Duality

Barnabas Poczos & Ryan Tibshirani
Convex Optimization 10-725/36-725

Duality in linear programs

Suppose we want to find lower bound on the optimal value in our convex problem, $B \le \min_{x \in C} f(x)$

E.g., consider the following simple LP

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

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

But didn't we get "lucky"?

Try again:

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

$$\begin{array}{rr} & x + y \geq 2 \\ + & 2y \geq 0 \\ = & x + 3y \geq 2 \end{array}$$

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:

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

Primal LP

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

Dual LP

Note: in the dual problem, c is unconstrained

General form LP

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

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

Primal LP

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

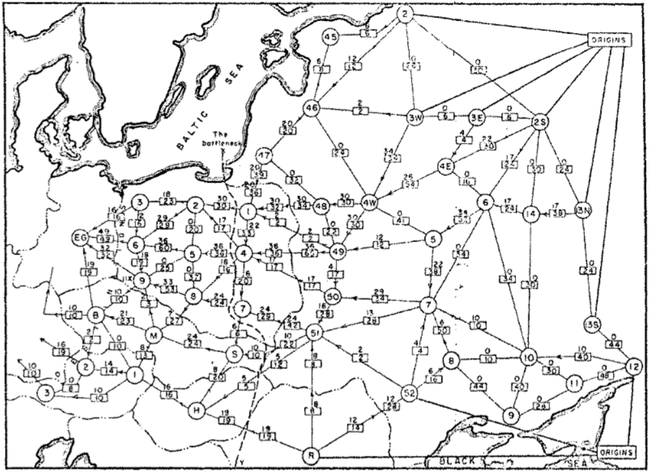
Dual LP

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

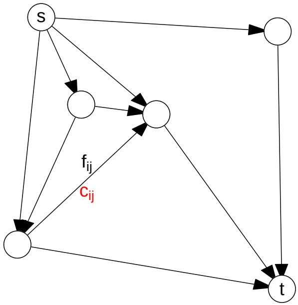
$$
u^T(Ax - b) + v^T(Gx - h) \le 0, \quad \text{i.e.,}
$$
$$
(-A^T u - G^T v)^T x \ge -b^T u - h^T v
$$

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

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:

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

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

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

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

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

Derive the dual, in steps:

- Note that

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

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

- Rearrange as

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

  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

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

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

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

Suppose that at the solution, it just so happened

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

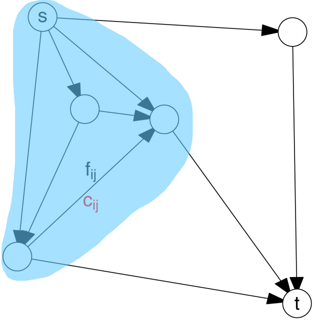
Call $A = \{i : x_i = 1\}$ and $B = \{i : x_i = 0\}$, note that $s \in A$ and $t \in B$. Then the constraints

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

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:

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



Therefore, from what we 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 later

(From F. Estrada et al. (2004), “Spectral embedding and min cut for image segmentation”)

Another perspective on LP duality

$$\min_{x \in \mathbb{R}^n} \; c^T x \quad \text{subject to} \quad Ax = b, \;\; Gx \le h$$

Primal LP

$$\max_{u \in \mathbb{R}^m,\, v \in \mathbb{R}^r} \; -b^T u - h^T v \quad \text{subject to} \quad -A^T u - G^T v = c, \;\; v \ge 0$$

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 $v \ge 0$,

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

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)

Outline

Rest of today:

- Lagrange dual function
- Langrange dual problem
- Examples
- Weak and strong duality

Lagrangian

Consider general minimization problem

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

Need not be convex, but of course we will pay special attention to convex case

We define the Lagrangian as

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

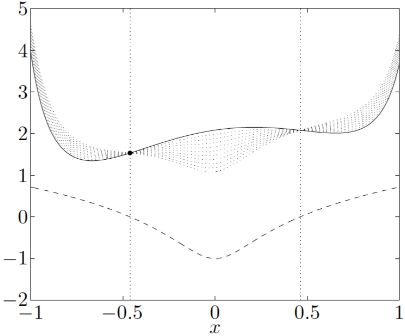
New variables $u \in \mathbb{R}^m, v \in \mathbb{R}^r$, with $u \ge 0$ (implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$)

Important property: for any $u \geq 0$ and v,

$$f(x) \geq L(x,u,v) \quad \text{at each feasible } x$$

Why? For feasible x,

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



- Solid line is f
- Dashed line is h, hence feasible set $\approx [-0.46, 0.46]$
- Each dotted line shows $L(x,u,v)$ for different choices of $u \geq 0$ and v

(From B & V page 217)

Lagrange dual function

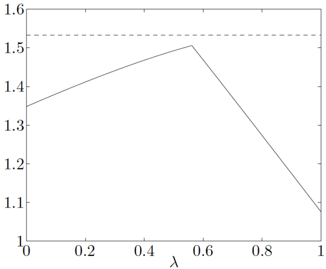
Let C denote primal feasible set, $f^\star$ denote primal optimal value. Minimizing $L(x, u, v)$ over all $x \in \mathbb{R}^n$ gives a lower bound:

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

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, v

- Dashed horizontal line is $f^\star$
- Dual variable λ is (our u)
- Solid line shows $g(\lambda)$

(From B & V page 217)



Quadratic program

Consider quadratic program (QP, step up from LP!)

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

where $Q \succ 0$. Lagrangian:

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

Lagrange dual function:

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

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

Same problem

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

but now $Q \succeq 0$. Lagrangian:

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

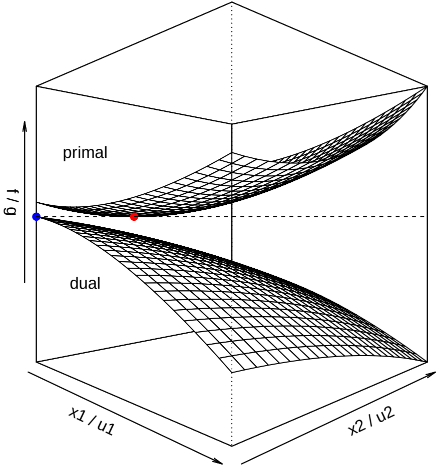
Lagrange dual function:

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

where Q^+ denotes generalized inverse of Q. For any $u \geq 0$, v, and $c - u + A^T v \perp \mathrm{null}(Q)$, $g(u, v)$ is a nontrivial lower bound on $f^\star$

Quadratic program in 2D

We choose $f(x)$ to be quadratic in 2 variables, subject to $x \geq 0$. Dual function $g(u)$ is also quadratic in 2 variables, also subject to $u \geq 0$



Dual function $g(u)$ provides a bound on $f^\star$ for every $u \geq 0$

Largest bound this gives us: turns out to be exactly $f^\star$ ... coincidence?

More on this later

Lagrange dual problem

Given primal problem

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

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

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

Key property, called weak duality: if dual optimal value $g^\star$, then

$$f^\star \geq g^\star$$

Note that this always holds (even if primal problem is nonconvex)

Another key property: the dual problem is a convex optimization problem (as written, it is a concave maximization problem)

Again, this is always true (even when primal problem is not convex)

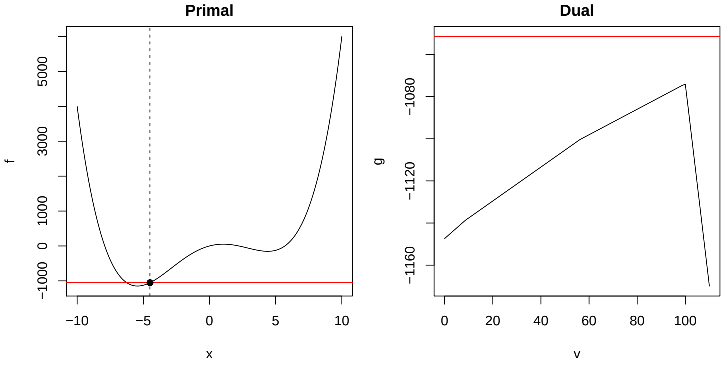
By definition:

$$\begin{aligned} g(u,v) &= \min_{x \in \mathbb{R}^n} \Big\{ f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j \ell_j(x) \Big\} \\ &= -\underbrace{\max_{x \in \mathbb{R}^n} \Big\{ -f(x) - \sum_{i=1}^m u_i h_i(x) - \sum_{j=1}^r v_j \ell_j(x) \Big\}}_{\text{pointwise maximum of convex functions in } (u,v)} \end{aligned}$$

I.e., g is concave in (u,v), and $u \geq 0$ is a convex constraint, hence dual problem is a concave maximization problem

Nonconvex quartic minimization

Define $f(x) = x^4 - 50x^2 + 100x$ (nonconvex), minimize subject to constraint $x \geq -4.5$



Dual function g can be derived explicitly (via closed-form equation for roots of a cubic equation). Form of g is quite complicated, and would be hard to tell whether or not g is concave ... but it must be!

Strong duality

Recall that we always have $f^\star \geq g^\star$ (weak duality). On the other hand, in some problems we have observed that actually

$$f^\star = g^\star$$

which is called strong duality

Slater's condition: if the primal is a convex problem (i.e., f and $h_1, \ldots h_m$ are convex, $\ell_1, \ldots \ell_r$ are affine), and there exists at least one strictly feasible $x \in \mathbb{R}^n$, meaning

$$h_1(x) < 0, \ldots h_m(x) < 0 \quad \text{and} \quad \ell_1(x) = 0, \ldots \ell_r(x) = 0$$

then strong duality holds

This is a pretty weak condition. (And it can be further refined: need strict inequalities only over functions h_i that are not affine)

Back to where we started

For linear programs:

- Easy to check that the dual of the dual LP is the primal LP
- Refined version of Slater's condition: strong duality holds for an LP if it is feasible
- Apply same logic to its dual LP: strong duality holds if it is feasible
- Hence strong duality holds for LPs, except when both primal and dual are infeasible

In other words, we pretty much always have strong duality for LPs

Mixed strategies for matrix games

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

		R			
		1	2	...	n
	1	P_{11}	P_{12}	...	P_{1n}
B	2	P_{21}	P_{22}	...	P_{2n}
	...				
	m	P_{m1}	P_{m2}	...	P_{mn}

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

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

$$x: \quad \mathbb{P}(\text{B chooses } i) = x_i, \ i = 1, \dots m$$

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

The expected payout then, from B 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 B is wiser, he will allow R to know his strategy x ahead of time. In this case, R will definitely choose y to maximize $x^T P y$, which results in B paying off

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

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

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

In an alternate universe, if R were somehow wiser than B, then he might allow B to know his strategy y beforehand

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

$$\max_{y \in \mathbb{R}^n} \min_{j=1,\ldots m} (Py)_j$$
$$\text{subject to } \; y \geq 0, \; 1^T y = 1$$

Call B'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, $f_1^\star \geq f_2^\star$

We can show using strong duality that $f_1^\star = f_2^\star$ ... which may come as a surprise!

Recast first problem as an LP

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

Lagrangian and Lagrange dual function

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

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

Hence dual problem is

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

This is exactly the second problem, and we have strong LP duality

References

- S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 5
- R. T. Rockafellar (1970), "Convex analysis", Chapters 28–30