Advanced Introduction to Machine Learning CMU-10715

Duality

Barnabás Póczos, 2015 Fall

Credits

Many of these slides are taken from Ryan Tibshirani's convex optimization class

References:

S. Boyd and L. Vandenberghe (2004), "Convex optimization", Chapter 5

R. T. Rockafellar (1970), "Convex analysis", Chapters 28–30

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"?

Duality in linear programs

Try again:

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

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

Lower bound $B = 2$

More generally:

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

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

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

Duality in linear programs

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

Duality in linear programs

Try another one:

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

Primal LP

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

Dual LP

Note: in the dual problem, c is unconstrained

Duality in linear programs

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

Another perspective on LP duality

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

Primal LP

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

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 $v \ge 0$,

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

Another perspective on LP duality

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)

Lagrangian

Consider general minimization problem

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

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 $u < 0$)

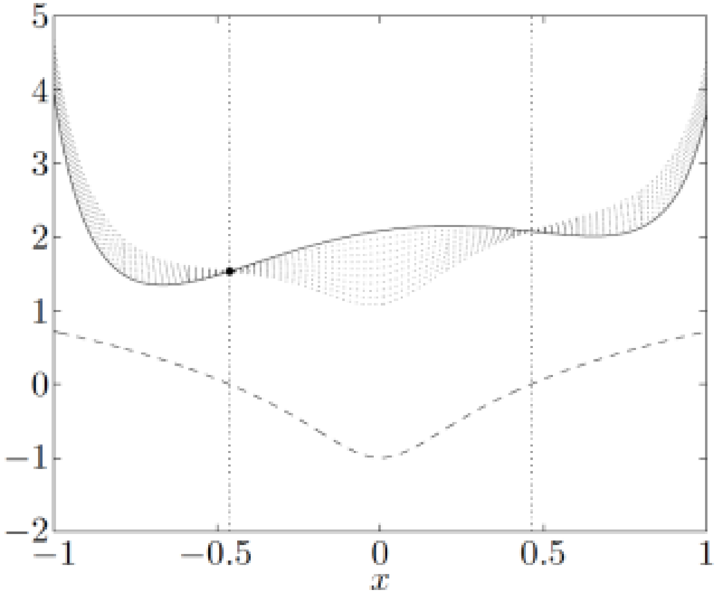
Lagrangian

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

Quadratic program

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

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

where $Q \succ 0$. Lagrangian:

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

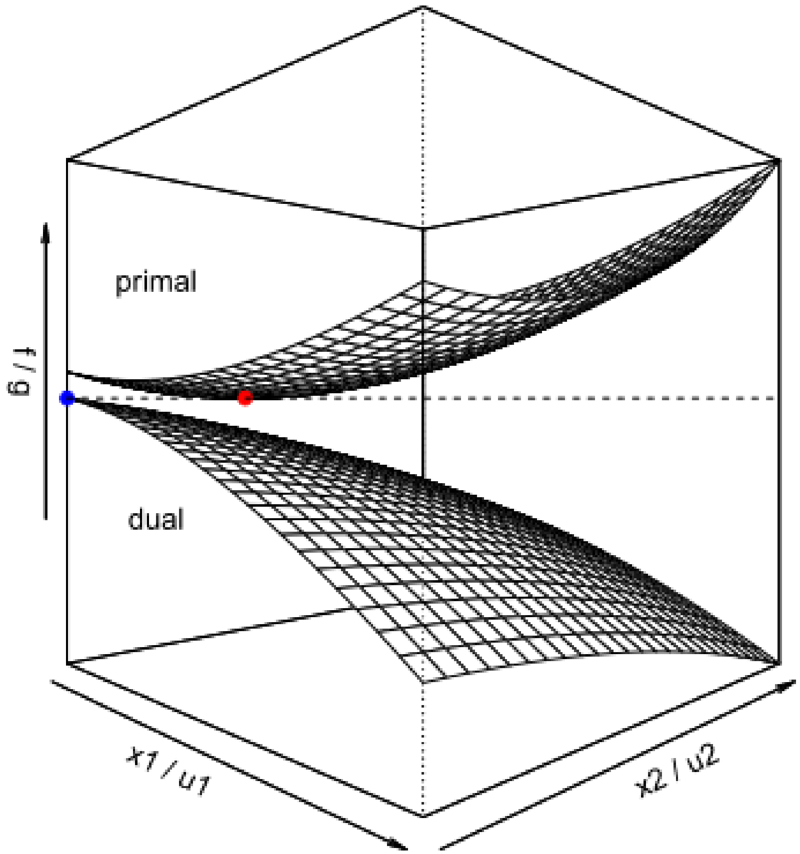
Lagrange dual function:

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

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

QP 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

Weak duality

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)

Dual is Convex

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

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)

Strong duality for LPs

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

KKT Conditions

What we have seen so far

Given a minimization problem

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

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:

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

What we have seen so far

The subsequent dual problem is:

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

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, $i = 1, \ldots m$)

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 ... more dual uses in coming lectures

Subgradients

Remember that for convex $f : \mathbb{R}^n \to \mathbb{R}$,

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{all } x, y$$

I.e., linear approximation always underestimates f

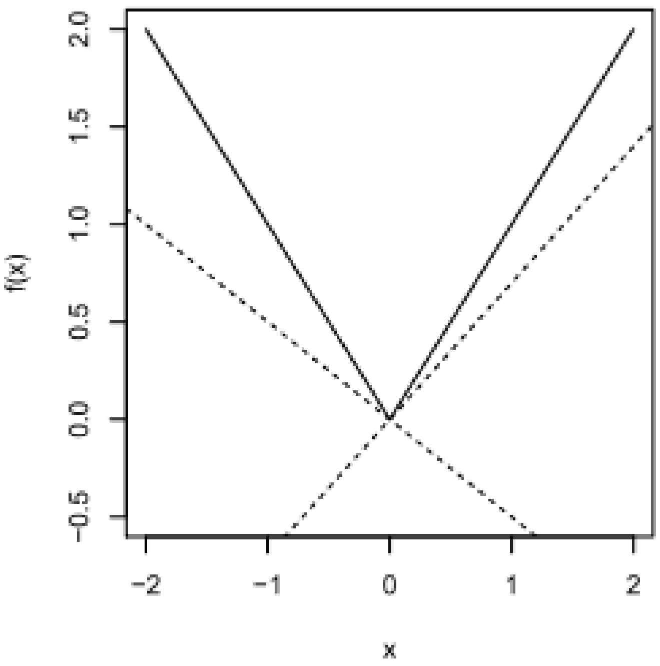
A subgradient of convex $f : \mathbb{R}^n \to \mathbb{R}$ at x is any $g \in \mathbb{R}^n$ such that

$$f(y) \geq f(x) + g^T (y - x), \quad \text{all } y$$

- Always exists
- If f differentiable at x, then $g = \nabla f(x)$ uniquely
- Actually, same definition works for nonconvex f (however, subgradients need not exist)

Subgradients - Example

Consider $f : \mathbb{R} \to \mathbb{R}$, $f(x) = |x|$



- For $x \neq 0$, unique subgradient $g = \mathrm{sign}(x)$
- For $x = 0$, subgradient g is any element of $[-1, 1]$

Subdifferential

Set of all subgradients of convex f is called the subdifferential:

$$\partial f(x) = \{g \in \mathbb{R}^n : g \text{ is a subgradient of } f \text{ at } x\}$$

- $\partial f(x)$ is closed and convex (even for nonconvex f)
- Nonempty (can be empty for nonconvex f)
- If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$
- If $\partial f(x) = \{g\}$, then f is differentiable at x and $\nabla f(x) = g$

Optimality condition

For any f (convex or not),

$$f(x^\star) = \min_{x \in \mathbb{R}^n} f(x) \iff 0 \in \partial f(x^\star)$$

I.e., $x^\star$ is a minimizer if and only if 0 is a subgradient of f at $x^\star$

Why? Easy: $g = 0$ being a subgradient means that for all y

$$f(y) \geq f(x^\star) + 0^T(y - x^\star) = f(x^\star)$$

Note implication for differentiable case, where $\partial f(x) = \{\nabla f(x)\}$

Karush-Kuhn-Tucker conditions

Given general problem

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

The Karush-Kuhn-Tucker conditions or KKT conditions are:

- $0 \in \partial f(x) + \sum_{i=1}^{m} u_i \partial h_i(x) + \sum_{j=1}^{r} v_j \partial \ell_j(x)$ (stationarity)
- $u_i \cdot h_i(x) = 0$ for all i (complementary slackness)
- $h_i(x) \le 0, \; \ell_j(x) = 0$ for all i, j (primal feasibility)
- $u_i \ge 0$ for all i (dual feasibility)

Necessity – Part 1

Let $x^\star$ and $u^\star, v^\star$ be primal and dual solutions with zero duality gap (strong duality holds, e.g., under Slater's condition). Then

$$
\begin{aligned}
f(x^\star) &= g(u^\star, v^\star) \\
&= \min_{x \in \mathbb{R}^n} f(x) + \sum_{i=1}^m u_i^\star h_i(x) + \sum_{j=1}^r v_j^\star \ell_j(x) \\
&\le f(x^\star) + \sum_{i=1}^m u_i^\star h_i(x^\star) + \sum_{j=1}^r v_j^\star \ell_j(x^\star) \\
&\le f(x^\star)
\end{aligned}
$$

In other words, all these inequalities are actually equalities

Necessity - Part 2

Two things to learn from this:

- The point $x^\star$ minimizes $L(x, u^\star, v^\star)$ over $x \in \mathbb{R}^n$. Hence the subdifferential of $L(x, u^\star, v^\star)$ must contain 0 at $x = x^\star$—this is exactly the stationarity condition
- We must have $\sum_{i=1}^m u_i^\star h_i(x^\star) = 0$, and since each term here is ≤ 0, this implies $u_i^\star h_i(x^\star) = 0$ for every i—this is exactly complementary slackness

Primal and dual feasibility obviously hold. Hence, we've shown:

If $x^\star$ and $u^\star, v^\star$ are primal and dual solutions, with zero duality gap, then $x^\star, u^\star, v^\star$ satisfy the KKT conditions

(Note that this statement assumes nothing a priori about convexity of our problem, i.e., of f, h_i, ℓ_j)

Sufficiency

If there exists $x^\star, u^\star, v^\star$ that satisfy the KKT conditions, then

$$\begin{aligned} g(u^\star, v^\star) &= f(x^\star) + \sum_{i=1}^m u_i^\star h_i(x^\star) + \sum_{j=1}^r v_j^\star \ell_j(x^\star) \\ &= f(x^\star) \end{aligned}$$

where the first equality holds from stationarity, and the second holds from complementary slackness

Therefore duality gap is zero (and $x^\star$ and $u^\star, v^\star$ are primal and dual feasible) so $x^\star$ and $u^\star, v^\star$ are primal and dual optimal. I.e., we've shown:

> If $x^\star$ and $u^\star, v^\star$ satisfy the KKT conditions, then $x^\star$ and $u^\star, v^\star$ are primal and dual solutions

Putting it together

In summary, KKT conditions:

- always sufficient
- necessary under strong duality

Putting it together:

For a problem with strong duality (e.g., assume Slater's condition: convex problem and there exists x strictly satisfying non-affine inequality contraints),

$$\begin{aligned} & x^\star \text{ and } u^\star, v^\star \text{ are primal and dual solutions} \\ \Leftrightarrow \quad & x^\star \text{ and } u^\star, v^\star \text{ satisfy the KKT conditions} \end{aligned}$$

(Warning, concerning the stationarity condition: for a differentiable function f, we cannot use $\partial f(x) = \{\nabla f(x)\}$ unless f is convex)

Quadratic with equality constraints

Consider for $Q \succeq 0$,

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^T Q x + c^T x \\ \text{subject to} \quad & Ax = 0 \end{aligned}$$

E.g., as in Newton step for $\min_{x \in \mathbb{R}^n} f(x)$ subject to $Ax = b$

Convex problem, no inequality constraints, so by KKT conditions: x is a solution if and only if

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} = \begin{bmatrix} -c \\ 0 \end{bmatrix}$$

for some u. Linear system combines stationarity, primal feasibility (complementary slackness and dual feasibility are vacuous)