Duality in General Programs

Ryan Tibshirani
Convex Optimization 10-725/36-725

Last time: duality in linear programs

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

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

Primal LP

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

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

Explanation # 2: 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 $v \geq 0$. Note that

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

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

Outline

Today:

- Lagrange dual function
- Langrange dual problem
- Weak and strong duality
- Examples
- Preview of duality uses

Lagrangian

Consider general minimization problem

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

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

We define the Lagrangian as

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

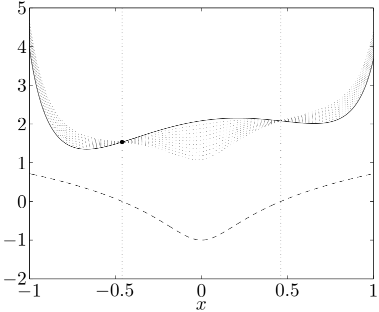
New variables $u \in \mathbb{R}^m, v \in \mathbb{R}^r$, with $u \ge 0$ (implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$)

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

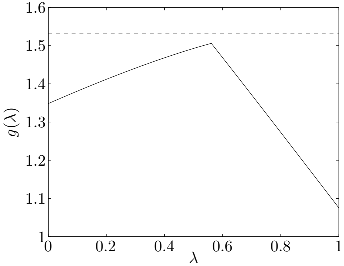
Let C denote primal feasible set, $f^\star$ denote primal optimal value. Minimizing $L(x,u,v)$ over all x gives a lower bound:

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

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 λ is (our u)
- Solid line shows $g(\lambda)$

(From B & V page 217)



Example: quadratic program

Consider quadratic program:

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

where $Q \succ 0$. Lagrangian:

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

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

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

but now $Q \succeq 0$. Lagrangian:

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

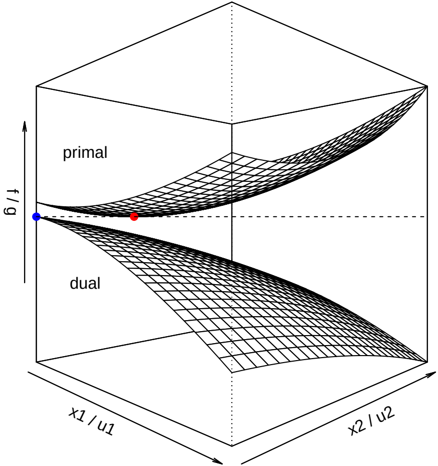
Lagrange dual function:

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

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

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

Our constructed dual function $g(u, v)$ satisfies $f^\star \geq g(u, v)$ 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:

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

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

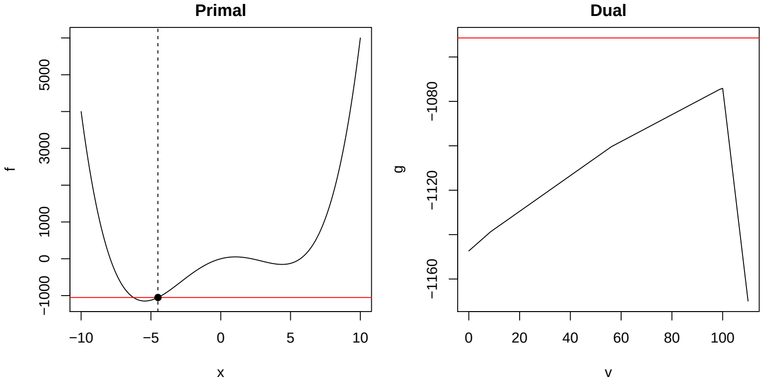
By definition:

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

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 quite complicated:

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

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. (Further refinement: only require strict inequalities 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: support vector machine dual

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

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

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

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

Minimizing over β, β_0, ξ gives Lagrange dual function:

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

where $\tilde{X} = \mathrm{diag}(y) X$. Thus SVM dual problem, eliminating slack variable v, becomes

$$\begin{aligned} \max_w \quad & -\frac{1}{2} w^T \tilde{X} \tilde{X}^T w + 1^T w \\ \text{subject to} \quad & 0 \le w \le C1, \; w^T y = 0 \end{aligned}$$

Check: Slater's condition is satisfied, and we have strong duality. Further, from study of SVMs, might recall that at optimality

$$\beta = \tilde{X}^T w$$

This is not a coincidence, as we'll later via the KKT conditions

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

Dual norms

Let $\|x\|$ be a norm, e.g.,

- ℓ_p norm: $\|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}$, for $p \geq 1$
- Trace norm: $\|X\|_{\mathrm{tr}} = \sum_{i=1}^r \sigma_i(X)$

We define its dual norm $\|x\|_*$ as

$$\|x\|_* = \max_{\|z\| \leq 1} \; z^T x$$

Gives us the inequality $|z^T x| \leq \|z\| \|x\|_*$, like Cauchy-Schwartz. Back to our examples,

- ℓ_p norm dual: $(\|x\|_p)_* = \|x\|_q$, where $1/p + 1/q = 1$
- Trace norm dual: $(\|X\|_{\mathrm{tr}})_* = \|X\|_{\mathrm{op}} = \sigma_{\max}(X)$

Dual norm of dual norm: it turns out that $\|x\|_{**} = \|x\|$ ... we'll see connections to duality (including this one) in coming lectures

References

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