Duality (đối ngẫu)

Hoàng Nam Dũng

Khoa Toán - Cơ - Tin học, Đại học Khoa học Tự nhiên, Đại học Quốc gia Hà Nội

Lagrangian

Consider general minimization problem

$$\begin{array}{ll} \min_x & f(x) \\ \text{subject to} & h_i(x) \le 0,\ i = 1, \ldots, m \\ & l_j(x) = 0,\ j = 1, \ldots, r. \end{array}$$

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

We define the Lagrangian as

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

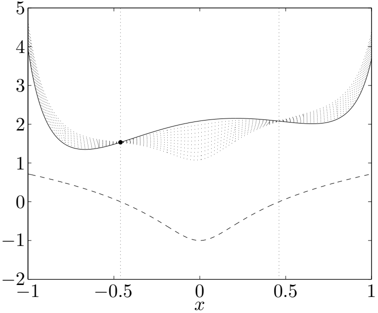
New variables $u \in \mathbb{R}^m, v \in \mathbb{R}^r$, with $u \ge 0$ (implicitly, we define $L(x, u, v) = -\infty$ for $u < 0$).

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,

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



- 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 choice of $u \geq 0$.

(From B & V page 127).

Lagrange dual function

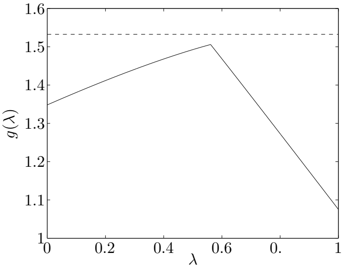
Let C denote primal feasible set, f^* denote primal optimal value. Minimizing $L(x, u, v)$ over all x gives a lower bound:

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

We call $g(u, v)$ the Lagrange dual function, and it gives a lower bound on f^* for any $u \geq 0$ and v, called dual feasible u, v.

- Dashed horizontal line is f^*.
- Dual variables λ is (our u).
- Solid line shows $g(\lambda)$.

  (From B & V page 127).



Example: quadratic program

Consider quadratic program:

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

where $Q \succ 0$. Lagrangian:

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

Lagrange dual function:

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

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

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:

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

Lagrange dual function:

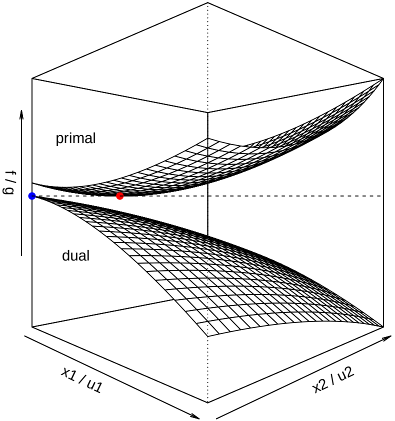
$$g(u, v) = \begin{cases} -\frac{1}{2}(c - u + A^Tv)^TQ^+(c - u + A^Tv) - b^Tv \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } c - u + A^Tv \perp \text{null}(Q) \\ -\infty \qquad\qquad\qquad\qquad\qquad \text{otherwise,} \end{cases}$$

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

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^* for every $u \geq 0$.

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

More on this later, via KKT conditions.

Lagrange dual problem

Given primal problem

$$\begin{array}{ll} \min_x & f(x) \\ \text{subject to} & h_i(x) \le 0, \; i = 1, \ldots, m \\ & l_j(x) = 0, \; j = 1, \ldots, r. \end{array}$$

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

$$\begin{array}{ll} \max_{u,v} & g(u, v) \\ \text{subject to} & u \ge 0. \end{array}$$

Key property, called weak duality: if dual optimal value is g^*, then

$$f^* \ge g^*.$$

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).

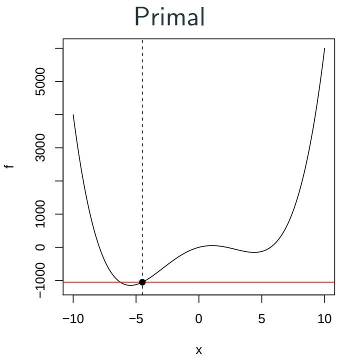
By definition:

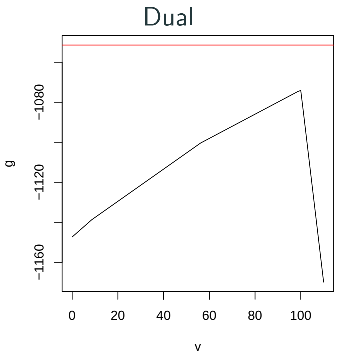
$$
\begin{aligned}
g(u,v) &= \min_x \left\{ f(x) + \sum_{i=1}^m u_i h_i(x) + \sum_{j=1}^r v_j l_j(x) \right\} \\
&= \underbrace{-\max_x \left\{ -f(x) - \sum_{i=1}^m u_i h_i(x) - \sum_{j=1}^r v_j l_j(x) \right\}}_{\text{pointwise maximum of convex functions in } (u,v)}
\end{aligned}
$$

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 constrains $x \geq -4.5$.





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 $i = 1, 2, 3$,

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

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

$$f^* = g^*,$$

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.

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:

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

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

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

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

$$\begin{array}{ll} \max_w & -\frac{1}{2} w^T \tilde{X} \tilde{X}^T w + 1^T w \\ \text{subject to} & 0 \leq w \leq C1, w^T y = 0. \end{array}$$

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

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

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 $f(x) - f^* \leq \epsilon$.

Very useful, especially in conjunction with iterative methods ... more dual uses in coming lectures.

References

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