10-725/36-725: Convex Optimization **Fall 2019**

Lecture 11: October 2

Lecturer: Ryan Tibshirani *Scribes: Mengdi Xu, Pengji Zhang, Ian Chiu*

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

The topic for this scribed noted is Duality in General Programs. We will introduce Lagrange dual function which can be applied to arbitrary optimization problems.

11.1 Review: Duality in Linear Programs

Given $c \in \mathbb{R}^n, A \in \mathbb{R}^{m,n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{r,n}$, $h \in \mathbb{R}^r$:

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

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

11.1.1 Explanation 1

The dual problem of a linear problem is generated by introducing dual variables to each constraints. By multiplying the dual variable with corresponding constraint and then adding them together (equalities and inequalities), we can rearrange the results to be equal to the primal problem and then get the bound.

For any u and $v \le 0$, and x primal feasible.

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

So if $c = -A^T u - G^T v$, we can get a bound on primal optimal value. Note that the primal problem is a minimization problem and its tightest bound is obtained by maximizing the dual problem.

11.1.2 Explanation 2

Another explanation is using Lagrange dual function which is completely general and applied to arbitrary optimization problems (including non-convex problems).
For any u and $v \ge 0$, and x primal feasible

$$c^T x \ge c^T x + u^T \overbrace{(Ax - b)}^{=0} + v^T \overbrace{(Gx - h)}^{\le 0} := L(x, u, v)$$

So if C denotes primal feasible set, $f^\star$ primal optimal value, then for any u and $v \geq 0$,

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

Note that

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

Note that $L(x,u,v) = (c + A^T u + G^T v)^T x - b^T u - h^T v$ is linear w.r.t. x. When $c \neq -A^T u - G^T v$, there exists x such that L approaches negative infinity.
In other words, $g(u,v)$ is a lower bound on $f^\star$ for any u and $v \geq 0$

11.2 Lagrangian

Consider general minimization problem (no need to be convex)

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

The Lagrangian is defined as

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

where $u \in \mathbb{R}^m \geq 0$, $v \in \mathbb{R}^r$. Note that if $u \leq 0$, $L(x,u,v) \to -\infty$.
An important property is that for any $u \geq 0$ and v, $f(x) \geq L(x,u,v)$ at each feasible x.
The relationship of f and $L(x,u,v)$ is not determined outside the feasible set.

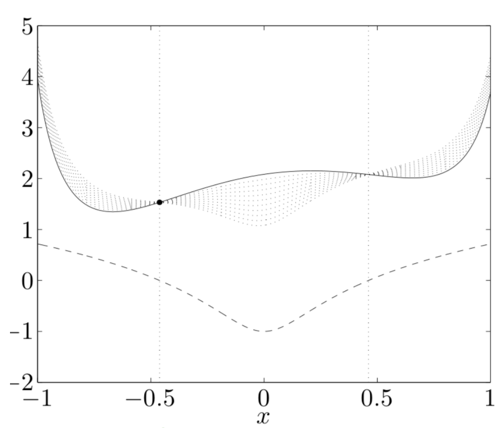


Figure 11.1: A one dimension optimization problem. Solid line is f. Dashed line is h. Each dotted line shows $L(x,u,v)$ for different choice of $u \geq 0$.

11.2.1 Lagrange dual function

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

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

We call $g(u,v)$ the Lagrange dual function, and it gives a lower bound on $f^\star$ for any $u \geq 0$ and v, called dual feasible u and v.

For $f^\star \geq \min_{x \in C} L(x,u,v)$, we can get strictly inequality if nonconvex problems.
For $\min_{x \in C} L(x,u,v) \geq \min_x L(x,u,v)$, when $u = 0$, $g(u,v)$ gives a tight lower bound. When $u \neq 0$, in general, duality will not be tight.

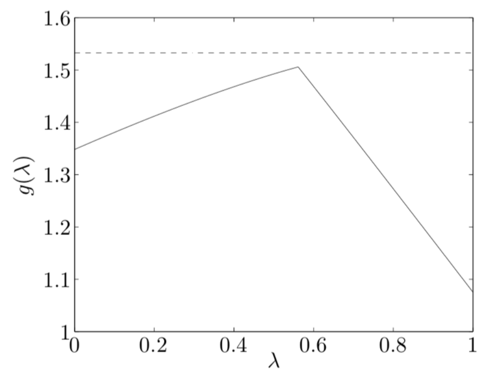


Figure 11.2: Dashed horizontal line is $f^\star$. Dual variable is λ

11.2.2 Example: Quadratic Program

Consider quadratic program where $Q \succ 0$:

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

The Lagrangian in this case is:

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

To get the Lagrange dual function of L, we need to take the gradient $\nabla_x L$ and set it to 0. The minimizer is $x^\star = -Q^{-1}(c - u + A^T v)$. By taking the minimizer back to the Langrangian, we get the Lagrange dual function:

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

For any $u \leq 0$ and any v, this lower bounds primal optimal value $f^\star$.

For the same problem but with $Q \succeq 0$

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

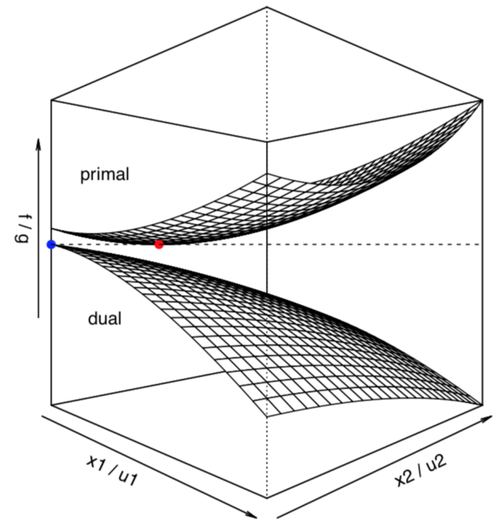


Figure 11.3: $f(x)$ here is quadratic with 2 variables, subject to $x \geq 0$. Dual function $g(u)$ is also quadratic in 2 variables, subject to $u \geq 0$. Dual function $g(u)$ provides a bound on $f^\star$ for every $u \geq 0$. The largest value of $g(u)$ is exactly $f^\star$ if $f^\star$ obtained interior.

The Lagrangian:

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

Follow the same procedure, we have $Qx = -(c - u + A^T v)$
If $(c - u + A^T v) \in \text{col}(Q)$ which is equavalent to $(c - u + A^T v) \perp \text{null}(Q)$, we can use the generalized inverse Q^+ of Q to get the minimizer.
Otherwise, there is no such x qualified and so there is no unique minimizer $x^\star$. Note that $L(x, u, v)$ is quadratic in x, hence the minimum should be $L(x, u, v) = -\infty$.

To sum up, Lagrange dual function:

$$g(u, v) = \begin{cases} -\frac{1}{2}(c - u + A^T v)^T Q^+ (c - u + A^T v) - b^T u & \text{if } c - u + A^T v \perp \text{null} Q \\ -\infty & \text{otherwise} \end{cases}$$

11.2.3 Lagrange dual problem

Given a primal problem

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

we have $f^* \geq g(u, v)$ for all $u \geq 0$ and v. Therefore, we can get a best lower bound for f^*—maximize $g(u, v)$ over feasible u, v, yielding the Lagrange dual problem:

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

A key property of the primal and dual problems is call "weak duality":

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

where f^* and g^* are the optimal value for the primal and dual problems respectively. Note that this property always holds for any primal and dual problems, even if the primal problem is nonconvex.

A second key property is that the dual problem is always a *convex optimization* problem (a concave maximization problem as written), even if the primal problem is nonconvex. This is because by definition

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

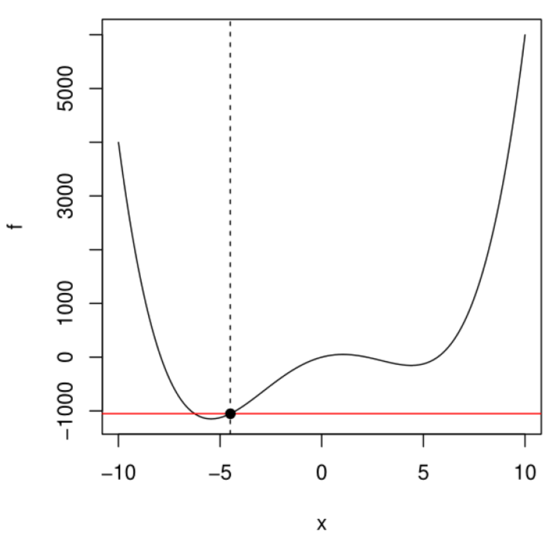
which is a concave function. Furthermore, the constraint $u \geq 0$ is convex, so the dual problem is a concave maximization problem.

11.2.4 Example: non-convex quartic minimization

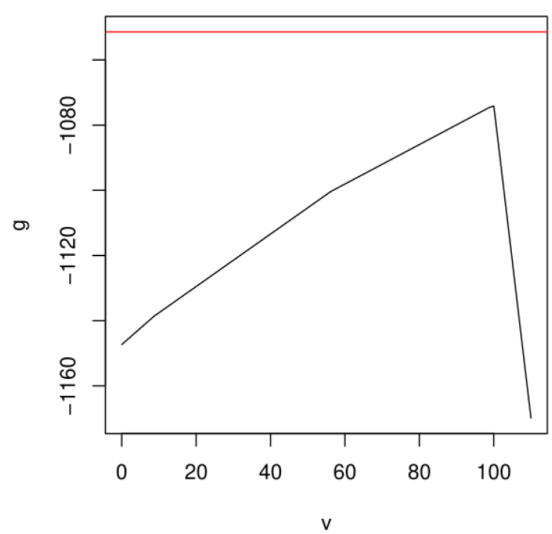
Consider a problem

$$\begin{aligned} \min_x \quad & x^4 - 50x^2 + 100x \\ \text{subject to} \quad & x \geq -4.5, \end{aligned}$$

where the objective $f(x) = x^4 - 50x^2 + 100x$ is not convex (Figure 11.4a).



(a) Primal objective function.



(b) Dual objective function.

Figure 11.4: Primal and dual objective functions of the quartic minimization problem.

The dual function $g(u)$ can be derived explicitly using the closed-form equation for roots of a cubic equation, and it is

$$g(u) = \min_{i=1,2,3} \Big\{ F_i^4(u) - 50F_i^2(u) + 100F_i(u) \Big\},$$

where

$$\begin{aligned} F_i(u) =& \frac{-a_i}{12 \cdot 2^{1/3}} \Big(432(100-u) - \big(432^2(100-u)^2 - 4 \cdot 1200^3\big)^{1/2} \Big)^{1/3} \\ &- 100 \cdot 2^{1/3} \frac{1}{\Big(432(100-u) - \big(432^2(100-u)^2 - 4 \cdot 1200^3\big)^{1/2} \Big)^{1/3}}, \quad i = 1,2,3 \\ & a_1 = 1, a_2 = \big(-1 + i\sqrt{3}\big)/2, a_3 = \big(-1 - i\sqrt{3}\big)/2. \end{aligned}$$

It is very hard to tell if $g(u)$ is concave without the context of duality. However, here we know $g(u)$ is the dual function of the primal problem, so we know that $g(u)$ must be concave (Figure 11.4b).

11.3 Strong Duality

Recall that by weak duality, we always have the condition that $f^* \geq g^*$. However, in some problems we will have $f^* = g^*$, which is what we call **strong duality**

Slater's condition: if the primal is a convex, where f and h_i all convex for $i \in [1, m]$ and ℓ_j affine $j \in [1, r]$, then there exists at least one strictly feasible $x \in \mathbb{R}^n$, then strong duality holds. In other words, the condition is that $h_i(x) < 0, \ell_j(x) = 0$ for all $i \in [1, m], j \in [1, r]$.

In fact, we actually observe that the only strict inequalities are needed for the non-affine h_i.

11.3.1 Example: support vector machine dual

Given $y \in \{-1, 1\}^n$, $X \in \mathbb{R}^{n \times p}$ with rows $x_1, ..., x_n$. Recall the support vector machine problem:

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

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

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

Minimizing over β, β_0, ξ gives Lagrangian:

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

for $\tilde{X} = \text{diag}(y)X$. Thus SVM, eliminating slack variable v:

$$\max_{w} -\frac{1}{2}w^T\tilde{X}\tilde{X}^T w + 1^T w$$
$$\text{subject to} \;\; 0 \le w \le C1, \;\; w^T y = 0$$

Check: Slater's condition is satisfied, and we have strong duality. Further, from study of SVMs, might recall that at optimality

$$\beta = \tilde{X}^T w$$

This is not a coincidence, as we will see via KKT conditions.

11.4 Duality Gap

Given primal x and dual feasible u, v, the quantity

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

is called the **duality gap** between x and u, v. Note that

$$f(x) - f^* \le f(x) - g(u, v)$$

so if the duality gap is zero, then x is primal optimal and u, v are dual optimal.

From an algorithmic viewpoint, it provides a stopping criterion. If $f(x) - g(u, v) \le \epsilon$, then we are guaranteed that $f(x) - f^* \le \epsilon$.