Lecture 09

Introduction to Duality

Duality in Linear Programs

Lower bounds in linear programs

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

E.g., consider the following simple LP

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

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

But didn't we get "lucky"?

Try again:

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

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

Lower bound $B = 2$

More generally:

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

the constraint can be equivalently represented as

$$\begin{aligned} & ax + ay \geq 2a, \\ & b\,x \geq 0, \qquad\qquad a, b, c \geq 0. \\ & cy \geq 0, \end{aligned}$$

Adding them together, we have that

$$\underbrace{(a+b)}_{=p}x + \underbrace{(a+c)}_{=q}y \geq 2a.$$

⇒

$$a + b = p$$
$$a + c = q$$
$$a, b, c \geq 0$$

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

Try another one:

Primal LP

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

Dual LP

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

The constraint of the linear program can be equivalently represented as

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

Note: in the dual problem, c is unconstrained

Adding them together, we have

$$\underbrace{(a + 3c)}_{=p}x + \underbrace{(-b + c)}_{=q}y \geq -b + 2c.$$

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

Another perspective on LP duality

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

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

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

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

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

$$g(u,v) = \min_x c^T x + u^T(Ax - b) + v^T(Gx - h) =$$

$$\min_x \underbrace{(c + A^T u + G^T v)^T x}_{linear\ function\ of\ x} - b^T u - h^T v$$

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

This second explanation reproduces the same dual, but is actually completely general and applies to arbitrary optimization problems (even nonconvex ones)

$$\max_{u,v} g(u,v) \quad s.t.\ v \geq 0$$

⟺

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

Dual LP

Lagrangian

Consider general minimization problem

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

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

We define the Lagrangian as

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

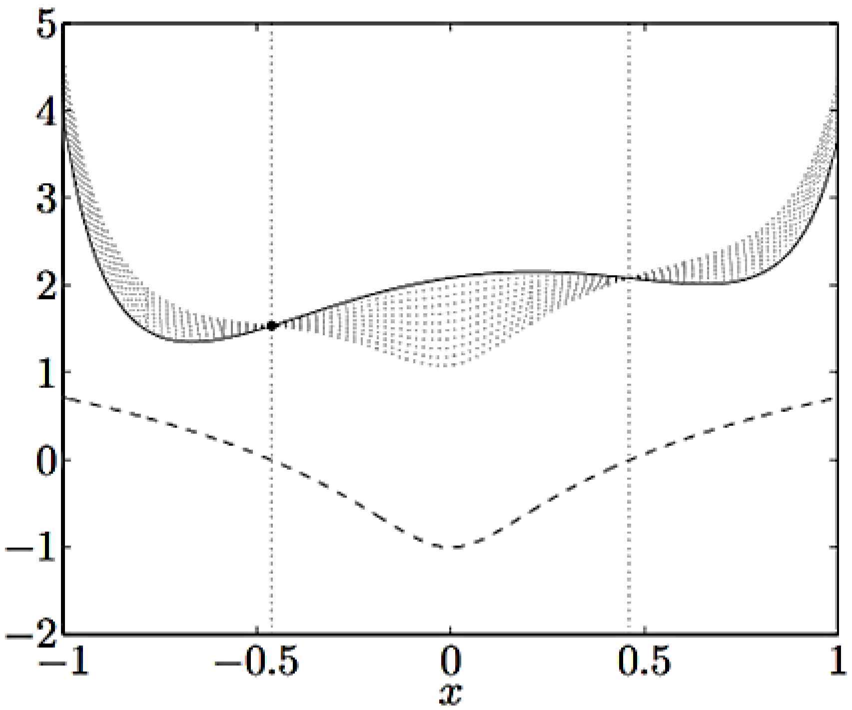
New variables $u \in \mathbb{R}^m, v \in \mathbb{R}^r$, with $u \geq 0$ (implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$)

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

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

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 $u \geq 0$

(From B & V page 217)

Lagrange dual function

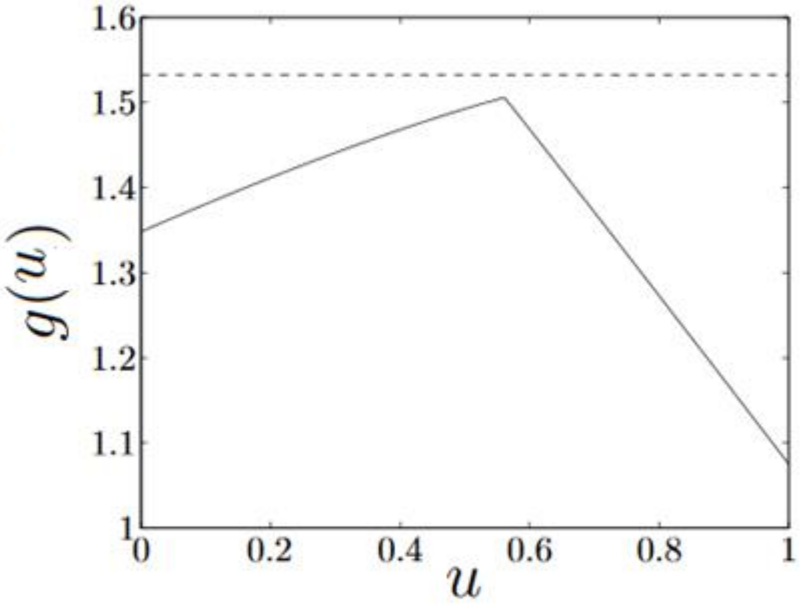
Let C denote primal feasible set, $f^\star$ denote primal optimal value. Minimizing $L(x,u,v)$ over all x gives a lower bound:

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

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, v

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

(From B & V page 217)



Example: quadratic program

Consider quadratic program:

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

where $Q \succ 0$. Lagrangian:

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

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

To compute the dual function $g(u, v) = \min_x L(x, u, v)$, we minimize the Lagrangian above by taking the gradient with respect to x and setting it equal to zero, and we get that

$$x^* = -Q^{-1}(c - u + A^T v)$$

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

$$x^* = -Q^{-1}(c - u + A^T v)$$

Lagrange dual function:

$$g(u, v) = \min_x \; L(x, u, v) = \; L(x^*, u, v)$$

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

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

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

Same problem

$$\begin{array}{ll} \min_{x} & \frac{1}{2}x^TQx + c^Tx \\ \text{subject to} & Ax = b,\ x \geq 0 \end{array}$$

but now $Q \succeq 0$. Lagrangian:

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

if we try to minimize the Lagrangian above by setting the gradient to 0, we get the following constraint at the optimum:

$$Qx = -(c - u + A^Tv)$$

$$Qx = -(c - u + A^T v) \qquad (*)$$

- Now, there are two cases:

(i) $c - u + A^T v \in \text{col}(Q)$.

- Then, we can use the pseudo-inverse $Q^\dagger$ of Q.

(ii) $c - u + A^T v \notin \text{col}(Q)$,

- But in this case, there is no x that satisfies eq. (*) and so there is no unique minimizer x^*.

Consider again case (ii), where

$$c - u + A^T v \notin \mathrm{col}(Q),$$

We can still find a min of $L(x, u, v)$:

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

- let $c - u + A^T v = z_1 + z_2$, where

$$z_1 \in \mathrm{col}(Q),$$

$$z_2 \in \mathrm{null}(Q), z_2 \neq 0.$$

If we take x to be a multiple of $-z_2$, we'll have: $\frac{1}{2} x^T Q x = 0$

But, we can minimize the term $(c - u + A^T v)^T x$ as much as we like $\Rightarrow \min_x L(x, u, v) = -\infty$

So,

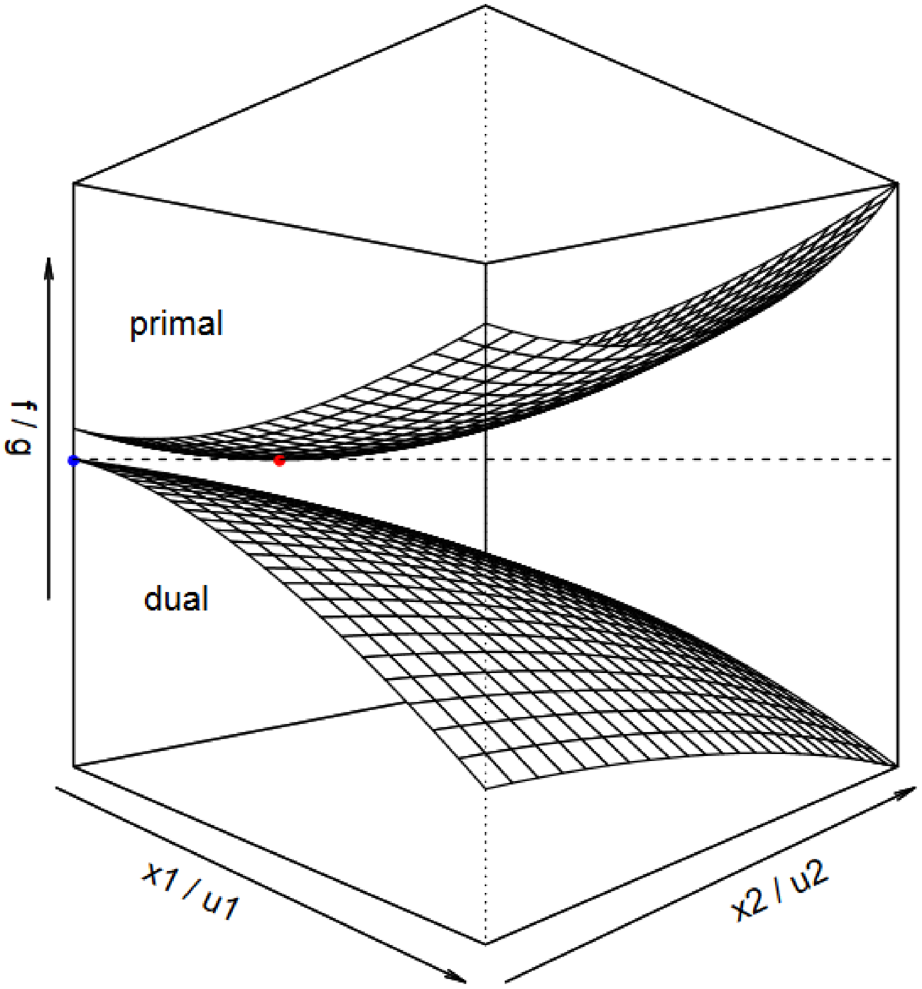
Lagrange dual function:

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

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

Example: 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, via KKT conditions

Lagrange dual problem

Given primal problem

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

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

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

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

$$f^\star \ge g^\star$$

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)

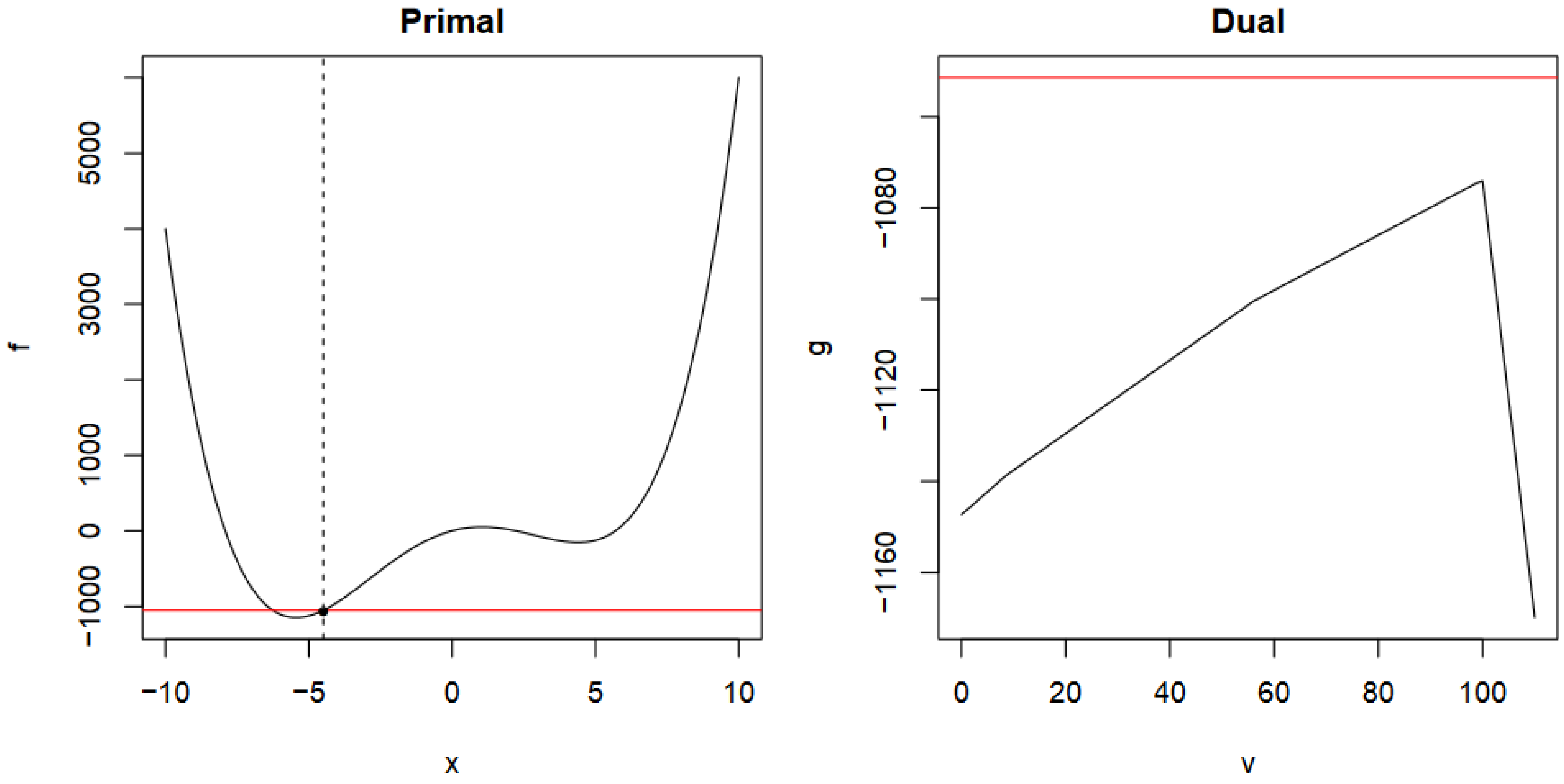
By definition:

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

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

Example: 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 rather complicated:

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

where for $i = 1, 2, 3$,

$$F_i(u) = \frac{-a_i}{12 \cdot 2^{1/3}} \left(432(100-u) - \left(432^2(100-u)^2 - 4 \cdot 1200^3\right)^{1/2}\right)^{1/3}$$
$$-100 \cdot 2^{1/3} \frac{1}{\left(432(100-u) - \left(432^2(100-u)^2 - 4 \cdot 1200^3\right)^{1/2}\right)^{1/3}},$$

and $a_1 = 1$, $a_2 = (-1 + i\sqrt{3})/2$, $a_3 = (-1 - i\sqrt{3})/2$

Without the context of duality it would be difficult to tell whether or not g is concave ... but we know 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. An important refinement: strict inequalities only need to hold over functions h_i that are not affine

LPs: 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 nearly always have strong duality for LPs)

Example

Given $y \in \{-1,1\}^n$, $X \in \mathbb{R}^{n\times p}$, rows $x_1,\ldots x_n$, recall the problem:

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

Introducing dual variables $v, w \ge 0$, we form the Lagrangian:

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

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

$$= \frac{1}{2}\beta^T\beta - w^T diag(y)X\beta + \underbrace{(C1 - v - w)^T}_{affine}\xi - \underbrace{w^T y}_{affine}\beta_0 + 1^T w$$

Since $\beta, \beta_0,$ and ξ have no interactions, $L(.)$ can be minimized separately on these variables!

- We first minimize on β:

Define $\tilde{X} \stackrel{\text{def}}{=} diag(y)X$

$$\nabla_\beta L = 0 \Longrightarrow \beta^{*,T} - w^T\tilde{X} = 0 \Rightarrow \begin{array}{l} \beta^{*,T} = w^T\tilde{X}, \\ \beta^* = \tilde{X}^T w \end{array}$$

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

$$g(v,w) = \begin{cases} \frac{1}{2}\beta^{*,T}\beta - \underbrace{w^T\widetilde{X}}_{=\beta^{*,T}}\beta^* + 1^T w & \text{if } w = C1 - v,\ w^T y = 0 \\ -\infty & \text{otherwise} \end{cases}$$

$$= -\frac{1}{2}\beta^{*,T}\beta$$

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

Thus dual problem, eliminating slack variable v, becomes

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

Check: Slater's condition is satisfied, and we have strong duality.

Duality gap

Given primal feasible 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

$$f(x) - f^\star \leq f(x) - g(u, v)$$

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

From an algorithmic viewpoint, provides a stopping criterion: if $f(x) - g(u, v) \leq \epsilon$, then we are guaranteed that $f(x) - f^\star \leq \epsilon$

Very useful, especially in conjunction with iterative methods ...

Summary

Given a minimization problem

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

we defined the Lagrangian:

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

and Lagrange dual function:

$$g(u, v) = \min_x \; L(x, u, v)$$

The subsequent dual problem is:

$$\begin{array}{ll} \max\limits_{u,v} & g(u,v) \\ \text{subject to} & u \geq 0 \end{array}$$

Important properties:

- Dual problem is always convex, i.e., g is always concave (even if primal problem is not convex)
- The primal and dual optimal values, $f^\star$ and $g^\star$, always satisfy weak duality: $f^\star \geq g^\star$
- Slater's condition: for convex primal, if there is an x such that

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

then strong duality holds: $f^\star = g^\star$. Can be further refined to strict inequalities over the nonaffine h_i, $i = 1, \ldots m$

Appendix

Some notes from linear algebra

Pseudo-inverse

- For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, we can define the pseudo-inverse $A^\dagger$ in terms of its Decomposition.
- we can write A as

$$A = UDU^T$$

- If A was invertible, we can directly invert the decomposition above:

$$A^{-1} = (UDU^T)^{-1} = (U^T)^{-1}D^{-1}U^{-1} = UD^{-1}U^T$$

where

$$D^{-1} = \begin{bmatrix} \frac{1}{d_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{d_2} & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{d_n} \end{bmatrix}$$

Pseudo-inverse

- If A is not invertible, we're going to see that for $k = \text{rank}(A)$,

$$d_{k+1} = d_{k+2} = \cdots = d_n = 0.$$

- In this case, we can construct a pseudo-inverse $(D^\dagger)$ of D as follows:

$$D^\dagger = \begin{bmatrix} \frac{1}{d_1} & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{d_2} & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ddots & \vdots & 0 & 0 & 0 \\ 0 & 0 & \ldots & \frac{1}{d_k} & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 \\ 0 & 0 & \ldots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 0 & 0 & 0 \end{bmatrix}$$

- And our pseudo-inverse, then, is

$$A^\dagger = UD^\dagger U^T$$

For symmetric matrices, $\boldsymbol{NULL(A) \perp Col(A)}$

$$Col(A) = \{v | \exists x: Ax = v\}$$

$$NULL(A) = \{u | Au = 0\}$$

$$if\ A = A^T \Rightarrow Col(A) \perp NULL(A), \qquad \boldsymbol{why}?$$

$$Let\ v \in Col(A). then: \quad Ax = v$$

$$Let\ u \in NULL(A). then: \quad Au = 0 \implies u^T A^T = 0 \xRightarrow{A=A^T} u^T A = 0$$

$$Now, we\ have: Ax = v \implies u^T Ax = u^T v = 0. \quad \checkmark$$

We say that for a symmetric matrix A, NULL(A) and Col(A) are "orthogonal complements".