Duality in Linear Programs

Ryan Tibshirani
Convex Optimization 10-725

Last time: stochastic gradient descent

Consider

$$\min_x \frac{1}{m}\sum_{i=1}^m f_i(x)$$

Stochastic gradient descent or SGD: let $x^{(0)} \in \mathbb{R}^n$, repeat:

$$x^{(k)} = x^{(k-1)} - t_k \cdot \nabla f_{i_k}(x^{(k-1)}), \quad k = 1, 2, 3, \ldots$$

where $i_k \in \{1, \ldots m\}$ is chosen uniformly at random. Step sizes t_k chosen to be fixed and small, or diminishing

Compare to full gradient, which would use $\frac{1}{m}\sum_{i=1}^m \nabla f_i(x)$. Upside of SGD: much (potentially much, much) cheaper iterations

Downside: can be slow to converge, attains suboptimal rates. Can be improved (more later?)

Lower bounds in linear programs

Suppose we want to find lower bound on the optimal value in our convex problem, $B \leq \min_x f(x)$

E.g., consider the following simple LP

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

What's a lower bound? Easy, take $B = 2$

But didn't we get "lucky"?

Try again:

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

$$\begin{aligned} & x + y \geq 2 \\ + \quad & 2y \geq 0 \\ = \quad & x + 3y \geq 2 \end{aligned}$$

Lower bound $B = 2$

More generally:

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

$$\begin{aligned} a + b &= p \\ a + c &= q \\ a, b, c &\geq 0 \end{aligned}$$

Lower bound $B = 2a$, for any a, b, c satisfying above

What's the best we can do? Maximize our lower bound over all possible a, b, c:

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

Called primal LP

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

Called dual LP

Note: number of dual variables is number of primal constraints

Try another one:

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

Primal LP

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

Dual LP

Note: in the dual problem, c is unconstrained

Outline

Today:

- Duality in general LPs
- Max flow and min cut
- Second take on duality
- Matrix games

Duality for general form LP

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

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

Primal LP

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

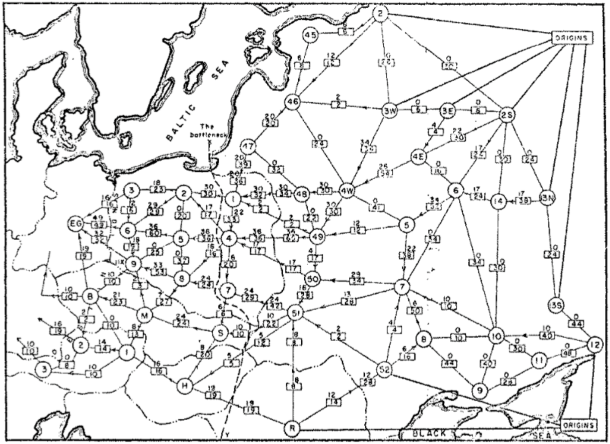
Dual LP

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

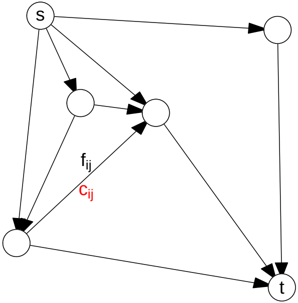
$$u^T(Ax - b) + v^T(Gx - h) \leq 0, \quad \text{i.e.,}$$
$$(-A^T u - G^T v)^T x \geq -b^T u - h^T v$$

So if $c = -A^T u - G^T v$, we get a bound on primal optimal value

Example: max flow and min cut



Soviet railway network (from Schrijver (2002), “On the history of transportation and maximum flow problems”)



Given graph $G = (V, E)$, define flow f_{ij}, $(i,j) \in E$ to satisfy:

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

Max flow problem: find flow that maximizes total value of the flow from s to t. I.e., as an LP:

$$\max_{f\in\mathbb{R}^{|E|}} \quad \sum_{(s,j)\in E} f_{sj}$$

$$\text{subject to} \quad 0 \leq f_{ij} \leq c_{ij} \quad \text{for all } (i,j) \in E$$

$$\sum_{(i,k)\in E} f_{ik} = \sum_{(k,j)\in E} f_{kj} \quad \text{for all } k \in V \setminus \{s,t\}$$

Derive the dual, in steps:

- Note that

$$\sum_{(i,j)\in E} \Big(-a_{ij}f_{ij} + b_{ij}(f_{ij} - c_{ij})\Big) + \sum_{k\in V\setminus\{s,t\}} x_k\Big(\sum_{(i,k)\in E} f_{ik} - \sum_{(k,j)\in E} f_{kj}\Big) \le 0$$

  for any $a_{ij}, b_{ij} \ge 0$, $(i,j) \in E$, and x_k, $k \in V \setminus \{s,t\}$

- Rearrange as

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

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

- Want to make LHS in previous inequality equal to primal objective, i.e.,

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

- We've shown that

$$\text{primal optimal value} \le \sum_{(i,j) \in E} b_{ij} c_{ij},$$

subject to a, b, x satisfying constraints. Hence dual problem is (minimize over a, b, x to get best upper bound):

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

Suppose that at the solution, it just so happened that

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

Let $A = \{i : x_i = 1\}$, $B = \{i : x_i = 0\}$; note $s \in A$, $t \in B$. Then

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

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

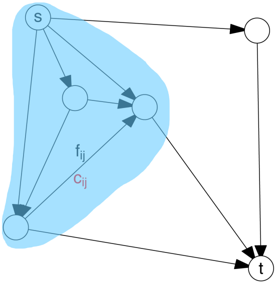
I.e., we've argued that the dual is the LP relaxation of the min cut problem:

$$\min_{b \in \mathbb{R}^{|E|},\, x \in \mathbb{R}^{|V|}} \quad \sum_{(i,j)\in E} b_{ij} c_{ij}$$

$$\text{subject to} \quad b_{ij} \geq x_i - x_j$$

$$b_{ij}, x_i, x_j \in \{0, 1\}$$

$$\text{for all } i, j$$



Therefore, from what we know so far:

value of max flow $\leq$

optimal value for LP relaxed min cut $\leq$

capacity of min cut

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, we get all equalities. In particular, we get that the primal LP and dual LP have exactly the same optimal values, a phenomenon called strong duality

How often does this happen? More on this soon

Another perspective on LP duality

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

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

Explanation # 2: for any u and $v \ge 0$, and x primal feasible

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

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

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

In other words, $g(u,v)$ is a lower bound on $f^\star$ for any u and $v \geq 0$

Note that

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

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)

Example: mixed strategies for matrix games

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

$$
\begin{array}{cc|cccc}
 & & & \text{R} & & \\
 & & 1 & 2 & \dots & n \\
\hline
 & 1 & P_{11} & P_{12} & \dots & P_{1n} \\
\text{J} & 2 & P_{21} & P_{22} & \dots & P_{2n} \\
 & \dots & & & & \\
 & m & P_{m1} & P_{m2} & \dots & P_{mn}
\end{array}
$$

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 probability distribution, and then

$$x: \quad \mathbb{P}(\text{J chooses } i) = x_i, \ i = 1, \dots m$$
$$y: \quad \mathbb{P}(\text{R chooses } j) = y_j, \ j = 1, \dots n$$

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 $x^T P y$, which results in J paying off

$$\max\{x^T P y : \ y \geq 0, \ 1^T y = 1\} = \max_{i=1,\ldots n} (P^T x)_i$$

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

$$\begin{array}{ll} \min\limits_{x} & \max\limits_{i=1,\ldots n} (P^T x)_i \\ \text{subject to} & x \geq 0, \ 1^T x = 1 \end{array}$$

In an alternate 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

$$\begin{array}{ll} \max\limits_{y} & \min\limits_{j=1,\dots m} (Py)_j \\ \text{subject to} & y \geq 0,\ 1^T y = 1 \end{array}$$

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

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

Recast first problem as an LP:

$$\begin{aligned} \min_{x,t} \quad & \\ \text{subject to} \quad & x \geq 0,\ 1^T x = 1 \\ & P^T x \leq t \end{aligned}$$

Now form what we call the Lagrangian:

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

and what we call the Lagrange dual function:

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

Hence dual problem, after eliminating slack variable u, is

$$\begin{aligned} \max_{y,v} \quad & v \\ \text{subject to} \quad & y \geq 0, \ 1^T y = 1 \\ & Py \geq v \end{aligned}$$

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 28–30