University of California, Berkeley Lecturer: Benjamin Recht

CS227C/STAT260 Convex Optimization and Approximation: Optimization for Modern Data Analysis

March 3, 2016 Notes: Ben Recht and Ashia Wilson

Lecture 12: Duality Theory

1 Optimality Conditions for Equality Constrained Optimization

Recall that $x_\star$ minimizes a smooth, convex function f over a closed convex set Ω if and only if

$$\langle \nabla f(x_\star), x - x_\star \rangle \geq 0 \quad \forall x \in \Omega\,. \tag{1}$$

Let's specialize this to the special case where Ω is an affine set. Let A be an $n \times d$ matrix with rank n such that $\Omega = \{x \ : \ Ax = b\}$ for some $b \in \mathbb{R}^n$. Note that we can always assume that $\operatorname{rank}(A) = n$ or else we would have redundant constraints. We could also parameterize Ω as $\Omega = \{x_0 + v : Av = 0\}$ for any $x_0 \in \Omega$. Then using (1), we have

$$\langle \nabla f(x_\star), x - x_\star \rangle \geq 0 \ \ \forall x \in \Omega \quad \text{if and only if} \quad \langle \nabla f(x_\star), u \rangle \geq 0 \ \ \forall u \in \ker(A)\,.$$

But since $\ker A$ is a subspace, this can hold if and only if $\langle \nabla f(x_\star), u \rangle = 0$ for all $u \in \ker(A)$. In particular, this means, $\nabla f(x_\star)$ must lie in $\ker(A)^\perp$. Since we have that $\mathbb{R}^d = \ker(A) \oplus \operatorname{Im}(A^T)$, this means that $\nabla f(x_\star) = A^T\lambda$ for some $\lambda \in \mathbb{R}^n$.

To summarize, this means that $x_\star$ is optimal for f over Ω if and only if there $\exists \lambda_\star \in \mathbb{R}^m$ such that

$$\begin{cases} \nabla f(x_\star) + A^T\lambda_\star = 0 \\ Ax_\star = b \end{cases}$$

These optimality conditions are known as the *Karush-Kuhn-Tucker Conditions* or *KKT Conditions*.

As an example, consider the equality constrained quadratic optimization problem

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

The KKT conditions can be expressed in matrix form

$$\begin{bmatrix} Q & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} c \\ b \end{bmatrix}.$$

1.1 Nonlinear constraints

Suppose we want to minimize a smooth function over an intersection of an affine set and a closed convex set

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in X \\ & Ax = b \end{array} \tag{2}$$

where A is again a full rank $n \times d$ matrix. In this section, we will generalize (1) to show

Proposition 1. *$x_\star$ is optimal for* (2) *if and only if there exists a $\lambda_\star$ in $\mathbb{R}^n$ such that*

$$\begin{cases} \langle \nabla f(x_\star) + A^T\lambda_\star, x - x_\star\rangle \geq 0 \quad \forall x \in \Omega \\ Ax_\star = b \end{cases} .$$

The key to our analysis here will be to rely on convex analytic arguments. Let Ω be a closed convex set. Let's define the *tangent cone* of Ω at x as

$$\mathcal{T}_\Omega(x) = \operatorname{cone}\{z - x : z \in \Omega\}$$

The tangent cone is the set of all directions that can move from x and remain in Ω. We can also define the *normal cone* of Ω at x to be the set

$$\mathcal{N}_\Omega = \mathcal{T}_\Omega(x)^\circ = \{u \ : \ \langle u, v\rangle \leq 0 \ \forall v \in \mathcal{T}_\Omega(x)\} \ .$$

Note that when there is no equality constraint, our constrained optimality condition is completely equivalent to the assertion

$$-\nabla f(x_\star) \in \mathcal{N}_\Omega(x_\star) \,. \tag{3}$$

Thus, to prove Proposition 1, it will suffice for us to understand the normal cone of the set

$$\Omega \cap \{z \ : \ Az = b\}$$

at the point $x_\star$.

To obtain a reasonable characterization, we begin by proving a general fact.

Proposition 2. *Let Ω be a closed, convex set. Let $\mathcal{A}$ denote the affine set $\{x \ : \ Ax = b\}$. Suppose there $\exists x \in \operatorname{ri}(\Omega) \cap \mathcal{A}$. Then,*

$$\mathcal{N}_{\Omega\cap\mathcal{A}}(x) = \mathcal{N}_\Omega(x) + \{A^T\lambda \ : \ \lambda \in \mathbb{R}^n\} \,.$$

Proof. The "$\subseteq$" assertion is straightforward. To see this, suppose $z \in \Omega$ satisfies $z \in \operatorname{null}(A)$. Then if $u \in \mathcal{N}_\Omega(x)$ and $\lambda \in \mathbb{R}^n$, we have

$$\langle z - x, u + A^T\lambda\rangle = \langle z - x, u\rangle \leq 0$$

implying that $u + A^T\lambda \in \mathcal{N}_{\Omega\cap\mathcal{A}}(x)$.

For the reverse inclusion, let $v \in \mathcal{N}_{\Omega\cap\mathcal{A}}(x)$. Then we have

$$v^T(z - x) \leq 0$$

for all $z \in \Omega \cap \mathcal{A}$. Now define the sets

$$\begin{aligned} C_1 &= \left\{(y, \mu) \in \mathbb{R}^{d+1} \ : \ y = z - x, \ z \in \Omega, \ \mu \leq v^T y\right\} \\ C_2 &= \left\{(y, \mu) \in \mathbb{R}^{d+1} \ : \ y \in \ker(A), \ \mu = 0\right\} . \end{aligned}$$

Note that $\operatorname{ri}(C_1) \cap C_2 = \emptyset$ because otherwise there would exist a $(\hat{y}, \hat{\mu})$ such that

$$v^T\hat{y} > \hat{\mu} = 0$$

and $\hat{y} \in \mathcal{T}_{\Omega\cap\mathcal{A}}(x)$. This would contradict our assumption that $v \in \mathcal{N}_{\Omega\cap\mathcal{A}}(x)$. Since their intersection is empty, we can properly separate $\mathrm{ri}(C_1)$ from C_2. Indeed, since C_2 is a subspace and C_1 has nonempty relative interior, there must be a (w, γ) such that

$$\inf_{(y,\mu)\in C_1} \{w^T y + \gamma\mu\} < \sup_{(y,\mu)\in C_1} \{w^T y + \gamma\mu\} \leq 0$$

while

$$w^T u = 0$$

for all $u \in \ker(A)$. In particular, this means that $w = A^T\lambda$ for some $\lambda \in \mathbb{R}^n$.

Now, γ must be nonnegative, as otherwise,

$$\sup_{(y,\mu)\in C_1} \{w^T y + \gamma\mu\} = \infty$$

(which can be seen by letting μ tend to negative infinity). If $\gamma = 0$, then

$$\sup_{y\in C_1} w^T y \leq 0$$

but the set $\{y \ : \ w^T y = 0\}$ does not contain the set $\{z - x \ : \ z \in \Omega\}$ as the infimum is strictly negative. This means that the relative interior of $\Omega - \{x\}$ cannot intersect the kernel of A which contradicts our assumptions. Thus, we can conclude that γ is strictly positive. By homogeneity, we may as well assume that $\gamma = 1$.

To complete the argument, note that we now have

$$(w + v)^T (z - x) \leq 0$$

for all $z \in \Omega$. This means that $v + w \in \mathcal{N}_\Omega(x)$. And we have already shown that $w = A^T\lambda$. Thus,

$$v = (v + w) - w \in \mathcal{N}_\Omega(x) + \mathcal{N}_\mathcal{A}(x)\,.$$

$\square$

Let's now translate the consequence of this proposition for our problem. Using (3) and Proposition 2, we have that $x_\star$ is optimal for

$$\min f(x) \quad s.t \quad x \in \Omega, \quad Ax = b$$

if and only if $Ax_\star = b$ and there exists a $\lambda \in \mathbb{R}^n$ such that

$$\langle \nabla f(x_*) + A^T\lambda, x - x_* \rangle \geq 0 \quad \forall x \in \Omega\,.$$

This reduction is not immediately useful to us, as it doesn't provide an algorithmic approach to solving the constrained optimization problem. However, it will form the basis of our dive into duality.

2 Duality

Duality lets us associate to any constrained optimization problem, a concave maximization problem whose solutions lower bound the optimal value of the original problem. In particular, under mild assumptions, we will show that one can solve the primal problem by first solving the dual problem.

We'll continue to focus on the standard primal problem for an equality constrained optimization problem:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega \\ & Ax = b \end{array} \tag{4}$$

Here, assume that Ω is a closed convex set, f is differentiable, and A is full rank.

The key behind duality (here, Lagrangian duality) is that problem (4) is equivalent to

$$\min_{x \in \Omega} \max_{\lambda \in \mathbb{R}^n} f(x) + \lambda^T(Ax - b)$$

To see this, just note that if $Ax \neq b$, then the max with respect to λ is infinite. On the other hand, if $Ax = b$ is feasible, then the max with respect to λ is equal to $f(x)$.

The *dual problem* associated with (4) is

$$\max_{\lambda \in \mathbb{R}^n} \min_{x \in \Omega} f(x) + \lambda^T(Ax - b)$$

Note that the function

$$q(\lambda) := \min_{x \in \Omega} f(x) + \lambda^T(Ax - b)$$

is always a concave function as it is a minimum of linear functions. Hence the dual problem is a concave maximization problem, regardless of what form f and Ω take. We now show that it always lower bounds the primal problem.

2.1 Weak Duality

Proposition 3. *For any function $\varphi(x, z)$,*

$$\min_x \max_z \varphi(x, z) \geq \max_z \min_x \varphi(x, z)\,.$$

Proof. The proof is essentially tautological. Note that we always have

$$\varphi(x, z) \geq \min_x \varphi(x, z)$$

Taking the maximization with respect to the second argument verifies that

$$\max_z \varphi(x, z) \geq \max_z \min_x \varphi(x, z) \quad \forall x\,.$$

Now, minimizing the left hand side with respect to x proves

$$\min_x \max_z \varphi(x, z) \geq \max_z \min_x \varphi(x, z)$$

which is precisely our assertion. □

2.2 Strong duality

For convex optimization problems, we can prove a considerably stronger result. Namely, that the primal and dual problems attain the same optimal value. And, moreover, that if we know a dual optimal solution, we can extract a primal optimal solution from a simpler optimization problem.

Theorem 4 (Strong Duality)**.**

1. *If* $\exists z \in \mathrm{relint}(\Omega)$ *that also satisfies our equality constraint, and the primal problem has an optimal solution, then the dual has an optimal solution and the primal and dual values are equal*

2. *In order for* $x_\star$ *to be optimal for the primal and* $\lambda_\star$ *optimal for the dual, it is necessary and sufficient that* $Ax_\star = b$, $x_\star \in \Omega$ *and*

$$x_\star \in \arg\min_{x \in \Omega} \ \mathcal{L}(x, \lambda_\star) = f(x) + {\lambda_\star}^T (Ax - b)$$

Proof. For all λ and all feasible x

$$q(\lambda) \leq f(x) + \lambda(Ax - b) = f(x)$$

where the second equality holds because $Ax = b$.

Now by Proposition 1, $x_\star$ is optimal if and only if there exists a $\lambda_\star$ such that

$$\langle \nabla f(x_\star) + A^T \lambda_\star, x - x_\star \rangle \geq 0 \quad \forall x \in \Omega$$

and $Ax_\star = b$. Note that this condition means that $x_\star$ minimizes $\mathcal{L}(x, \lambda_\star)$ over Ω.

By preceding argument, it now follows that

$$\begin{aligned} q(\lambda_\star) &= \inf_{x \in \Omega} \mathcal{L}(x, \lambda_\star) \\ &= \mathcal{L}(x_\star, \lambda_\star) \\ &= f(x_\star) + {\lambda_\star}^T (Ax_\star - b) = f(x_\star) \end{aligned}$$

which completes the proof. □