
Lecture 16

For convex problems of the form

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{f}(\mathbf{x}) \leq \mathbf{0} \\ & && \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{1}$$

we introduced the Lagrangian $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ and defined the Lagrange dual as

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}).$$

We saw that $g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ is a lower bound on the optimal value of (1). Note that here we wrote the conditions $h_j(\mathbf{x}) = 0$ as system of linear equations $\mathbf{A}\mathbf{x} = \mathbf{b}$, since for the problem to be convex, we require that the h_j be linear functions.

Example 16.1. Consider a linear programming problem of the form

$$\begin{aligned} & \text{minimize} && \langle \mathbf{c}, \mathbf{x} \rangle \\ & \text{subject to} && \mathbf{A}\mathbf{x} = \mathbf{b} \\ & && \mathbf{x} \geq \mathbf{0}. \end{aligned}$$

The inequality constraints are $-x_i \leq 0$, while the equality constraints are $\mathbf{a}_i^\top \mathbf{x} - b_i$. The Lagrangian has the form

$$\begin{aligned} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) &= \langle \mathbf{c}, \mathbf{x} \rangle - \sum_{i=1}^n \lambda_i x_i + \sum_{j=1}^m \mu_j (\mathbf{a}_j^\top \mathbf{x} - b_j) \\ &= (\mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\mu})^\top \mathbf{x} - \mathbf{b}^\top \boldsymbol{\mu}. \end{aligned}$$

The infimum over \mathbf{x} of this function is $-\infty$ unless $\mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0}$. The Lagrange dual is therefore

$$g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \begin{cases} -\boldsymbol{\mu}^\top \mathbf{b} & \text{if } \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0} \\ -\infty & \text{else.} \end{cases}$$

From Lemma 15.5 we conclude that

$$\max\{-\mathbf{b}^\top \boldsymbol{\mu} : \mathbf{c} - \boldsymbol{\lambda} + \mathbf{A}^\top \boldsymbol{\mu} = \mathbf{0}, \boldsymbol{\lambda} \geq \mathbf{0}\} \leq \min\{\mathbf{c}^\top \mathbf{x} : \mathbf{Ax} = \mathbf{0}, \mathbf{x} \geq \mathbf{0}\}.$$

Note that if we write $\mathbf{y} = -\boldsymbol{\mu}$ and $\mathbf{s} = \boldsymbol{\lambda}$, then we get the dual version of the linear programming problem we started out with, and in this case we know that

$$\max_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) = p^*.$$

16.1 Constraint qualification

In the example of linear programming, we have seen that the optimal value of the dual problem is equal to the optimal value of the primal problem. In general, we have

$$d^* = \sup_{\boldsymbol{\lambda} \geq \mathbf{0}, \boldsymbol{\mu}} g(\boldsymbol{\lambda}, \boldsymbol{\mu}) \leq \inf_{\mathbf{x} \in \mathcal{D}} \{f(\mathbf{x}) : f_i(\mathbf{x}) \leq 0, \mathbf{Ax} = \mathbf{b}\} = p^*.$$

Once certain conditions, called *constraint qualifications*, hold, we can ensure that *strong duality* holds, which means $d^* = p^*$. One particular such constraint qualification is Slater's Theorem.

Theorem 16.2. (*Slater conditions*) Assume that the interior of the domain \mathcal{D} of (1) is non-empty, that the problem (1) is convex, and that there exists a point $\mathbf{x} \in \mathcal{D}$ such that

$$f_i(\mathbf{x}) < 0, \quad 1 \leq i \leq m, \quad \mathbf{Ax} = \mathbf{b}, \quad 1 \leq j \leq p.$$

Then $d^* = p^*$, the primal optimal value coincides with the dual optimal value.

Proof. (Optional) Assume \mathbf{A} has rank p (the number of rows). Assume moreover that p^* is finite, since if $p^* = -\infty$, then by weak duality we already have $d^* = p^*$.

Define the convex set

$$\mathcal{A} = \{(\mathbf{u}, \mathbf{v}, t) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : \forall \mathbf{x} \in \mathcal{D}, f_i(\mathbf{x}) \leq u_i, \mathbf{Ax} - \mathbf{b} = \mathbf{v}, f(\mathbf{x}) \leq t\}.$$

Then the optimal value of (1) is

$$p^* = \inf\{t : (\mathbf{0}, \mathbf{0}, t) \in \mathcal{A}\}.$$

Define the convex set \mathcal{B} as

$$\mathcal{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} : s < p^*\}.$$

The sets \mathcal{A} and \mathcal{B} are disjoint. To see this, assume to the contrary that there is a point $\mathbf{w} \in \mathcal{A} \cap \mathcal{B}$. Then since $\mathbf{w} \in \mathcal{B}$, $\mathbf{w} = (0, 0, s)$ with $s < p^*$, but also since $\mathbf{w} \in \mathcal{A}$, there exists an $\mathbf{x} \in \mathcal{D}$ with $f_i(\mathbf{x}) \leq 0$, $\mathbf{Ax} - \mathbf{b} = \mathbf{0}$, and $f(\mathbf{x}) \leq s < p^*$, in contradiction to the optimality of p^* as a value.

By the separating hyperplane theorem, there exist a hyperplane separating \mathcal{A} and \mathcal{B} (but not necessarily strictly!), defined by a vector $(\tilde{\lambda}, \tilde{\mu}, \nu) \neq \mathbf{0}$ and $\alpha \neq 0$ with the property that

$$(u, v, t) \in \mathcal{A} \implies \tilde{\lambda}^\top u + \tilde{\mu}^\top v + \nu t \geq \alpha \quad (16.1)$$

and

$$(u, v, t) \in \mathcal{B} \implies \tilde{\lambda}^\top u + \tilde{\mu}^\top v + \nu t \leq \alpha. \quad (16.2)$$

If $\tilde{\lambda} < \mathbf{0}$ or $\nu < 0$, we could make the right-hand side of (16.1) arbitrary small, contradicting the bound by α . It follows that $\tilde{\lambda} \geq \mathbf{0}$ and $\nu \geq 0$. Condition (16.2) simply means that $\nu t \leq \alpha$ for all $t < p^*$, so that $\nu p^* \leq \alpha$. Combining this bound with (16.1), we get the two inequalities, valid for any $\mathbf{x} \in \mathcal{D}$,

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) + \tilde{\mu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) + \nu f(\mathbf{x}) \geq \alpha \geq \nu p^*. \quad (16.3)$$

Note that the left-hand side has the form of a Lagrangian function scaled by ν . If $\nu > 0$ we can divide by ν and set $\lambda_i = \tilde{\lambda}_i/\nu$, $\mu_i = \tilde{\mu}_i/\nu$, to obtain

$$\mathcal{L}(\mathbf{x}, \lambda, \mu) \geq p^*.$$

By weak duality we have $p^* \geq g(\lambda, \mu)$, so that we get strong duality if $\nu > 0$. If $\nu = 0$, then (16.3) implies

$$\sum_{i=1}^m \tilde{\lambda}_i f_i(\mathbf{x}) + \tilde{\mu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) \geq 0, \quad (16.4)$$

and for a point $\tilde{\mathbf{x}}$ satisfying the conditions $\mathbf{A}\tilde{\mathbf{x}} = \mathbf{b}$ and $f_i(\tilde{\mathbf{x}}) < 0$ for $1 \leq i \leq m$, this means that $\tilde{\lambda} = \mathbf{0}$. As $(\tilde{\lambda}, \tilde{\mu}, \nu) \neq \mathbf{0}$, we have $\tilde{\mu} \neq \mathbf{0}$. Since $\tilde{\mu}^\top (\mathbf{A}\tilde{\mathbf{x}} - \mathbf{b}) \geq 0$ and $= 0$ for some $\tilde{\mathbf{x}}$ in the interior of \mathcal{D} , there must be a \mathbf{x} in the interior such that $\tilde{\mu}^\top (\mathbf{A}\mathbf{x} - \mathbf{b}) < 0$, in contradiction to (16.4) (unless $\tilde{\mu}^\top \mathbf{A} = \mathbf{0}$, which would contradict the condition that \mathbf{A} has maximal rank p). \square

Example 16.3. The problem minimize e^{-x} subject to $x^2/y \leq 0$, $y \geq 0$ is an example of a convex problem that does not satisfy strong duality.

Example 16.4. Consider the problem

$$\text{minimize } \mathbf{x}^\top \mathbf{x} \quad \text{subject to } \mathbf{A}\mathbf{x} = \mathbf{b}.$$

The Lagrangian is $\mathcal{L}(\mathbf{x}, \mu) = \mathbf{x}^\top \mathbf{x} + \mu^\top (\mathbf{A}\mathbf{x} - \mathbf{b})$. For any μ , we can find the infimum

$$g(\mu) = \inf_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu)$$

by setting the derivative of the Lagrangian to \mathbf{x} to zero:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \mu) = 2\mathbf{x} + \mathbf{A}^\top \mu = \mathbf{0},$$

which gives the solution

$$\mathbf{x} = -\frac{1}{2}\mathbf{A}^\top \boldsymbol{\mu}.$$

The dual function is therefore

$$g(\boldsymbol{\mu}) = -\frac{1}{4}\boldsymbol{\mu}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\mu} - \mathbf{b}^\top \boldsymbol{\mu}.$$

As the negative of a positive semidefinite quadratic function, it is concave. Moreover, we get the lower bound

$$-\frac{1}{4}\boldsymbol{\mu}^\top \mathbf{A}^\top \mathbf{A} \boldsymbol{\mu} - \mathbf{b}^\top \boldsymbol{\mu} \leq \inf\{\mathbf{x}^\top \mathbf{x} : \mathbf{A}\mathbf{x} = \mathbf{b}\}.$$

The problem we started out with is convex, and if we assume that there exists a feasible primal point, then the above inequality is in fact an equality by Slater's conditions.

16.2 Karush-Kuhn-Tucker optimality conditions

Consider now a not necessarily convex problem of the form

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{subject to} && \mathbf{f}(\mathbf{x}) \leq \mathbf{0} \\ & && \mathbf{h}(\mathbf{x}) = \mathbf{0}, \end{aligned} \tag{16.1}$$

If p^* is the optimal solution of (16.1) and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ dual variables, then we have seen that (this holds even in the non-convex case)

$$p^* \geq g(\boldsymbol{\lambda}, \boldsymbol{\mu}).$$

From this it follows that for any primal feasible point \mathbf{x} ,

$$f(\mathbf{x}) - p^* \leq f(\mathbf{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\mu}).$$

The difference $f(\mathbf{x}) - g(\boldsymbol{\lambda}, \boldsymbol{\mu})$ between the primal objective function at a primal feasible point and the dual objective function at a dual feasible point is called the *duality gap* at \mathbf{x} and $(\boldsymbol{\lambda}, \boldsymbol{\mu})$. For any such points we know that

$$p^*, q^* \in [g(\boldsymbol{\lambda}, \boldsymbol{\mu}), f(\mathbf{x})],$$

and if the gap is small we have a good approximation of the primal and dual optimal values. The duality gap can be used in iterative algorithms to define stopping criteria: if the algorithm generates a sequence of primal-dual variables $(\mathbf{x}^k, \boldsymbol{\lambda}^k, \boldsymbol{\mu}^k)$, then we can stop if the duality gap is less than, say, a predefined tolerance ε .

Now suppose that we have points $(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ such that the duality gap is zero. Then

$$\begin{aligned} f(\mathbf{x}^*) &= g(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) \\ &= \inf_{\mathbf{x}} \left(f(\mathbf{x}) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}) + \sum_{j=1}^p \mu_j^* h_j(\mathbf{x}) \right) \\ &\leq f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j^* h_j(\mathbf{x}^*) \\ &\leq f(\mathbf{x}^*), \end{aligned}$$

where the last inequality follows from the fact that $h_j(\mathbf{x}^*) = 0$ and $\lambda_i^* f_i(\mathbf{x}^*) \leq 0$ for $1 \leq j \leq p$ and $1 \leq i \leq m$. It follows that the inequalities are in fact equalities. From the identity

$$f(\mathbf{x}^*) = f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* f_i(\mathbf{x}^*)$$

and $\lambda_i^* \geq 0$ and $f_i(\mathbf{x}^*) \leq 0$ we also conclude that at such optimal points,

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad 1 \leq i \leq m.$$

This condition is known as *complementary slackness*. From the above we also see that \mathbf{x}^* minimizes the Lagrangian $\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$, so that the gradient of that function is zero:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*) = \mathbf{0}.$$

Collecting these conditions (primal and dual feasibility, complementary slackness, vanishing gradient), we arrive at a set of optimality conditions known as the Karush-Kuhn-Tucker (KKT) conditions.

Theorem 16.5. (*KKT conditions*) Let \mathbf{x}^* and $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ be primal and dual optimal solutions of (16.1) with zero duality gap. The the following conditions are satisfied:

$$\begin{aligned} \mathbf{f}(\mathbf{x}^*) &\leq \mathbf{0} \\ \mathbf{h}(\mathbf{x}^*) &= \mathbf{0} \\ \boldsymbol{\lambda}^* &\geq \mathbf{0} \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad 1 \leq i \leq m \\ \nabla_{\mathbf{x}} f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i^* \nabla_{\mathbf{x}} f_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j^* \nabla_{\mathbf{x}} h_j(\mathbf{x}^*) &= \mathbf{0}. \end{aligned}$$

Moreover, if the problem is convex, then any points satisfying the KKT conditions have zero duality gap.