10-725: Optimization **Fall 2012**

Lecture 15: October 16

Lecturer: Geoff Gordon/Ryan Tibshirani *Scribes: Chunlei Liu,Seth Flaxman,Junier Oliva*

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

15.1 Part 1

15.1.1 Introduction to Duality

It may be useful to find a lower bound to the optimal value of a minimizing LP or convex problem. That is, one may wish to find a B such that $B \leq \min_{x \in C} f(x)$. For example, such a lower bound may be used as a stopping criterea if an algorithm solving the original problem (the primal problem) ever reaches or comes close to the lower bound.

In the case of LPs such a lower bound may be found by manipulating our inequality constraints, emposing new constraints, and now maximizing. In general, one may find a lower bound via a Lagrangian, which will be discussed below. In both cases, the task of maximizing the lowerbound is called the dual problem.

We will see that formulating the dual problem is often enlightening to non-obvious characteristics of the of the primal problem and its solution(s). A particular case of this is the equivalence of max-flow and min-cut problems, which is explored below.

15.1.2 Duality in Linear Programs

Suppose we had the following LP:

$$\begin{aligned} \min_{x,y} \quad & x+y && (15.1)\\ \text{subject to} \quad & x+y \geq 2 && (15.2)\\ & x,y \geq 0 && (15.3) \end{aligned}$$

Obviously, $B = 2$ serves as a lower bound; however, can we derive a lower bound in a more general case? Suppose we had the following LP:

$$\begin{aligned} \min_{x,y} \quad & px+qy && (15.4)\\ \text{subject to} \quad & x+y \geq 2 && (15.5)\\ & x \geq 0 && (15.6)\\ & y \geq 0 && (15.7) \end{aligned}$$

Then we may take multiply our contraints (15.5), (15.6), (15.7) by positive reals a, b, c respectively to arrive at:

$$ax + ay \geq 2a, \tag{15.8}$$
$$bx \geq 0, \tag{15.9}$$
$$cy \geq 0. \tag{15.10}$$

Adding yeilds:

$$(a+b)x + (a+c)y \geq 2a. \tag{15.11}$$

Thus, if we restrict a, b, c to be such that $a + b = p$ and $a + c = q$ we will have lower bound on our primal problem. In other words, our dual problem is:

$$\max_{a,b,c} \quad 2a \tag{15.12}$$
$$\text{subject to} \quad a + b = p \tag{15.13}$$
$$a + c = q \tag{15.14}$$
$$a, b, c \geq 0 \tag{15.15}$$

Note that number of dual variables will be the number of primal constraints. In general a primal LP will be of the form:

$$\min_{x\in\mathbb{R}^n} \quad c^{\mathrm{T}}x \tag{15.16}$$
$$\text{subject to} \quad Ax - b = 0 \tag{15.17}$$
$$Gx - h \leq 0 \tag{15.18}$$

where $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{r\times n}$, $h \in \mathbb{R}^r$. Using the same principles as with (15.4), we can manipulate our constraints to derive a lower bound. Consider multiplying (15.17) by any u and (15.18) by any $v \geq 0$ and adding these constraints, then for any feasible x one has:

$$u^{\mathrm{T}}(Ax - b) + v^{\mathrm{T}}(Gx - h) \leq 0 \quad \Longrightarrow \tag{15.19}$$
$$(-A^{\mathrm{T}}u - G^{\mathrm{T}}v)^{\mathrm{T}}x \geq -b^{\mathrm{T}}u - h^{\mathrm{T}}v. \tag{15.20}$$

Thus, $-b^{\mathrm{T}}u - h^{\mathrm{T}}v$ will serve as a lowerbound on our primal objective so long as $-A^{\mathrm{T}}u - G^{\mathrm{T}}v = c$. That is, the corresponding dual LP is:

$$\max_{u\in\mathbb{R}^m, v\in\mathbb{R}^r} \quad -b^{\mathrm{T}}u - h^{\mathrm{T}}v \tag{15.21}$$
$$\text{subject to} \quad -A^{\mathrm{T}}u - G^{\mathrm{T}}v = c \tag{15.22}$$
$$v \geq 0 \tag{15.23}$$

15.2 Part 2

15.2.1 Max flow / min cut

As a motivating example of duality, consider a railway network, used for moving goods between cities (nodes) along train tracks (edges) subject to capacity constraints. (For a historical discussion in the context of the Soviet railway network, see [S02]).

Formally, given a graph $G = (V, E)$ with a source node s and a sink (destination) node t, we wish the find the flow along each edge f_{ij} for an edge $(i, j) \in E$ so as to maximize the total flow from s to t subject to the following constraints:

- $f_{ij} \geq 0, \forall (i,j) \in E$
- $f_{ij} \leq c_{ij}, \forall (i,j) \in E$
- $\sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{kj}, k \in V\backslash\{s,t\}$

Conveniently, we can write this as a linear program:

$$\begin{aligned}
\max_{f\in R^{|E|}} \quad & \textstyle\sum_{(s,j)\in E} f_{sj} & & (15.24)\\
\text{subject to} \quad & f_{ij} \geq 0, & \forall (i,j) \in E \quad & (15.25)\\
& f_{ij} \leq c_{ij}, & \forall (i,j) \in E \quad & (15.26)\\
& \textstyle\sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{kj}, & \forall k \in V\backslash\{s,t\} \quad & (15.27)
\end{aligned}$$

Now we could solve this LP using the simplex method, but instead we'll use duality. Since we're maximizing flow, our dual problem will be to construct an upper bound on our primal, and we'll be minimizing this dual. (Note that this is the opposite of when we start with a minimization problem as is our usual practice with convex problems. Nothing fundamental changes for concave problems, as we'll see.) We'll create dual variables $a_{ij} \geq 0$ for equation 15.28, $b_{ij} \geq 0$ for equation 15.29, and x_k for equation 15.30. (We introduce one variable per constraint, so e.g. equation 15.30 says $\forall k \in V\backslash\{s,t\}$, thus we introduce variables x_k, one for each constraint in equation 15.30, i.e. $\forall k \in V\backslash\{s,t\}$). We multiply the constraints by these objectives and make all inequalities $\leq$:

$$-a_{ij} f_{ij} \leq 0, \quad \forall (i,j) \in E \tag{15.28}$$

$$b_{ij}(f_{ij} - c_{ij}) \leq 0, \quad \forall (i,j) \in E \tag{15.29}$$

$$x_k \left(\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} \right) = 0, \quad \forall k \in V\backslash\{s,t\} \tag{15.30}$$

Now we can sum together these constraints:

$$\sum_{(i,j)\in E} (-a_{ij} f_{ij} + b_{ij}(f_{ij} - c_{ij})) + \sum_{k\in V\backslash\{s,t\}} x_k \left(\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj} \right) \leq 0 \tag{15.31}$$

$$\text{subject to} \;\; a_{ij}, b_{ij} \geq 0 \tag{15.32}$$

Recall that we're looking for an upper bound on 15.24 so we rearrange 15.31 as:

$$\sum_{(i,j)\in E} M_{ij} f_{ij} \leq \sum_{(i,j)\in E} b_{ij} c_{ij} \tag{15.33}$$

where M_{ij} gives the coefficient of f_{ij} which we can express for three separate cases:

- For edges from the source: $M_{sj} = -a_{sj} + b_{sj} + x_j, \;\; \forall j$
- For edges into the sink: $M_{it} = -a_{it} + b_{it} - x_i, \;\; \forall i$
- For internal edges: $M_{ij} = -a_{ij} + b_{ij} + x_j - x_i, \;\; \text{otherwise}$

Comparing 15.33 and 15.24 we see that 15.24 is only concerned with maximizing the flow out of the source (which makes sense), so we can find an upper bound on this objective using 15.33 if we set the appropriate M_{ij} coefficients to 0, i.e. we set $M_{ij} = 0$ for $i \neq s$, thus:

- $M_{sj} = 1, \quad \forall j$
- $M_{it} = 0, \quad \forall i$
- $M_{ij} = 0, \quad \text{otherwise}$

We could stop here and we'd have a perfectly good dual program which would be a valid upper bound for the primal optimal value. But further simplification is possible (and worthwhile, as we'll see). Considering the following:

$$M_{sj} = b_{sj} - a_{sj} + x_j = 1 \tag{15.34}$$
$$M_{it} = b_{it} - a_{it} - x_i = 0 \tag{15.35}$$
$$M_{ij} = b_{ij} - a_{ij} + x_j - x_i = 0 \tag{15.36}$$

We can treat the a_{ij} variables as slacks (recall that $a_{ij} \geq 0$) and turn equalities into inequalities:

$$b_{sj} + x_j \geq 1 \tag{15.37}$$
$$b_{it} - x_i \geq 0 \tag{15.38}$$
$$b_{ij} + x_j - x_i \geq 0 \tag{15.39}$$

If we further restrict $x_s = 1$ and $x_t = 0$ (and assuming there is no edge (s, t) directly from source to sink) then we can summarize in one line:

$$b_{ij} + x_j - x_i \geq 0, \quad \forall (i, j) \in E \tag{15.40}$$

Putting this all together gives us the dual problem:

$$\min_{b \in \mathcal{R}^{|E|}} \quad \textstyle\sum_{(i,j) \in E} b_{ij} c_{ij} \tag{15.41}$$
$$\text{subject to} \quad b_{ij} + x_j - x_i \geq 0, \quad \forall (i, j) \in E \tag{15.42}$$
$$b \geq 0, x_s = 1, x_t = 0 \tag{15.43}$$

Now we can think about what this program says. Consider the special case in which each variable x_i (of which there is one for each node) is a binary indicator $\in \{0, 1\}$ and let A be the set of nodes for which $x_i = 1$ and B be the set of nodes for which $x_i = 0$. Immediately we see that $s \in A$ and $t \in B$. Furthermore, if we rearrange the inequality constraint we get $b_{ij} \geq x_i - x_j$ meaning that $b_{ij} = 1$ for $i \in A, j \in B$ and $b_{ij} = 0$ otherwise (-1 is not possible because another constraint enforces non-negativity on the b's). So qualitatively, we're looking for a way to cut the graph into two sets A and B with $s \in A$ and $t \in B$ such that the capacity of the cut, defined as the sum of the edge capacities crossing the cut, is minimized.

What have we accomplished? Our dual program doesn't actual enforce $x_i \in \{0, 1\}$, so really it's a "relaxed" version of the min cut problem. This relaxed version is the dual of the max flow problem. Relaxing a constraint means that we've removed it from our program, so the optimal value will either stay the same or increase. Putting this all together we see that: max flow (primal) optimal value $\leq$ optimal for (relaxed) dual problem $\leq$ min cut (binary) problem.

The **max flow min cut theorem**, which we do not prove, says that the value of the max flow is equal to the value of the min cut (meaning that the relaxed version has the same optimal value as the min cut problem). This is a specific (and famous, and important) example of a more general idea, that of strong duality: the primal LP and the dual LP have the same optimal values (though not necessarily at the same values of the variables).

If you've seen the min cut problem and the max flow problem before it might seem natural to connect them, but a priori there's no reason to suspect they're dual (as the historical record suggests). Min cut has applications to image segmentation: if we choose a disimilarity metric for pixel values, we can segment a picture into two parts (corresponding to sets A and B above) by finding the min cut. Then we can repeat this recursively for each set. With an appropriate stopping rule, we'll end up with a segmented image.

15.2.2 Another perspective on LP duality

Let us inspect completely general primal/dual LPs. Here is the primal LP:

$$\min_{x\in\mathcal{R}^n} \quad c^T x \tag{15.44}$$
$$\text{subjec to} \quad Ax = b \tag{15.45}$$
$$Gx \le h \tag{15.46}$$

And here is the dual LP, where we've added variables u for the equality constraint and $v \ge 0$ for the inequality constraint:

$$\max_{u\in\mathcal{R}^m, v\in\mathcal{R}^r} \quad -b^T u - h^T v \tag{15.47}$$
$$\text{subject to} \quad -A^T u - G^T v = c \tag{15.48}$$
$$v \ge 0 \tag{15.49}$$

Another way to get this (which works even when the program isn't an LP!) is to lower bound the primal objective, $c^T x$ where we introduce u (unconstrained) and $v \ge 0$:

$$c^T x \ge c^T x + u^T(Ax - b) + v^T(Gx - h)$$

The RHS is so important, we'll give it a name:

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

For a primal optimal value $f^\star$, primal feasible set $C = \{x | Ax = b, Gx \le h\}$, any u, and $v \ge 0$, we see that (where the second inequality is because we've removed constraints so it's possible we'll find a smaller value when we minimize):

$$f^\star \ge \min_{x\in C} L(x, u, v) \ge \min_{x\in\mathcal{R}^n} L(x, u, v)$$

Once again, we'll give this last term a name:

$$g(u, v) := \min_{x\in\mathcal{R}^n} L(x, u, v)$$

Now we see that for any u, $v \ge 0$, $g(u, v)$ is a lower bound on $f^\star$.

Let us rearrange terms:

$$g(u, v) := \min_{x\in\mathcal{R}^n} L(x, u, v) = \min_{x\in\mathcal{R}^n} (A^T u + G^T v + c)^T x - b^T u - h^T v$$

Now we see that if $(A^T u + G^T v + c)^T \ne 0$ then $g(u, v) = -\infty$. Otherwise, $g(u, v) = -b^T u - h^T v$. So if we consider maximizing $g(u, v)$ (subject to $v \ge 0$) then assuming we can find a value other than ∞, we're really maximizing $-b^T u - h^T v$ which is exactly the dual objective from Equation 15.47. So we've found another way of characterizing the dual function, which we'll describe in more detail in the general (non-LP) case in the next section.

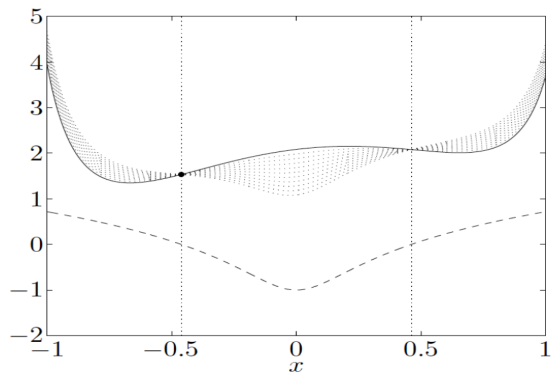

Figure 15.1: $L(x, u, v)$ vs $f(x)$ over all x ranges.

15.3 Part 3, Lagrangian

We consider the following general minimization problem :

$$\begin{array}{ll} \min_{x \in R^n} & f(x) \\ \text{subject to} & h_i(x) \le 0, i = 1, ..., m \\ & l_j(x) = 0, j = 1, ..., r \end{array}$$

Notice that, these functions are not necessarily convex, but we will pay special attention to the convex case. We then define the Lagrangian as the following:

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

where new variables $u \in R^m$ and $v \in R^r$ are introduced, with $u \ge 0$ (implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$). With the above definition, one realize an important property:

$$f(x) \ge L(x, u, v) \text{ at each feasible x} \tag{15.50}$$

This is because, with any feasible x, we have $h_i(x) \le 0$ and $l_j(x) = 0$ from the constraints, thus the second and third term in $L(x, u, v)$ is at most 0. This property can be illustrated in the following example (from B & V page 217) as shown in Fig. 15.1, where $f(x)$ is represented by the solid line, and the only inequality constraint $h(x)$ is shown in the dashed line, hence the feasible set is around $x \in [-0.46, 0.46]$. Each dotted line shows $L(x, u, v)$ for different choices of u, from which one can verify that $L(x, u, v) \le f(x)$ at each feasible x, while it's not necessarily true outside of the feasible set.

Now let C denote the primal feasible set, and x^* is the primal optimal value. Minimizing $L(x, u, v)$ over all $x \in R^n$ gives a lower bound:

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

where the first inequality sign is obtained because this: from Eqn. 15.50, we know $f(x^*) \ge L(x^*, u, v)$, and it is also true that $L(x^*, u, v) \ge \min_{x \in C} L(x, u, v)$. The second inequality sign is obtained is because: minimizing over the entire R^n is definitely to get a lower or equal value. $g(u, v)$ is called the Lagrange dual

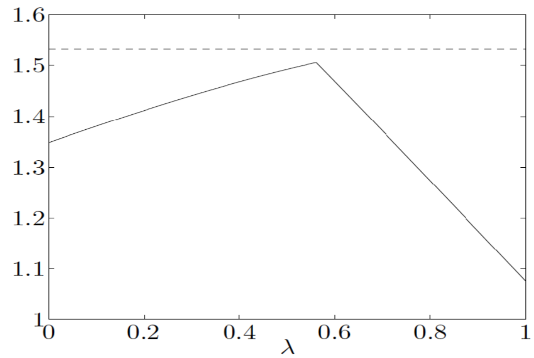

Figure 15.2: Dual function versus different λ.

function, and it gives a lower bound on $f(x^*)$ for any dual feasible $u \geq 0$ and v. Obviously if we want to get a tighter lower bound, we should maximize $g(u, v)$. Using the above example as shown in Fig. 15.1, we can also plot the dual function versus the primal optimal value in Fig. 15.2, where the dashed horizontal line is $f(x^*)$, λ is the dual variable (u in our case), and the solid line shows $g(\lambda)$. From Fig. 15.2, one can see that the dual function provides a lower bound for all feasible λ.

In the following we consider the quadratic program as another example, which is as the following:

$$\begin{array}{rl} \min_{x \in R^n} & \frac{1}{2} x^T Q x + c^T x \\ \text{subject to} & Ax = b, x \geq 0 \end{array}$$

where $Q \succ 0$ (i.e., strictly convex). The corresponding Lagrangian is then:

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

which achieves the minimum over x at $x = -Q^{-1}(c - u + A^T v)$. Bring this solution to the Lagrangian, we obtain the dual function as:

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

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

If in the above example of quadratic program, Q satisfies $Q \succeq 0$, we will have the dual function as the following:

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

where Q^+ denotes the generalized inverse of Q. So for any $u \geq 0$, v, and $c - u + A^v \perp \text{null}(Q)$, $g(u, v)$ is a nontrivial lower bound on f^*.

We continue to look at an example of the quadratic program in two dimensions as the following:

$$\begin{array}{rl} \min_{x \in R^2} & \frac{1}{2}(x_1, x_2) Q \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + c^T x \\ \text{subject to} & x \geq 0 \end{array}$$

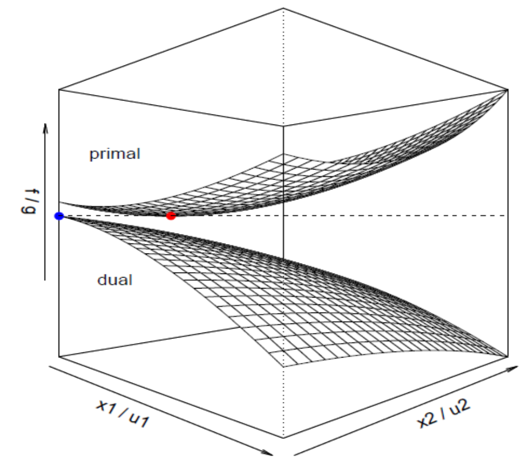

Figure 15.3: Primal and dual value in the same plot.

The dual function $g(u)$ is also quadratic in two variables as the following:

$$\begin{aligned} \min_{u\in R^2} \quad & -\frac{1}{2}(c-u)^T Q^{-1}(c-u) \\ \text{subject to} \quad & u \geq 0 \end{aligned}$$

The corresponding plot about the primal value and dual value is shown in Fig. 15.3. As one can see the dual function $g(u)$ provides a bound a on f^* for every $u \geq 0$. The largest bound turns to be exactly f^*. Notice that minimum and maximum doesn't touch at the same point. In fact, the dual function is maximized at $u = 0$ in this case. The reason is because of the property of complementary slackness, which will be clear in the following lecture.

Given the above examples, we can summarize the Lagrange dual problem. Given the primal problem

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

The constructed dual function $g(u, v)$ satisfies $f^* \geq g(u, v)$ for all $u \geq 0$ and v. Hence the best lower bound is given by maximizing $g(u, v)$ over all dual feasible u and v, yielding the Lagrange dual problem:

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

The key property is called weak duality, which states that, if dual optimal value is g^*, the

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

which always holds even if the primal problem is nonconvex. Another key property is that, the dual problem is always a convex optimization problem (as written above, it is a concave maximization problem). This is can be seen from the following proof:

$$\begin{aligned} g(u,v) &= \min_{x\in R^n}\{f(x)+\sum_{i=1}^{m}u_ih_i(x)+\sum_{j=1}^{r}v_jl_j(x)\} \\ &= -\max_{x\in R^n}\{-f(x)-\sum_{i=1}^{m}u_ih_i(x)-\sum_{j=1}^{r}v_jl_j(x)\} \end{aligned}$$

since the last equation is a pointwise maximum of convex functions in (u, v), $g(u, v)$ is concave in (u, v), and $u \geq 0$ is a convex constraint, hence the dual problem is a concave maximization problem.

References

[CW87] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Proceedings of the 19th ACM Symposium on Theory of Computing*, 1987, pp. 1–6.

[CW87] D. Coppersmith and S. Winograd, "Matrix multiplication via arithmetic progressions," *Proceedings of the 19th ACM Symposium on Theory of Computing*, 1987, pp. 1–6.

[S02] A. Schrijver, "On the history of the transportation and maximum flow problems," *Mathematical Programming*, 2002, pp. 437–445.