\hat{u}$, any lasso solution $\hat{\beta}$ satisfies (from the z block of the stationarity condition) $\hat{z} - y + \hat{\beta} = 0$, i.e.,

$$X\hat{\beta} = y - \hat{u}$$

So the lasso fit is just the dual residual

Conjugates and dual problems

Conjugates appear frequently in derivation of dual problems, via

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

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

$$\begin{aligned} & \min_{x \in \mathbb{R}^n} \; f(x) + g(x) \\ \iff \quad & \min_{x \in \mathbb{R}^n,\, z \in \mathbb{R}^n} \; f(x) + g(z) \;\; \text{subject to} \;\; x = z \end{aligned}$$

Lagrange dual function:

$$g(u) = \min_{x \in \mathbb{R}^n,\, z \in \mathbb{R}^n} \; f(x) + g(z) + u^T(z - x) = -f^*(u) - g^*(-u)$$

Hence dual problem is

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

Examples of this last calculation:

- Indicator function: dual of

$$\min_{x \in \mathbb{R}^n} \; f(x) + I_C(x)$$

is

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

where I_C^* is the support function of C

- Norms: the dual of

$$\min_{x \in \mathbb{R}^n} \; f(x) + \|x\|$$

is

$$\max_{u \in \mathbb{R}^n} \; -f^*(u) \quad \text{subject to} \quad \|u\|_* \le 1$$

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

Cones

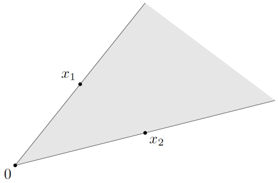
A set $K \in \mathbb{R}^n$ is called a cone if

$$x \in K \quad \Rightarrow \quad \theta x \in K \quad \text{for all } \theta \geq 0$$

It is called a convex cone if

$$x_1, x_2 \in C \quad \Rightarrow \quad \theta_1 x_1 + \theta_2 x_2 \in C \quad \text{for all } \theta_1, \theta_2 \geq 0$$

i.e., K is convex and a cone



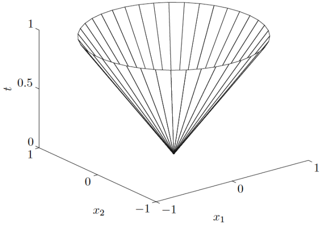
(From B & V page 26)

Examples:

- Linear subspace: any linear subspace is a convex cone
- Norm cone: if $\|\cdot\|$ is a norm then

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

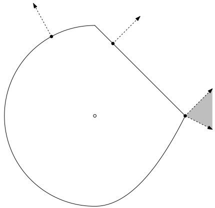
is a convex cone, called a norm cone (epigraph of norm function). Under ℓ_2, called second-order cone, e.g.,



(From B & V page 31)

- Normal cone: given a set C, recall we defined its normal cone at a point $x \in C$ as

$$\mathcal{N}_C(x) = \{g \in \mathbb{R}^n : g^T x \geq g^T y \text{ for any } y \in C\}$$



This is always a convex cone, regardless of C

- Positive semidefinite cone: consider the set of (symmetric) positive semidefinite matrices

$$\mathbb{S}^n_+ = \{X \in \mathbb{R}^{n \times n} : X = X^T,\ X \succeq 0\}$$

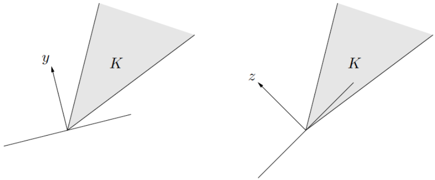
This is a convex cone, because for $A, B \succeq 0$ and $\theta_1, \theta_2 \geq 0$, $x^T(\theta_1 A + \theta_2 B)x = \theta_1 x^T A x + \theta_2 x^T B x \geq 0$

Dual cones

For a cone $K \in \mathbb{R}^n$,

$$K^* = \{y \in \mathbb{R}^n : y^T x \geq 0 \text{ for all } x \in K\}$$

is called its dual cone. This is always a convex cone (even if K is not convex)



Note that $y \in K^* \iff$ the halfspace $\{x \in \mathbb{R}^n : y^T x \geq 0\}$ contains K

(From B & V page 52)

Important property: if K is a closed convex cone, then $K^{**} = K$

Examples:

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

- 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, meaning $(\mathbb{S}^n_+)^* = \mathbb{S}^n_+$. Why? Check that

$$Y \succeq 0 \iff \mathrm{tr}(YX) \ge 0 \text{ for all } X \succeq 0$$

  by looking at the eigenvalue decomposition of X

Dual cones and dual problems

Consider the constrained problem

$$\min_{x \in K} f(x)$$

Recall that its dual problem is

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

where recall $I_K^*(y) = \max_{z \in K} z^T y$, the support function of K. If K is a cone, then this is simply

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

where K^* is the dual cone of K, because $I_K^*(-u) = I_{K^*}(u)$

This is quite a useful observation, because many different types of constraints can be posed as cone constraints

Dual subtleties

- Often, we will 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 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

  Usually there is ambiguity in how to do this, and different choices lead to different dual problems!

Double dual

Consider general minimization problem with linear constraints:

$$\min_{x \in \mathbb{R}^n} \ f(x) \quad \text{subject to} \ Ax \le b, \ \ Cx = d$$

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

$$\max_{u \in \mathbb{R}^m, \, v \in \mathbb{R}^r} \ -f^*(-A^T u - C^T v) - b^T u - d^T v \quad \text{subject to} \ u \ge 0$$

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, the connection (between duals of duals and conjugates) runs much deeper than this, beyond linear constraints. Consider

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) \leq 0, \ \ i = 1, \ldots m \\ & \ell_j(x) = 0, \ \ j = 1, \ldots r \end{aligned}$$

If f and $h_1, \ldots h_m$ are closed and convex, and $\ell_1, \ldots \ell_r$ 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 (for more, read Chapters 29 and 30 of Rockafellar)

References

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapters 2, 3, 5
- R. T. Rockafellar (1970), "Convex analysis", Chapters 12, 13, 14, 16, 28–30