10-725/36-725: Convex Optimization **Fall 2015**

Lecture 11: October 5

Lecturer: Ryan Tibshirani *Scribes: Achal Dave, Anirudh Vemula, Vishal Dugar*

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.*

11.1 Lagrangian

Consider any general minimization problem

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

Let's define the Lagrangian, introducing variables $u \in \mathbb{R}^m, v \in \mathbb{R}^r$, with $u \geq 0$.

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

It turns out that for any $u \geq 0$ and any v, we have that

$$L(x, u, v) = f(x) \sum_{i=1}^{m} u_i \underbrace{h_i(x)}_{\leq 0} + \sum_{j=1}^{r} v_j \underbrace{l_j(x)}_{=0} \leq f(x)$$

Thus, we can observe that the Lagrangian $L(x, u, v)$ is always a lower bound for the primal criterion $f(x)$ for any value of $u \geq 0$ and v. An example for this is shown in the figure 11.1.

And so, we have that if f^* be the primal optimal value and C is the primal feasible set, then

$$f^* \geq \min_{x \in C} L(x, u, v) \geq \min_x L(x, u, v) \triangleq g(u, v)$$

This $g(u, v)$ is the **Lagrange dual function**, and it provides a lower bound on the optimal value f^* for any dual feasible u, v (i.e. $u \geq 0$ and any v).

Generally, duality will provide us with a tight lower bound in the *convex* case, but this need not be the true in the *non-convex* case. One such example is shown in the figure 11.2, where the lower bound is not tight.

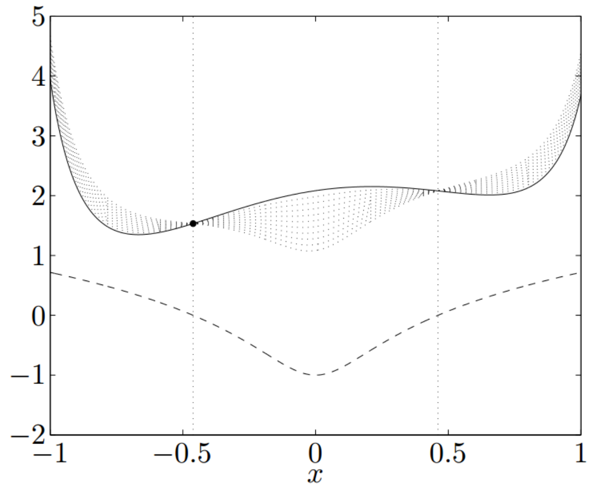

Figure 11.1: Solid line is f, dashed line is h. Each dotted line shows $L(x, u, v)$ for different choices of $u \geq 0$ and v. Note that the feasible set is $x \in [-0.46, 0.46]$

11.1.1 Example: Quadratic Program

Consider a quadratic program where $Q \succ 0$:

$$\begin{aligned} \min_x \quad & \frac{1}{2}x^T Q x + c^T x \\ \text{subject to } & Ax = b, x \geq 0 \end{aligned}$$

In this case, our Lagrangian is simply

$$L(x, u, v) = \frac{1}{2}x^T Q x + c^T x - u^T x + v^T (Ax - b)$$

To compute the dual function $g(u, v) = \min_x L(x, u, v)$, we minimize the Lagrangian above by taking the gradient with respect to x and setting it equal to zero, and we get that

$$\begin{aligned} x^* &= -Q^{-1}(c - u + A^T v) \\ \implies \min_x L(x, u, v) &= L(x^*, u, v) \\ &= \frac{1}{2}(c - u + A^T v)^T Q^{-1}(c - u + A^T v) - (c - u + A^T v)^T Q^{-1}(c - u + A^T v) - b^T v \\ &= -\frac{1}{2}(c - u + A^T v)^T Q^{-1}(c - u + A^T v) - b^T v \end{aligned}$$

What if, instead, we had the same QP as above, except $Q \succeq 0$ (i.e. Q is only positive *semi*-definite). Then, if we try to minimize the Lagrangian above by setting the gradient to 0, we get the following

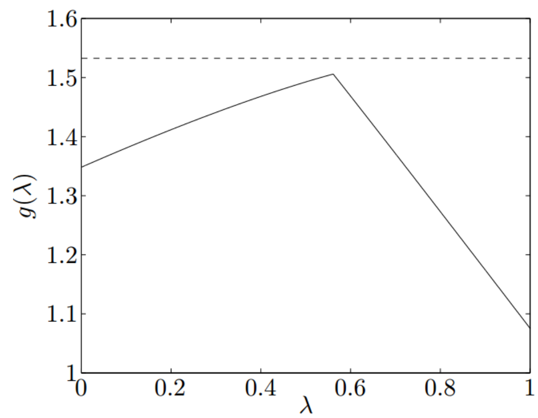

Figure 11.2: Dashed horizontal line is f^*, dual variable is λ and solid line shows $g(\lambda)$

constraint at the optimum:

$$Qx = -(c - u + A^T v) \tag{11.1}$$

Now, there are two cases:

(i) $c - u + A^T v \in \text{col}(Q)$. Then, we can use the pseudo-inverse (see below) $Q^\dagger$ of Q. This also implies that $P\left(c - u + A^T v\right) = 0$, where P is the projection matrix onto $\text{null}\,(Q)$.

(ii) $c - u + A^T v \notin \text{col}(Q)$, which implies that $c - u + A^T v$ is not orthogonal to the null space of Q. Then, let
$$c - u + A^T v = z_1 + z_2,$$
where $z_1 \in \text{col}(Q), z_2 \in \text{null}(Q), z_2 \neq 0$. But in this case, there is no x that satisfies eq. (11.1), and so there is no unique minimizer x^*. But $L(x, u, v)$ is quadratic in x, so if there is no minimizer of $L(x, u, v)$ in x, the minimum must be $L(x, u, v) = -\infty$.

So,

$$g(u, v) = \begin{cases} -\frac{1}{2}(c - u + A^T v)^T Q^\dagger (c - u + A^T v) - b^T v & \text{case (i)} \\ -\infty & \text{case (ii)} \end{cases}$$

11.1.1.1 Aside: Pseudo-inverse

For a general matrix $A \in \mathbb{R}^{n \times n}$, we can define the pseudo-inverse $A^\dagger$ in terms of its Singular Vector Decomposition. Using SVD, we can write A as

$$A = UDV^T$$

If A was invertible, we can directly invert the decomposition above:

$$A^{-1} = (UDV^T)^{-1} = (V^T)^{-1}D^{-1}U^{-1} = VD^{-1}U^T$$

where

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & \dots & 0 \\ 0 & \frac{1}{d_2} & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{d_n} \end{bmatrix}$$

If A is not invertible, we're going to see that for $k = \text{rank}(A)$, $d_{k+1} = d_{k+2} = \cdots = d_n = 0$. In this case, we can construct a pseudo-inverse ($D^\dagger$) of D as follows:

$$D^\dagger = \begin{bmatrix} \frac{1}{d_1} & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{d_2} & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \vdots & 0 & 0 & 0 \\ 0 & 0 & \dots & \frac{1}{d_k} & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & 0 \end{bmatrix}$$

And our pseudo-inverse, then, is

$$A^\dagger = VD^\dagger U^T$$

11.1.2 Example: Quadratic Program in 2D

In this example, we choose $f(x)$ to be quadratic in 2 variables, subject to $x \geq 0$. The dual function $g(u)$ is also quadratic in 2 variables, also subject to $u \geq 0$. In the figure 11.3, we can see that the dual function $g(u)$ provides a bound on f^* for every $u \geq 0$, and the largest bound $g(u)$ gives us turns out to be exactly f^*! In the future, we will see that this is not a coincidence and results from KKT conditions.

11.2 Lagrange Dual Problem

Given the primal problem

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

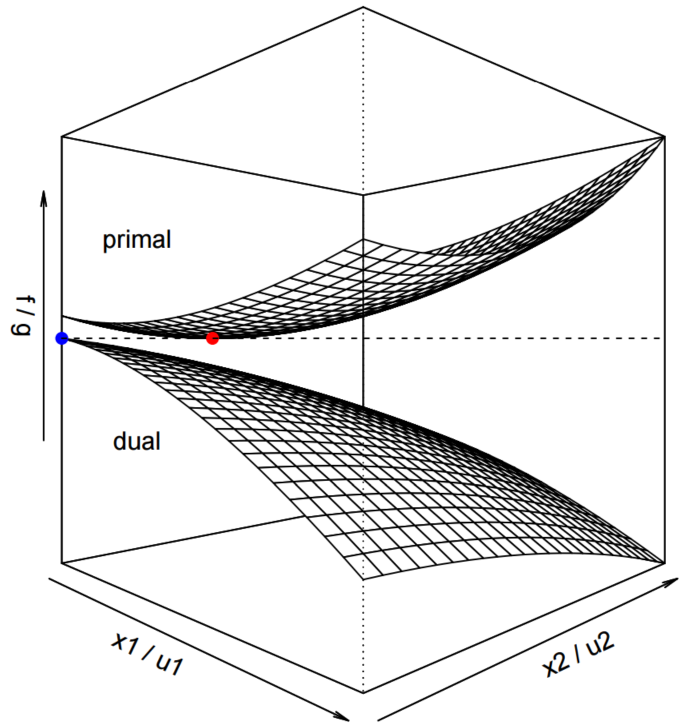

Figure 11.3: Blue dot is the optimal dual value and the red dot is the optimal primal value

We have shown that our constructed dual function $g(u, v)$ satisfies the property $f^* \geq g(u, v)$ for all $u \geq 0$ and v. Thus, we get the tightest lower bound on the optimal primal criterion f^* by maximizing $g(u, v)$ over all dual feasible u, v, yielding the Lagrange dual problem:

$$\begin{aligned} \max_{u,v} \quad & g(u, v) \\ \text{subject to } & u \geq 0 \end{aligned}$$

Note that, if the dual optimal value is g^*, then

$$f^* \geq g^*$$

This always holds (even if the primal problem is nonconvex) and is called the *weak duality* property.

11.2.1 Convexity of the dual

A very interesting property is that the dual problem is a convex optimization problem (even if the primal problem is non-convex) problem in u, v:

$$\begin{aligned} g(u,v) &= \min_x \{f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j l_j(x)\} \\ &= -\underbrace{\max_x \{-f(x) - \sum_{i=1}^m u_i h_i(x) - \sum_{j=1}^r v_j l_j(x)\}}_{\text{pointwise maximum of convex functions in } (u,v)} \end{aligned}$$

As can be observed from above, $g(u, v)$ can be expressed as the negative of pointwise maximum of convex functions in (u, v). Hence, g is concave in (u, v), and $u \geq 0$ is a convex constraint, hence the dual problem is a concave maximization problem (or a convex minimization problem if we consider $-g$)

So why don't we just always write down the dual problem and solve it, if it's convex? It turns out computing $g(u, v)$ is hard in and of itself, since it involves a maximization over x, especially for non-convex problems. In other words, we might not be able to write out $g(u, v)$ in the first place!

11.3 Strong Duality

We know that $f^* \geq g^*$, which is known as weak duality. In some problems, we see that $f^* = g^*$. This is known as **strong duality**.

Slater's condition: If the primal is a convex problem, and there exists at least one *strictly* feasible $x \in \mathbb{R}^n$, then strong duality holds.

That is, for a general convex primal problem,

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

where $h_i(x)$ is convex and $l_j(x)$ is affine, and a strictly feasible x is an x such that for every $i, h_i(x) < 0$ and for every $j, l_j(x) = 0$. This is a weak condition, and an important extension to Slater's condition maintains that strict inequalities *only* need to hold over $h_i(x)$ that are *not affine*.

11.3.1 Strong duality for Linear Programs

For linear programs

- The dual of the dual is the primal.

- Strong duality holds for the primal LP if it is feasible (refinement over Slater's conditions).
- Similarly, strong duality holds for the dual if it is feasible.
- Thus, strong duality holds for LPs, except when both primal and dual are infeasible.

11.3.2 Example: SVM dual

The SVM problem is as follows:

$$\begin{aligned} \min_{\beta,\beta_0,\xi} \quad & \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i \\ \text{subject to } & \xi_i \geq 0, i = 1,\dots,n \\ & y_i\left(x_i^T\beta + \beta_0\right) \geq 1 - \xi_i, i = 1,\dots,n \end{aligned}$$

We form the following Lagrangian with the dual variables $v, w \geq 0$:

$$L\left(\beta,\beta_0,\xi,v,w\right) = \frac{1}{2}\|\beta\|_2^2 + C\sum_{i=1}^n \xi_i - \sum_{i=1}^n v_i\xi_i + \sum_{i=1}^n w_i\left(1 - \xi_i - y_i\left(x_i^T\beta + \beta_0\right)\right)$$

Minimizing over β, β_0, ξ gives the Lagrange dual:

$$g(v,w) = \begin{cases} -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w & \text{if } w = C1 - v, w^T y = 0 \\ -\infty & \text{otherwise} \end{cases}$$

where $\tilde{X} = \text{diag}\left(y\right)X$. The SVM dual problem can thus be written as:

$$\begin{aligned} \max_{w} \quad & -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w \\ \text{subject to } & 0 \leq w \leq C1, \;\; w^T y = 0 \end{aligned}$$

Since $w = 0$ is a feasible solution, Slater's conditions are satisfied and we have strong duality.

11.4 Duality Gap

The duality gap $f\left(x\right) - g\left(u,v\right)$ refers to the difference between the primal (f) and dual (g) criterion values for corresponding x, u, v. An important property of the duality gap is the following:

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

This implies that a zero duality gap indicates an optimal value for both the primal and the dual. In practice, this provides a stopping criterion; if $f\left(x\right) - g\left(u,v\right) \leq \epsilon$, then $f\left(x\right) - f^* \leq \epsilon$.