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

Lecture 10: Duality in Linear Programs

Lecturer: Ryan Tibshirani *Scribes: Jingkun Gao and Ying Zhang*

Administrivia: The project proposals have been received and will be divided among TAs. In future, each group will have a TA served as project mentor, or point of contact if you have any questions so that progress can be made smoothly.

We will start duality and optimality unit today, which will last for four lectures. The course content has been revised to extend the discussion in duality than originally planned, because duality is a subtle topic which is worthy going through slowly. Duality topics are our second theoretical unit and after which, we will cover the second-order algorithms.

Last lecture we explained the numerical linear algebra, which goes through basic flop counts for operations (please review the scribe from last lecture). This lecture's notes focus on the duality in linear programming, and give an example of the dual problem for maximum flow problem.

10.1 Lower Bounds in Linear Programs(LP)

Linear programs arises very frequently in optimization, especially in the cases of minimizing quadratic functions, e.g least square problems. Why is minimizing quadratic functions so fundamental in optimization? First order methods do a quadratic approximation at each step; the quadratic form is actually the identity. Second order methods use a quadratic approximation, but they use the Hessians of the criterion function instead of the identity. Because of this ubiquitous presence of quadratic approximations, linear programs are frequently observed for optimization.

We have seen two ways of solving linear systems, i.e. Cholesky decomposition and QR decomposition. When solving least square problem with matrix X, Cholesky acts on the matrix X^TX while QR acts on the matrix X instead. Compared with QR decomposition, Cholesky decomposition is cheaper and uses less memory, but it is more sensitive to numerical errors. The factor for flops and memories is approximately 2(twice as faster, uses half of the memory). If the quadratic to be minimized is poorly conditioned, i.e. with large largest eigenvalue and small smallest eigenvalues (or singular values), QR decomposition might be a better candidate, sacrificing some speed and memory to achieve stability. Now we are leaving the algorithm behind and dive into the topic of duality and optimality. Duality is really a fabulous topic: the arguments are really simple without complicated mathematical formalism, but the results are pretty powerful.

Suppose we want to find out the lower bound on the optimal value in our linear programming problem, $B \le \min f(x)$. We will introduce a simple LP example which might seem to be abstract, but it will provides insight of duality formulation:

$$\begin{aligned} \min_{x,y} \quad & x+y \\ \text{subject to} \quad & x+y \ge 2 \\ & x,y \ge 0. \end{aligned}$$

What is the lower bound for the optimal value of $x + y$? It is 2. For any feasible point, it has to satisfy the constraint $x + y \geq 2$. It may look too obvious to believe, but there are no other tricks.

Maybe the previous example seems to be too much like artificial problem, because coefficients for the constraints may not be as simple as the objective function. Let's try a different mutation:

$$\begin{aligned} \min_{x,y} \quad & x + 3y \\ \text{subject to} \quad & x + y \geq 2 \\ & x, y \geq 0. \end{aligned}$$

$$\begin{aligned} & x + y \geq 2 \\ + \quad & \phantom{x+{}} 2y \geq 0 \\ & = x + 3y \geq 2 \\ \text{Lower Bound} \quad & B = 2 \end{aligned}$$

The left column specifies the problem of finding the lower bound of the optimal value of $x+3y$. The method is to express the objective function in terms of the constraint functions. The right column shows the specific procedure for this $x+3y$. We note $x+3y = (x+y)+2\cdot y$, thus successfully expressed the objective function as the linear combination of constraints. Again, the lower bound of the optimal value is 2.

Here is a slightly more general question of modifying the objective function:

$$\begin{aligned} \min_{x,y} \quad & px + qy \\ \text{subject to} \quad & x + y \geq 2 \\ & x \geq 0 \\ & y \geq 0. \end{aligned}$$

$$\begin{aligned} & a + b = p \\ & a + c = q \\ & a, b, c \geq 0 \\ \text{Lower Bound} \quad & B = 2a, \text{for } \forall a, b, c \text{ satisfying the above} \end{aligned}$$

Now,in the object function $px + qy$, p and q are arbitrary constants. The same procedures is shown in the right column, multiplying each of the three constraints by constant of a,b,c respectively, yields,

$$ax + ay \geq 2a, bx \geq 0, cy \geq 0.$$

Add the three equations together,

$$(a + b)x + (a + c)y \geq 2a.$$

If $a + b = p, a + c = q$, then the lower bound of the objective function is obtained. Therefore the bound is $B = 2a$ for any a, b, c, such that $a + b = p, a + c = q$ and $a, b, c \geq 0$.

Now the natural question arises: what is the best lower bound, or biggest lower bound? Among all the values of a, b, c, what will be the tightest bound for the objective function? The best lower bound B is the maximum value of 2a, satisfying the constraints. This newly derived optimization problem of maximizing $2a$ is called the dual LP, while the original problem is called the primal LP.

$$\begin{aligned} A = \min_{x,y} \quad & px + qy \\ \text{subject to} \quad & x{+}y \geq 2 \\ & x{,}y \geq 0 \end{aligned}$$

Called Primal LP

$$\begin{aligned} B = \max_{a,b,c} \quad & 2a \\ \text{subject to} \quad & a{+}b = p \\ & a{+}c = q \\ & a, b{,}c \geq 0 \end{aligned}$$

Called Dual LP

By construction, the objective function $A \geq B$, and we will see in future, in well conditioned problems, the equality of the two objective functions is always observed, i.e. $A = B$. What if there are no values of a, b, c satisfying the dual problem constraints, i.e the constraint set is empty? For example, if $p < 0$, then a, b cannot both be non negative. The optimal value of the dual problem over an empty set is minus infinity(definition). Similarly, in the primal problem, if $p < 0$, x can be chosen to be larger and larger, then the optimal value of $px + qy$ is minus infinity as well. This condition of empty set did not break the rule $A \geq B$ as they are both minus infinity.

Let's try one more example in LP.

$$\begin{aligned} \min_{x,y} \quad & px + qy \\ \text{subject to} \quad & x \geq 0 \\ & y \leq 1 \\ & 3x + y = 2 \end{aligned}$$

$$\begin{aligned} \max_{a,b,c} \quad & 2c - b \\ \text{subject to} \quad & a + 3c = p \\ & -b + c = q \\ & a, b \geq 0 \end{aligned}$$

Applying the same procedure, multiplying each of the three constraints by constant of a,b,c respectively, yields,

$$ax \geq 0, -by \geq -1, 3cx + cy = 2c.$$

Reorganizing the inequalities, we have:

$$(a + 3c)x + (-b + c)y \geq -b + 2c,$$

which gives us the dual problem in the right column.

Generally, the number of dual variables is equal to the number of inequality constraints and equality constraints in the primal problem. The number of inequality constraints for dual variables is equal to the number of inequalities in the primal problem. In contrast, there is no constraints for the dual variables which correspond to the primal equality constraints, as the addition of any equality does not affect inequality relations. Through the examples above manually constructing dual problems, we should get an idea of where they come from. Now we consider the duality in general LPs. Also we will learn halfway through another way of deriving dual problem, called Lagrange duality, which is different from duality constructed manually in LPs.

10.2 Duality in General LPs

We are going to construct dual problem for general form LPs

A general LP problem can be written in the left column, where the objective function and constraints are dependent on the following variables, with dimensions specified: $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{r \times n}$, $h \in \mathbb{R}^r$:

$$\begin{aligned} A = \min_{x \in \mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & Gx \leq h \end{aligned}$$

Primal LP

$$\begin{aligned} B = \max_{u \in \mathbb{R}^m, v \in \mathbb{R}^r} \quad & -b^T u - h^T v \\ \text{subject to} \quad & -A^T u - G^T v = c \\ & v \geq 0 \end{aligned}$$

Dual LP

The construction procedures are as follows: for any u and $v \geq 0$, and x primal feasible,

$$u^T(Ax - b) + v^T(Gx - h) \leq 0, \text{i.e.,}$$

$$(-A^T u - G^T v)^T x \geq -b^T u - h^T v$$

Thus we get a lower bound on the primal optimal value being $c = -A^T u - G^T v$. Let's illustrate the different meaning of primal and dual problem through the nice example of max flow and min cut.

10.3 Max Flow and Min Cut

There is a nice history on max flow and min cut. It turns out that people cared about this problem during the wartime(WWI or WWII),e.g in Figure10.1. Initially, people did not know the two problems were equivalent and tried to solve them independently.

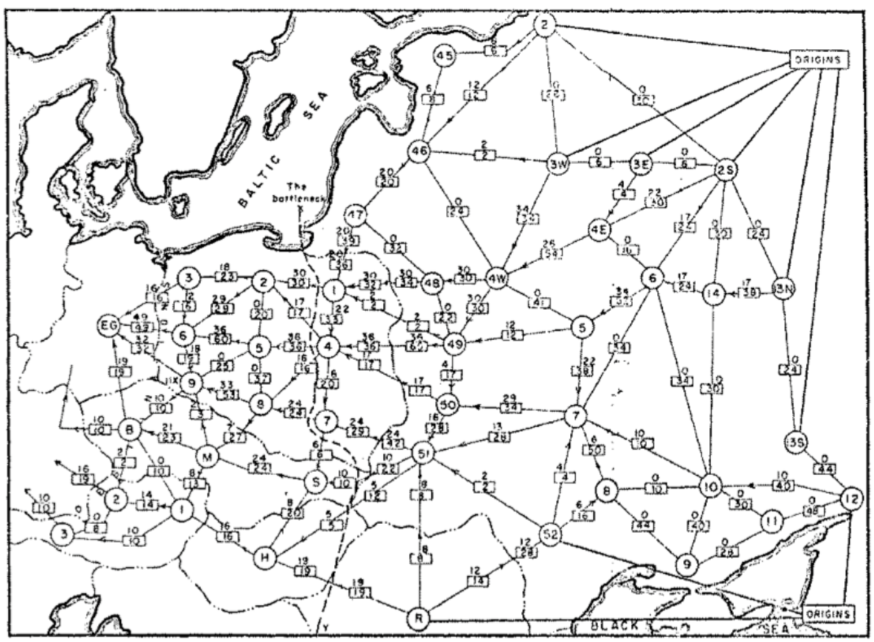

Figure 10.1: Soviet railway network (from Schrijver (2002), "On the history of transportation and maximum flow problems")

Here is a description of the maximum flow problem presented as a graph. Given a graph $G = (V, E)$,where V denotes nodes and E denotes edges.Node s denotes the source and node t is the sink as shown in Figure 10.2. The flow for any edge $(i, j) \in E$ is defined as f_{ij} to satisfy the following conditions (non-negative directed flow, capacity condition at each edge E and conservation at each internal node k,except source and sink):

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

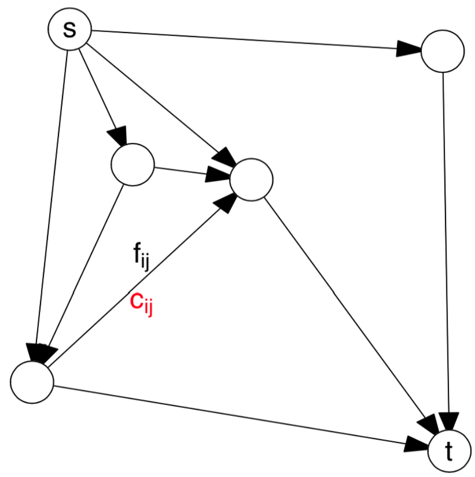

Figure 10.2: Max flow problem in graph representation

Max flow problem is to find flow that maximizes total value of the flow from s to t with the constraints. In other words, as an LP:

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

The derivation of the dual problem follows in steps: for inequality constraints, introduce dual variables a_{ij}, b_{ij}, and for the equality constraint, the dual variables to be multiplies are x_k. Summation of all these constraints yields,

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

for any a_{ij}, $b_{ij} \geq 0$, $(i,j)\in E$, and x_k, $k \in V\backslash\{s,t\}$
Rearranging the different terms gives,

$$\sum_{(i,j)\in E} M_{ij}(a,b,x) f_{ij} \leq \sum_{(i,j)\in E} b_{ij}c_{ij}$$

where $M_{ij}(a,b,x)$ collects terms multiplying f_{ij}.

To make LHS in previous inequality equal to primal objective, we need:

$$\begin{cases} M_{sj} = b_{sj} - a_{sj} + x_j & \text{want this} = 1 \\ M_{it} = b_{it} - a_{it} - x_i & \text{want this} = 0 \\ M_{ij} = b_{ij} - a_{ij} + x_j - x_i & \text{want this} = 0 \end{cases}$$

Thus, we have shown that upper bound condition for the primal optimal value, or the dual problem is then,

$$
\begin{aligned}
\min_{a,b,x} \quad & \sum_{(i,j)\in E} b_{ij}c_{ij} \\
\text{subject to} \quad & b_{sj} - a_{sj} + x_j = 1 \quad \text{for all } j \\
& b_{it} - a_{it} - x_i = 0 \quad \text{for all } i \\
& b_{ij} - a_{ij} + x_j - x_i = 0 \quad \text{for all } i, j \text{ where } i \neq s, j \neq t
\end{aligned}
$$

This dual problem can be further simplified in terms of inequalities by eliminating variables of a, which acts like a slack variable.Based on properties that$a \geq 0$,constraints can be rewritten as

$$
\begin{aligned}
b_{sj} + x_j = 1 + a_{sj} \geq 1 \quad & \text{for all } j \\
b_{it} - x_i = a_{it} \geq 0 \quad & \text{for all } i \\
b_{ij} + x_j - x_i = a_{ij} \geq 0 \quad & \text{for all } i, j \text{ where } i \neq s, j \neq t
\end{aligned}
$$

This is equivalent to the following dual problem. (Please stare at these equations to make sure the constraints are the same as the previous set.) This derivation is actually similar to the previous version for simple LP, except it's more detailed; but there is nothing intricate.

$$
\begin{aligned}
\min_{b\in\mathbb{R}^{|E|}, x\in\mathbb{R}^{|V|}} \quad & \sum_{(i,j)\in E} b_{ij}c_{ij} \\
\text{subject to} \quad & b_{ij} + x_j - x_i \geq 0 \text{ for all } (i,j) \in E \\
& b \geq 0,\ x_s = 1,\ x_t = 0
\end{aligned}
$$

Suppose that at the solution to the dual problem, it just so happened

$$x_i \in \{0, 1\} \text{ for all } i \in V,$$

which is kind of the relaxation of the previous dual problem. Call $A = \{i : x_i = 1\}$ and $B = \{i : x_i = 0\}$, note that $s \in A$ and $t \in B$.

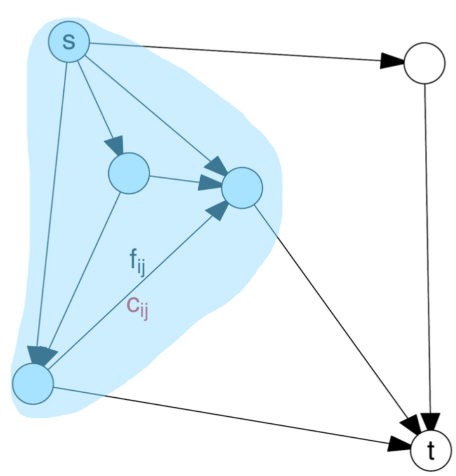

Figure 10.3: Min cut problem in graph representation

Compare Figure 10.3, then the constraints

$$b_{ij} \geq x_i - x_j \text{ for } (i,j) \in E,\, b \geq 0$$

imply that $b_{ij} = 1$ if $i \in A$ and $j \in B$, and 0 otherwise.

Moreover, the objective $\sum_{(i,j)\in E} b_{ij}c_{ij}$ is the capacity of cut defined by A, B. In other words, we've argued that the dual is the LP relaxation of the **min cut** problem:

$$\begin{aligned} \min_{b\in\mathbb{R}^{|E|}, x\in\mathbb{R}^{|V|}} \quad & \sum_{(i,j)\in E} b_{ij}c_{ij} \\ \text{subject to} \quad & b_{ij} \geq x_i - x_j \\ & b_{ij}, x_i, x_j \in \{0,1\} \text{ for all } (i,j) \end{aligned}$$

Therefore, from what we known so far:

$$\text{value of max flow} \;\leq\; \text{optimal value for LP relaxed min cut} \;\leq\; \text{capcity of min cut}$$

The second inequality says the dual problem is the relaxation of the capacity min cut problem which is actually an integer problem that don't know how to solve yet. There is a famous result, called **max flow min cut theorem**: value of max flow through a network is exactly the capacity of the min cut.

Hence in the above, all the inequalities turn out to be equalities.All the relaxations are tight; in particular, we get that the primal LP and dual LP have exactly the same optimal values, a phenomenon called **strong duality**. When we derive a dual and find out the bounds are actually tight, then this property is called strong duality.

How often does this happen? We will see in next class, the strong duality is more than just matching criterion values; the solutions for the dual problem can be expressed in terms of primal solution, vice versa.

10.4 Another Perspective on Duality

There is a problem with the previous derivation for duality in LP: how do we generalize this procedure to more general convex optimization problems, where constraints can be arbitrary convex functions where the objective function can not be expressed in terms of the linear combination of constraints. That's the roadblock in this particular perspective on duality. Fortunately, there is a completely equivalent take on the duality, which is called Lagrange duality.
Given $c \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $G \in \mathbb{R}^{r\times n}$, $h \in \mathbb{R}^r$:

$$\begin{aligned} A = \min_{x\in\mathbb{R}^n} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & Gx \leq h \end{aligned}$$

Primal LP

$$\begin{aligned} B = \max_{u\in\mathbb{R}^m, v\in\mathbb{R}^r} \quad & -b^T u - h^T v \\ \text{subject to} \quad & -A^T u - G^T v = c \\ & v \geq 0 \end{aligned}$$

Dual LP

There is a second explanation for this duality: for any u and $v \geq 0$, and x primal feasible, then the criterion,

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

Why is this true? Because $u^T(Ax-b)$ is always zero if x is primal feasible while $v^T(Gx-h)$ is always non-positive for feasible x. Actually we call the RHS $L(x,u,v)$ as Lagrangian, which is always the lower bound for the criterion, if x is primal feasible and u,v are also feasible($v \geq 0$, no constraints for u). So if C denotes primal feasible set,denote f^* as primal optimal value, then for any u and $v \geq 0$,

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

In other words, $g(u,v)$ is a lower bound on f^* for any u and $v \geq 0$, where $L(x,u,v)$ is called Lagrangian and $g(u,v)$ is called Lagrange dual function. The lagrangian can be rewritten as,

$$L(x,u,v) = (A^Tu + c + G^Tv)^T x - b^Tu - h^Tv$$

Note that

$$\begin{aligned} g(u,v) &= \min_{x} \quad L(x,u,v) \\ &= \begin{cases} -b^Tu - h^Tv & \text{if } c = -A^Tu - G^Tv \\ -\infty & \text{otherwise} \end{cases} \end{aligned}$$

Now we can maximize $g(u,v)$ over u and $v \geq 0$ to get the tightest bound, and this gives exactly the dual LP as before.

This last perspective is actually **completely general** and applies to arbitrary optimization problems (even nonconvex ones).

10.5 Matrix Games(not covered in class, slides are copied for completeness)

Setup: two players, J vs. R, and a payout matrix P:

R

		1	2	$\cdots$	n
	1	P_{11}	P_{12}	$\cdots$	P_{1n}
J	2	P_{21}	P_{22}	$\cdots$	P_{2n}
	$\cdots$				
	m	P_{m1}	P_{m2}	$\cdots$	P_{mn}

Game: if J chooses i and R chooses j, then J must pay R amount P_{ij} (don't feel bad for J–this can be positive or negative)

They use **mixed strategies**, i.e., each will first specify a probablity distribution, and then

$$x : \mathbb{P}(\text{J chooses } i) = x_i,\ i = 1, \cdots, m$$

$$y : \mathbb{P}(\text{R chooses } j) = y_j,\ j = 1, \cdots, n$$

The expected payout then, from J to R, is

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j P_{ij} = x^T P y$$

Now suppose that, because J is wiser, he will allow R to **know his strategy** x ahead of time. In this case, R will choose y to maximize x^TPy, which results in J paying off

$$\max\{x^TPy : y \geq 0, 1^Ty = 1\} = \max_{i=1,\cdots,n} (P^Tx)_i$$

J's best strategy is then to choose his distribution x according to

$$\begin{aligned} \min_{x \in \mathbb{R}^m} \quad & \max_{i=1,\cdots,n} (P^T x)_i \\ \text{subject to} \quad & x \geq 0, 1^T x = 1 \end{aligned}$$

In parallel universe, if R were somehow wiser than J, then he might allow J to know his strategy y beforehand.

By the same logic, R's best strategy is to choose his distribution y according to

$$\begin{aligned} \max_{y \in \mathbb{R}^n} \quad & \min_{j=1,\cdots,m} (P^T y)_j \\ \text{subject to} \quad & y \geq 0, 1^T y = 1 \end{aligned}$$

Call J's expected payout in first scenario f_1^*, and expected payout in second scenario f_2^*. Because it is clearly advantageous to know the other player's strategy, $f_1^* \geq f_2^*$.

But by **Von Neumman's minimax theorem**: we know that $f_1^* = f_2^*$, which may come as a surprise.

Recall first problem as an LP:

$$\begin{aligned} \max_{x \in \mathbb{R}^m, t \in \mathbb{R}} \quad & t \\ \text{subject to} \quad & x \geq 0, 1^T x = 1 \\ & P^T x \geq t \end{aligned}$$

Now form what we call the Lagranigian:

$$L(x, t, u, v, y) = t - u^T x + v(1 - 1^T x) + y^T (P^T x - t1)$$

and what we call the Lagrange dual function:

$$g(u, v, y) = \min_{x,t} L(x, t, u, v, y) = \begin{cases} v & \text{if } 1 - 1^T y = 0,\ Py - u - v1 = 0 \\ -\infty & \text{otherwise} \end{cases}$$

Hence dual problem, after eliminating slack variable u, is

$$\begin{aligned} \max_{y \in \mathbb{R}^n, v \in \mathbb{R}} \quad & v \\ \text{subject to} \quad & y \geq 0, 1^T y = 1 \\ & P^T y \geq v \end{aligned}$$

This is exactly the second problem, and therefore again we see that **strong duality** holds. So how often does strong duality hold? In LPs, as we'll see, strong duality holds unless both the primal and dual are infeasible.

References

- S. Boyd and L. Vandenberghe (2004), "Convex optimization, Chapter 5
- R. T. Rockafellar (1970), "Convex analysis, Chapters 2830