10-725: Optimization **Fall 2012**

Lecture 17: October 23

Lecturer: Geoff Gordon/Ryan Tibshirani *Scribes: Yifei Ma, Mahdi Pakdaman Naeini*

17.1 Three things talked in class

- How to construct the dual problem with/without constraints.
- What is the conjuate of a function and how to construct the dual via conjugate functions.
- What is the dual cone and software packages that utilize cone constraints.

17.2 Constructing the dual problem

17.2.1 Recall

The primal and dual of an optimization problem is essentially a change in the order of which variables to optimize,

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

$$\min_x \max_{u\geq 0, v|x} L(x,u,v) = \min_{h_i(x)\geq 0, l_j(x)=0} f(x) \qquad \text{(primal problem)}$$

$$\max_{u\geq 0, v} \min_{x|u,v} L(x,u,v) = \max_{u\geq 0, v} g(u,v) \qquad \text{(dual problem)}$$

Special with the dual problem, under strong duality, given any dual solution u^*, v^*, the minimizer x^* solves $g(u^*, v^*) = L(x^*, u^*, v^*)$, i.e.,

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

and conversely the minimizer for the above is the primal solution if the primal solution is unique.

17.2.2 Two running examples

Example:*Running example 1.* Consider the following convex problem when $f_i : \mathbb{R} \to \mathbb{R}$ is strictly convex,

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

The dual function is $g(v) = \min_{x \in \mathbb{R}^n} L(x, v) = \min_{x \in \mathbb{R}^n} \sum_{i=1}^{n} f_i(x_i) + v(b - a^T x) = bv + \sum_{i=1}^{n} \min_{x_i \in \mathbb{R}} (f_i(x_i) - a_i v x_i)$.

Now the dual problem is a univariate problem that is much easier to solve,

$$v^* = \arg\max_{v \in \mathbb{R}} g(v) = bv - \sum_{i=1}^{n} \max_{x_i \in \mathbb{R}} (a_i v x_i - f_i(x_i)), \tag{17.1}$$

$$x_i^* = \arg\max_{x_i} a_i v^* x_i - f_i(x_i). \tag{17.2}$$

Example: *Unconstrained optimization. Running example 2.* In order to derive *a* dual problem of the Lasso problem,

$$\min_{x \in \mathbb{R}^p} \frac{1}{2}\|y - Ax\|^2 + \lambda \|x\|_1$$

We mandate a constraint by introducing $z = Ax$ as,

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

Now the dual function is

$$\begin{aligned} g(u) &= \min_{x \in \mathbb{R}^p, z \in \mathbb{R}^n} \frac{1}{2}\|y - z\|^2 + \lambda \|x\|_1 + u^T(z - Ax) \\ &= \min_{z \in \mathbb{R}^n} (\frac{1}{2}\|y - z\|^2 + u^T z) + \min_{x \in \mathbb{R}^p} (\lambda \|x\|_1 - u^T A x) \end{aligned} \tag{17.3}$$

where

$$\begin{aligned} \text{first term} &= \min_{z \in \mathbb{R}^n} (\frac{1}{2}\|z\|^2 + (u - y)^T z + \frac{1}{2}\|y\|^2) \\ &= \min_{z \in \mathbb{R}^n} (\frac{1}{2}\|z + (u - y)\|^2 - \frac{1}{2}\|u - y\|^2 + \frac{1}{2}\|y\|^2) = \frac{1}{2}\|y\|^2 - \frac{1}{2}\|y - u\|^2, \end{aligned}$$

and because $\partial \lambda \|x\|_1 \in [-\lambda, \lambda]^p$, $\partial_x(\lambda \|x\|_1 - u^T A x) = \partial(\lambda \|x\|_1) - u^T A = 0 \Rightarrow u^T A \in [-\lambda, \lambda]^p$.

$$\text{second term} = \left\{ \begin{matrix} 0 & -\lambda \leq u^T A < \lambda \\ -\infty & otherwise \end{matrix} \right\} \approx I_{\{v : \|v\|_\infty \leq 1\}}(A^T u / \lambda)$$

Therefore the *lasso dual* function and dual problem are resp.

$$\begin{aligned} g(u) = \frac{1}{2}(\|y\|^2 - \|y - u\|^2) - I_{\{v : \|v\|_\infty \leq 1\}}(A^T u / \lambda) \\ \text{and } \max_{u \in \mathbb{R}^n} \frac{1}{2}(\|y\|^2 - \|y - u\|^2) \text{ subject to } \|A^T u\|_\infty \leq \lambda. \end{aligned}$$

The primal correspondance,

$$\begin{aligned} x^* &= \arg\max_{x \in \mathbb{R}^p} (u^T A x - \lambda \|x\|_1) \\ z^* &= \arg\max_{z \in \mathbb{R}^n} (-u^T z - \frac{1}{2}\|y - z\|^2). \end{aligned} \tag{17.4}$$

17.2.3 Summary

- (Disclamer). For clarity (or laziness), an equivalent problem of the dual problem is still called the dual problem, even though its optimal value does not necessarily equal the optimal primal value.
- The dummy variable z used in example 2 is a common trick in deriving duals for unconstrained problems.
- (Question for the reader). How to derive a dual problem for example 2 when the dummy variable is defined alternatively as $z = x$?

17.3 Conjugate functions: definition and properties

In both the above examples, (17.1), (17.2), (17.3), (17.4) fall in the framework of *conjugate functions*.

Definition 17.1 *Given a function $f : \mathbb{R}^n \to \mathbb{R}$, define its* conjugate $f^* : \mathbb{R}^n \to \mathbb{R}$ *as*

$$f^*(y) = \max_{x \in \mathbb{R}^n} y^T x - f(x). \tag{17.5}$$

The conjugate of differentiable f is called its Legendre transform.

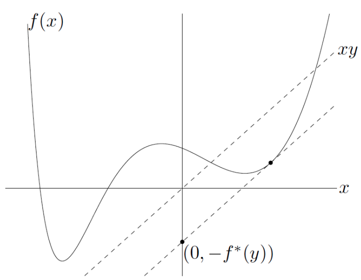


Figure 17.1: conjugate

Note that the $f^*(y)$ is always convex even when f is not. This is because (17.5) is a maximization of the family of linear functions $f(y; x, f(x)) = x^T y - f(x)$, where x and $f(x)$ are viewed as parameters.

Properties

- Fenchel's inequality: for any x, y, $f(x) + f^*(y) \geq x^T y$. (proof by definition)
- Hence, $f^{**} \leq f$. (proof left for the reader)
- If f is closed and convex, then $f^{**} = f$. (proof left for the reader)
- If f is closed and convex, then for any x, y

$$x \in \partial f^*(y) \Leftrightarrow y \in \partial f(x) \Leftrightarrow f(x) + f^*(y) = x^T y. \quad \text{(proof followed)}$$

- If $f(u, v) = f_1(u) + f_2(v)$ then $f^*(w, z) = f_1^*(w) + f_2^*(z)$. (c.f. (17.3))

Proof: Take ∂_x or ∂_y on both sides of $f(x)+f^*(y) = x^T y$, we get the first two formulations resp. It suffices to show that $x \in \partial f^*(y) \Rightarrow f(x) + f^*(y) = x^T y$.

Take the subdifferential of (17.5), which is a maximization over a family of convex linear functions, we have

$$\partial f^*(y) = conv(\bigcup\{x : y^T x - f(x) = \max_{z\in\mathbb{R}^n} y^T z - f(z)\})$$

Also note that $\forall x = \sum \lambda_i x_i$ with $\lambda_i > 0$, $\sum \lambda_i = 1$, and $y^T x_i - f(x_i) = \max_{z\in\mathbb{R}^n} y^T z - f(z)$, we have

$$f(x) \le \sum \lambda_i f(x_i)$$

$$\Rightarrow \qquad f^*(y) \ge y^T x - f(x) \ge \sum \lambda_i (y^T x_i - f(x_i)) = f^*(y).$$

■

17.4 The conjugate function as a way to form the dual problem

There are four groups of objective functions that may generalize to most common cases, namely indicator functions, norm functions, pairs of unconstraint functions, and convex functions with general linear constraints.

17.4.1 Conjugates of the indicator function and the norm function

Example: *indicator function* If $f(x) = I_C(x)$, then its conjugate is

$$f^*(y) = \max_{x\in\mathbb{R}^n} y^T x - I_C(x) = \max_{x\in C} y^T x = I^*_C(y). \tag{17.6}$$

Here, the $f^*(y)$ is the support function of C. Geometrically, it is the perpendicular distance from the origin to the supporting hyperplane of C that has normal vector y and that contains C on the same side as the origin. Specially, when $C = \|x\| \le \lambda$ a is norm constraint, then $I^*_C(-u) = \lambda\|u\|_*$ is a norm penalty.

Example: *norm* If $f(x) = \|x\|$, then its conjugate is

$$f^*(y) = \max_{x\in\mathbb{R}^n} y^T x - \|x\| = \begin{cases} 0 & \|y\|_* \le 1 \\ \infty & else \end{cases}. \tag{17.7}$$

Recall that $\|y\|_* = \max\limits_{\|x\|\le 1} y^T x$ and hence $\|x\|\|y\|_* \ge y^T x$. When $\|y\|_* \le 1$, $y^T x \le \|x\|\|y\|_* \le \|x\|$, equality (necessarily) when $x = 0$. When $\|y\|_* > 1$, the x that allows $y^T x = \|x\|\|y\|_* > \|x\|$ can make this inequality gap arbitrarily large.

17.4.2 Dual construction via conjugates of pairs of functions

Example: *pairs of functions* Consider the problem

$$\min_{x\in\mathbb{R}^n} f(x) + g(x) \Leftrightarrow \min_{x\in\mathbb{R}^n, z\in\mathbb{R}^n} f(x) + g(z) \text{ subject to } x = z \tag{17.8}$$

The Lagrange dual function of RHS, in a way similar to (17.3), is

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

Hence, the dual problem corresponding to (17.8) is

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

Example: *indicator function* The dual of

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

is, via utilizing (17.8) and (17.6),

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

Specially, when $C = \|x\| \leq \lambda$ a is norm constraint, then $I_C^*(-u) = \lambda\|u\|_*$ is a norm penalty.

Example: *norm* The dual of

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

is, via utilizing (17.8) and (17.7),

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

Notice that the above two examples solve the same problem with the definition of dual and primal altered, when C is exactly $\{x : \|x\|_* \leq 1\}$. Following the property $f^{**} = f$ under closeness and convexity, it is easy to understand that $I_C^*(-u) = \|u\|$, c.f. (17.7).

17.4.3 Linear constraints and double dual

Example: Consider the general convex problem with linear constraints,

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} f(x) \\ &\text{subject to } Ax \leq b, Cx = d \end{aligned}$$

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

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

Now, the Lagrangian of the dual should be the same as above but represented by f^*,

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

The dual of dual is

$$\begin{aligned}\max_{u\in\mathbb{R}^m, v\in\mathbb{R}^r} L^*(x,u,v) &= \max_{t\in\mathbb{R}}(x^T t - f^*(t)) + \max_{t\in\mathbb{R}^m}(x^T A^T u - b^T u - y^T u) + \max_{v\in\mathbb{R}^r}(x^T C^T v - d^T v) \\ &= f(x) + I_{\{(x,y):Ax-b-y=0\}}(x,y) + I_{\{x:Cx-d=0\}}(x).\end{aligned}$$

Hence the dual of dual problem is exactly the primal,

$$\begin{aligned}&\min_{x\in\mathbb{R}^n, y\geq 0} f(x) + I_{\{(x,y):Ax=b+y\}}(x,y) + I_{\{x:Cx=d\}}(x) \\ &= \min_{x\in\mathbb{R}^n} f(x) \text{ subject to } Ax \geq b, Cx = d.\end{aligned}$$

The above conclusion that *the dual of the dual is the primal* can be extended to the general case

$$\min_{x\in\mathbb{R}^n} f(x) \text{ subject to } h_i(x) \leq 0, l_j(x) = 0$$

where f and h_i are closed and convex and l_j are affine.

17.5 Cones and its application in optimization problems

Cone is one of the basic concepts in convex optimization which is going to be very useful in future lectures. The formal definition of cone is as follow.

Definition 17.2 *A set $K \in \mathbb{R}$ is called Cone iff $x \in K \Longrightarrow \theta x \in K$ for all $\theta \geq 0$*

Examples of cone could be a the origin itself or any single ray going through the origin. Or even any two rays intersecting at origin can be also considered as a cone.

If a cone K is a convex set then it is called **convex cone**, which is equivalent to say:

Definition 17.3 *A set $K \in \mathbb{R}$ is called a Convex Cone iff $x_1, x_2 \in C \Longrightarrow \theta_1 x_1 + \theta_2 x_2 \in C$ for all $\theta_1, \theta_2 \geq 0$*

here a simple convex cone has been depicted in figure 17.1

Some Examples of convex cone:

- Any Linear subspace
- Non negative Orthant $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x_i \geq 0 \text{ all } i\}$
- Positive Semi definite cone: Set of (symmetric) positive semidefine materices
- Norm Cone: epigraph of norm function
- Normal Cone of Set C at any point x is always a convex cone.

Figure 17.1: A picture of Convex Cone.

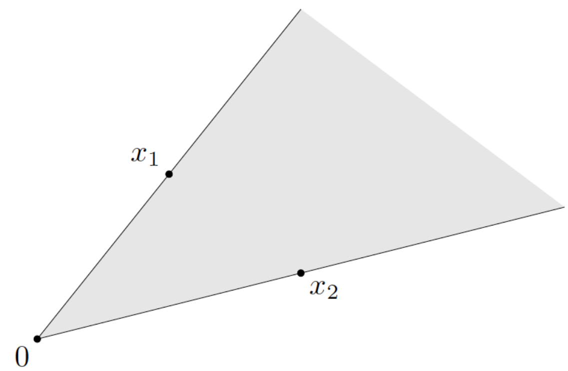
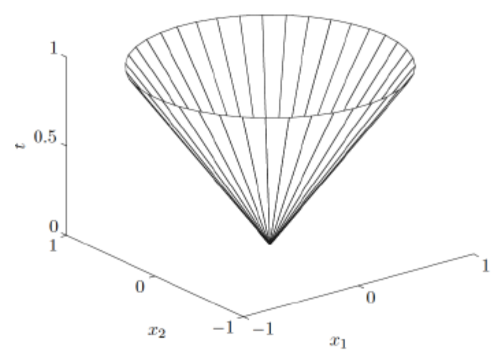


Figure 17.2: Second Order Cone.



Definition 17.4 *Norm Cone: if* $||.||$ *is a norm then the set defnied as* $K = \{(x,t)\mathbb{R}^{n+1} : ||x|| \leq t\}$ *is a convex cone called a norm cone.*

The second order cone which is a norm cone under 2-norm has be depicted in 17.2

Definition 17.5 *Proper cone: a cone* K *is called paroper if it is convex, closed, solid and pointed*

An example of proper cone could be nonnegative orthant.

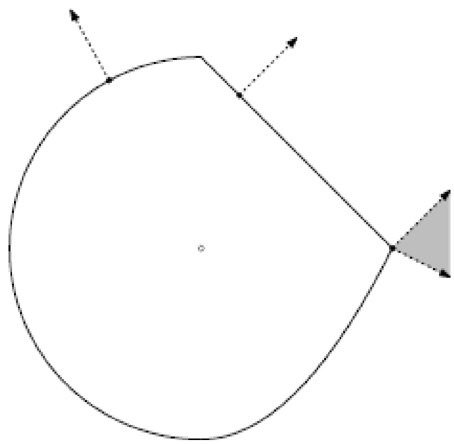
Definition 17.6 *Normal Cone: given a set* C *its normal cone at point* x *is defined as:* $\mathbb{N}_c(x) = \{g \in \mathbb{R}^n : g^T(y - x) \leq 0 \textit{for any } y \in C\}$

here normal cone for a set C at three different point been depicted in figure 17.3

17.5.1 generalized inequality

We can use the cone to generalized the definition of inequality as following:

Figure 17.3: A picture of Convex Cone.



Definition 17.7 *If $K \in \mathbb{R}^n$ is a proper cone then it induces a* ***generalized inequatilit*** $\leq_K$ *over* $\mathbb{R}^n$ *via* $x \leq_K y iff y - x \in K$

Some Examples of inequailities induced by proper cones:

- componentwise inequailty: for two vecto x and y, $x \leq_{\mathbb{R}^n_+}$ iff $x_i \leq y_i$ for all i
- Matrix inequaity: for two matrix X and Y, $X \leq_{\mathbb{S}^n_+} Y$ iff $Y - X \in \mathbb{S}^n_+$

17.5.2 Dual Cones

Dual cone is one of important concept in convex optimization that specially could be very useful in deriving the dual program. Here we first define the concept of dual cone:

Definition 17.8 *For a cone $K \in \mathbb{R}^n$ its dual cone K^* is defined as: $K^* = \{y \in \mathbb{R}^n : y^T x \geq 0$ for all $x \in K\}$.*

There are two important property for dual cones:

- the Dual cone K^* is always a convex cone;
- If K is closed and convex then $K^{**} = K$

Here is three popular examples of dual cones:

- Linear Subspace: the dual cone of any linear subspace is its orthogonal complement. The useful special case: $(RowSpace(A))^* = null(A)$
- The dual of a norm cone $K = \{(x,t) \in \mathbb{R}^{n+1} : ||x|| \leq t\}$ is the norm cone of its dual norm which is $K^* = \{(y,s) \in \mathbb{R}^{n+1} : ||y||_* \leq s\}$
- Non negative orthant is self dual $(\mathbb{R}^n_+)^* = \mathbb{R}^n_+$.
- Positive semidefinite cone is self dual: $((S)^n_+)^* = (S)^n_+$

17.5.2.1 Dual cone and Dual Program

We can use the dual cone concept to derive the dual program as following:

Consider the constraint poroblem $min_{x \in K} f(x)$, Recall that its dual program is $max_{u \in \mathbb{R}^n} - f^*(u) - I_K^*(-u)$, where $I_K^*(y) = max_{z \in K} z^T y$ is the support function of K. Now, if we assume that K is a cone, and its dual cone is K^*, then considering the fact that $I_K^*(-u) = I_{K^*}(u)$, the support funciton would be equal to $max_{u \in K^*} - f^*(u)$ where K^* is the dual cone of K.

The above observation could be very useful in findignt the dual program of optimization problems because many different types of constraints can be posed as cone constraints.

17.5.3 Conic Solvers

Two general convex optimization packages, that rely on transforming a convex problem into conic form (i.e. one with cone constraints) are **CVX** [1] and **TFOCS** [2] that both are implemented in matlab. Roughly speaking **CVX** is more general and use better transfromation for derinving coninc constraints, However **TFOCS** can be a lot faster even though its less general.

References

[S. Boyd, L. Vandenberghe] , (2004), "Convex Optimization", Cambridge University Press, Chapters 2, 3, 5

[B. Grunbaum] , (2003), Convex Polytopes, Springer, Chapters 2, 3

[R. T. Rockafellar] , (1970), Convex Analysis, Princeton University Press, Chapters 12, 13, 14, 16

[1] M. Grant and S. Boyd (2008), Graph implementations for nonsmooth convex problems, http://cvxr.com/cvx

[2] S. Becker and E. Candes and M. Grant (2010), Templates for convex cone problems with applications to sparse signal recovery, http://cvxr.com/tfocs