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

Lecture 12: October 3

Lecturer: Ryan Tibshirani *Scribes: Shubhranshu Shekhar, Xinkun Huang, Yuan Wang*

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

Shubhranshu Shekhar scribed from 12.1 to 12.4. Xiukun Huang scribed from 12.5 to 12.6. Yuan Wang scribed from 12.7 to 12.8.

In today's scribe, we will talk about duality in general convex optimization problems beyond Linear Programs. We will introduce Lagrangian dual function that would allow us to define a dual problem (which is convex) for general optimization problems including non-convex problems, and the solution to which provides a lower-bound on the solution to the primal problem.

12.1 Review: Duality in Linear Programs

A linear program can be transformed into a dual problem by introducing a dual variable for each constraint as: Given $c \in \mathbb{R}^n, A \in \mathbb{R}^{> \times \ltimes}, b \in \mathbb{R}^m, G \in \mathbb{R}^{r \times n}, h \in \mathbb{R}^r$:

$$\begin{aligned} \min_x \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & Gx \le h \end{aligned}$$

Primal LP

$$\begin{aligned} \max_{u,b} \quad & -b^T u - h^T v \\ \text{subject to} \quad & -A^T u - G^T v = c \\ & v \ge 0 \end{aligned}$$

Dual LP

Dual LP has few interesting features such as - (i) is an LP maximization, (ii) a dual variable is introduced for each constraint in primal (iii) solution to max dual $\le$ solution of min primal (tightest possible bound by construction).

Explanation: for any u and $v \ge 0$, and x primal feasible,

$$\underbrace{u^T(Ax - b)}_{=0} + v^T \overbrace{(Gx - h)}^{\le 0} \le 0$$
$$\implies (-A^T u - G^T v)^T x \ge -b^T u - h^T v$$

So if $c = -A^T u - G^T v$, we get a bound on primal optimal value. Thus the tightest possible bound is obtained by maximizing $-b^T u - h^T v$.

Note: If the primal criterion would have been quadratic, then by construction we wouldn't be able to construct dual.

12.2 Alternate perspective on LP duality

This alternate perspective reproduces the same dual, but is actually completely general and applies to arbitrary optimization problems (even non-convex ones). Consider the Primal LP defined above. For any u and $v \geq 0$, and $x \in R^n$ primal feasible

$$c^T x \geq c^T x + u^T(Ax - b) + v^T(Gx - h) := L(x, u, v)$$

So if C denotes primal feasible set, f^* primal optimal value, then we have

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

Here, $g(u, v) = \begin{cases} -b^T u - h^T v & \text{if} \quad c = -A^T u - G^T v \\ -\infty & \text{otherwise} \end{cases}$

In other words, for any u and $v \geq 0$, $g(u, v)$ is the lower bound on f^*. To find the tightest lower bound among all, we solve the Dual LP defined above.

12.3 The Lagrangian

Consider a general minimization problem

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

We introduce new variables $u \in R^m$, and $v \in R^r$, with $u \geq 0$, and define the Lagrangian to be

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

(implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$).

Observe that for a feasible x, and for any $u \geq 0$ and v,

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

This property is illustrated in the Fig 12.1

12.4 Lagrange Dual Function

Let C denote primal feasible set, f^* denote primal optimal value. Minimizing $L(x, u, v)$ over all xR^n gives a lower bound on f^*:

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

Similar to LP dual function, we call $g(u, v)$ Lagrangian dual function, and it gives a lower bound on f^* for any $u \geq 0$ and v, called dual feasible u, v. This is illustrated in Fig It is to note that the relation holds, in general, for any optimization problem.

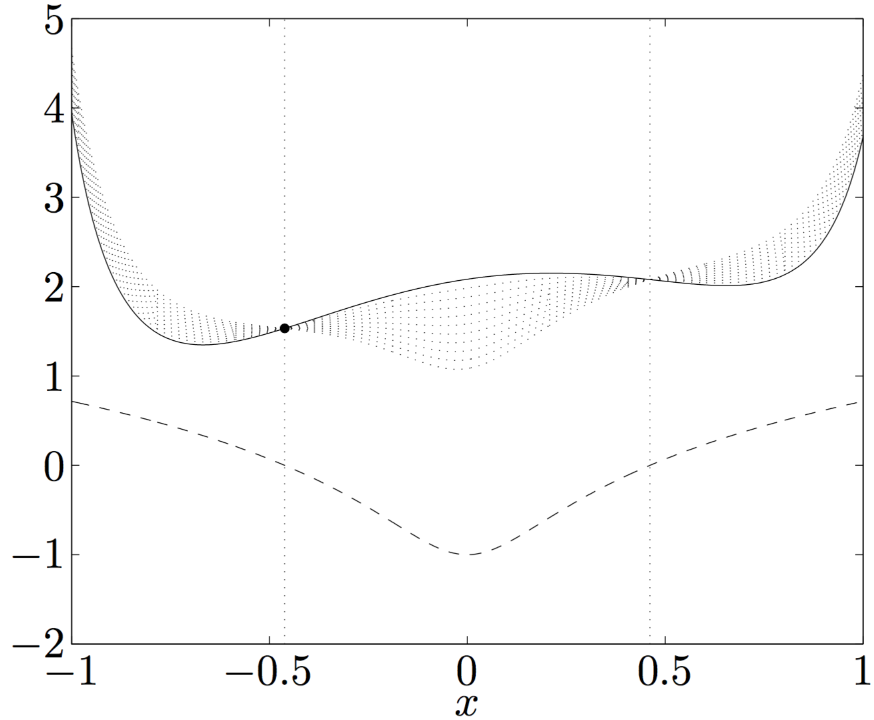

Figure 12.1: A one-dimensional optimization problem, with objective $f(x)$ (solid line). The constraint $h(x) \leq 0$ shown as dashed line/curve, thus the feasible range of x is approximately $[0.46, 0.46]$. Each dotted line/curve represents a Lagrangian $L(x, u)$ for different choices of $u \geq 0$. Note that these lie below the solid line within the feasible region.

12.5 Lagrange Dual Problem

Consider quadratic program:

$$\begin{aligned} &\min_x \frac{1}{2} x^T Q x + c^T x \\ &s.t. Ax = b, x \geq 0 \end{aligned}$$

when $Q \succ 0$. Lagrangian:

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

Because we know:

$$\begin{aligned} Qx &= -(c - u + A^T v) \\ x &= -Q^{-1}(c - u + A^T v) \end{aligned}$$

Put this x back into Lagrangian.

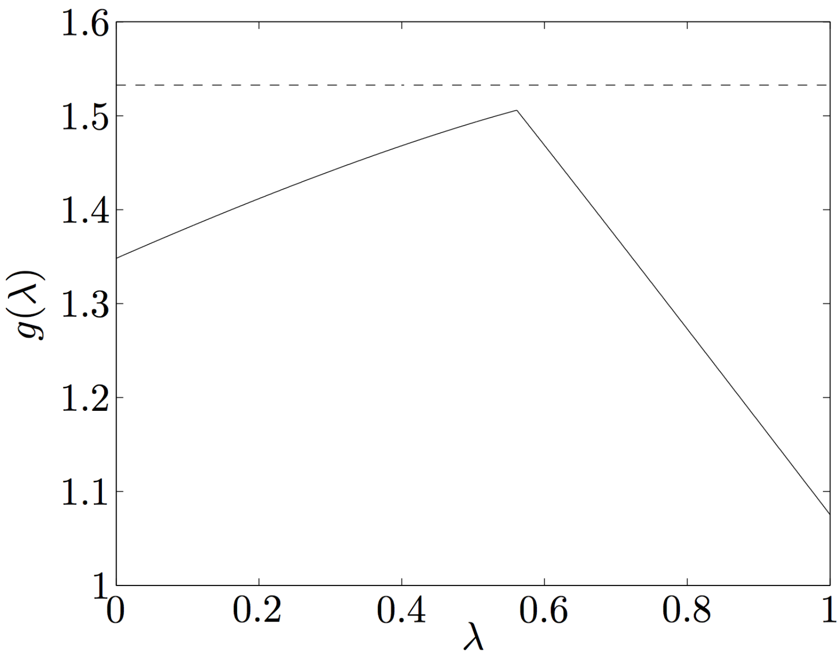

Figure 12.2: This figure shows the Lagrangian dual function $g(\lambda)$ (solid line) as a function of the dual variable λ (our u variable). The optimal value of the primal objective function is shown by the dashed line. Note that actual optimal value is > Lagrange optimal because, in this case, primal is non-convex.

$$\begin{aligned} g(u,v) &= \min_x L(x,u,v) \\ &= \frac{1}{2}(c-u+A^Tv)^TQ^{-T}QQ^{-1}(c-u+A^Tv) - (c-u+A^Tv)^TQ^{-1}(c-u+A^Tv) - v^Tb \\ &= -\frac{1}{2}(c-u+A^Tv)^TQ^{-1}(c-u+A^Tv) - b^Tv \end{aligned}$$

For any $u \geq 0$ and any v, this is lower a bound on primal optimal value f^*.

Same problem:

$$\begin{aligned} &\min_x \frac{1}{2}x^TQx + c^Tx \\ &s.t. Ax = b, x \geq 0 \end{aligned}$$

when $Q \succeq 0$. Lagrangian:

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

Similar as before, we know:

$$Qx = -(c-u+A^Tv)$$

If $-(c-u+A^Tv) \notin col(Q)$, then $L(x,u,v)$ may be negative infinity. If $-(c-u+A^Tv) \in col(Q)$, then $x = -Q^+(c-uA^Tv)$ where Q^+ is the generalized inverse of Q.

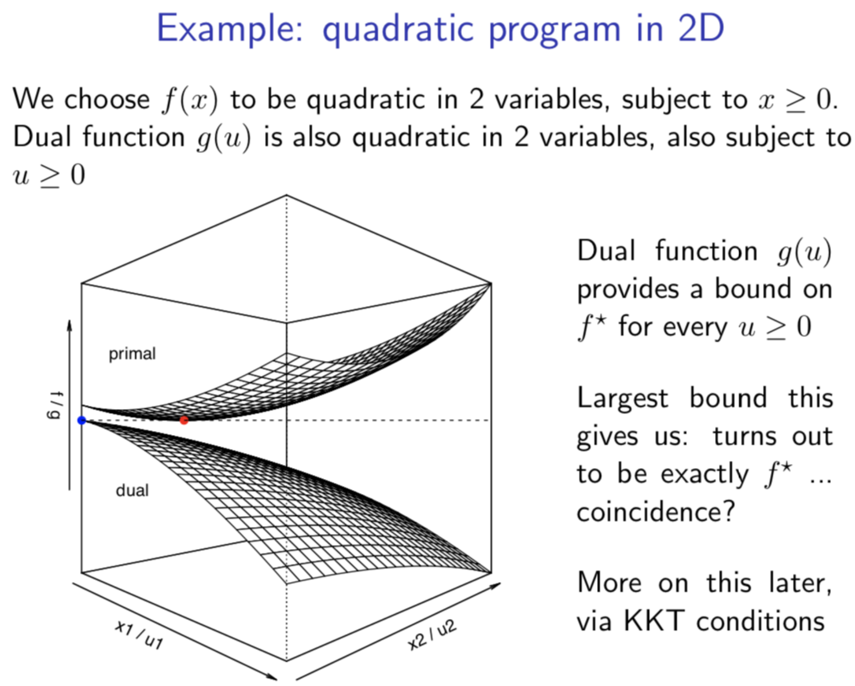

Figure 12.3: Slide page 10

The blue point in figure 12.3 is g^* which is on the bound. This is called complementary slackness which will be covered in KKT.

12.6 Weak Duality

Weak duality is that: if primal optimal value is f^*, and dual optimal value is g^*, then:

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

Note that this always holds even if the primal problem is nonconvex. There is an example on the slide page 13 and page 14.

12.7 Strong Duality

When you construct a dual where the primal criterion matches the dual criterion:

$$f^* = g^*$$

This means that the best bound that can be obtained through the Lagrange dual function is tight. Strong duality does not, in general, hold. One of the conditions where the strong duality holds is,
Slater's condition: if the primal is a convex problem (i.e., f and $h_1, \ldots, h_m$ are convex, and $l_1, \ldots, l_r$

are affine), and there exists at least one strictly feasible $x \in \mathbb{R}^n$, meaning

$$h_1(x) < 0, \dots, h_m(x) < 0 \text{ and } l_1(x) = 0, \dots, l_r(x) = 0$$

that strong duality holds (Proof refers to Boyd & V). An important refinement: strict inequalities only need to hold over functions hi that are not affine.

12.7.1 Slater's condition on LP cases

Fact: The dual of the dual problems is the primal problems. This is true for a lot of convex problems. (The sufficient condition is convexity, closeness of the constraints functions. This will be discussed later.)

Apply Slater's condition to P: if the primal problem is feasible, then $f^* = g^*$.
Apply Slater's condition to D: if the dual is feasible then $g^* = f^*$.
Put together the two statements from above, we get strong duality in linear programs. Strong duality fails only when BOTH primal and dual are infeasible (which would be hard to construct).

12.7.2 Example: SVM dual

The SVM problem

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

Introducing the dual variables, v, $w \geq 0$, we form the Lagrangian:

$$\begin{aligned} L(\beta_0,\beta,\xi,v,w) &= \frac{1}{2}\|\beta\|_2^2 + C\sum_{n=1}^{n}\xi_i - \sum_{i=1}^{n} v\xi_i + \sum_{n=1}^{n} w_i(1 - \xi_i - y_i(x_i^T\beta + \beta_0)) \\ &= \frac{1}{2}\beta^T\beta + (C\mathbf{1} - v - w)^T\xi + (\tilde{X}^T w)^T\beta + y^T w\beta_0 \end{aligned}$$

$$\begin{aligned} \xi_i \geq 0 & \quad v_i \\ y_i(x_i^T\beta + \beta_0) \geq 1 - \xi_i & \quad w_i \end{aligned}$$

$$\min L(\beta_0,\beta,\xi,v,w) = \begin{cases} -\frac{1}{2}w\tilde{X}\tilde{X}^T w + \mathbf{1}^T w & \text{if } y^T w = 0,\ w = C\mathbf{1} - v \\ -\infty & \end{cases}$$

Therefore the dual problem is: $\tilde{X}\tilde{X}^T$ is the kernal matrix of the kernel SVM.

A more practical solution: learn primal solution from dual solution. WIll revisit SVM when learning about KKT.

12.8 Duality Gap

Given primal feasible x and dual feasible u, v, duality gap is the quantity $f(x) - g(u, v)$. We also have that $f(x) - f^* \leq f(x) - g(u, v)$. So if the duality gap is zero, then x is primal optimal (and similarly, u, v are dual optimal). It provides an upper bound on distance from the optimality, which can provide an upper bounds for stopping criteria.