7 Constrained optimization and duality

So far we have considered unconstrained optimization problems of the form $\min\{f(x) : x \in \mathbb{R}^n\}$, where f is convex and defined on the whole of $\mathbb{R}^n$. We consider in this lecture

$$\min_{x \in C} f(x) \tag{1}$$

where $C \subset \mathbb{R}^n$ is a convex set, and, for simplicity, $\mathbf{dom}(f) = \mathbb{R}^n$.

Optimality conditions We can write our problem as

$$\min_{x \in C} f(x) = \min_{x \in \mathbb{R}^n} f(x) + I_C(x)$$

where $I_C(x)$ is the indicator function of C, which takes the value 0 for $x \in C$, and $+\infty$ otherwise. The optimality condition for the latter is $0 \in \partial(f + I_C)(x)$. Under nice assumptions on C (e.g., $\mathbf{int}\, C \neq \emptyset$ or C is polyhedral, see Lecture 5) we have $\partial(f + I_C)(x) = \partial f(x) + \partial I_C(x)$. In this case, the optimality condition reads

$$0 \in \partial f(x) + \partial N_C(x).$$

We have seen that $\partial I_C(x)$ is nothing but the *normal cone* of C at x:

$$\partial I_C(x) = N_C(x) = \{g \in \mathbb{R}^n : \langle g, x\rangle \geq \langle g, y\rangle \ \forall y \in C\}.$$

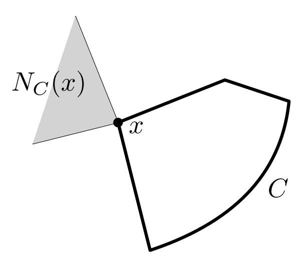

Figure 1: Normal cone

Assuming that f is smooth, a necessary and sufficient condition for x^* to be an optimal point of (1) is

$$\boxed{-\nabla f(x^*) \in N_C(x^*)}$$

Special case of linear constraints Consider the special case where C is a subspace, i.e. $C = \{x \in \mathbb{R}^n : Ax = b\}$ where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, so that the problem of interest is:

$$f^* = \min_{Ax=b} f(x). \tag{2}$$

Then one can easily verify that

$$N_C(x) = \mathbf{ker}(A)^\perp = \mathbf{im}(A^T) = \{A^T\lambda : \lambda \in \mathbb{R}^m\}.$$

(Note in particular that $N_C(x)$ is independent of x.) Thus in this case, a necessary and sufficient condition for x^* to solve $\min\{f(x) : Ax = b\}$ is that there exists $\lambda^* \in \mathbb{R}^m$ such that

$$\begin{cases} -\nabla f(x^*) = A^T \lambda^* \\ Ax^* = b. \end{cases} \tag{3}$$

The variable λ^* is the *Lagrange multiplier* or *dual variable*. The optimal dual variable λ^* happens to be the solution of another convex optimization problem, called the dual problem, which we now introduce.

The Lagrangian $L(x, \lambda)$ associated to the problem (2) is

$$L(x, \lambda) = f(x) + \langle \lambda, Ax - b \rangle .$$

Note that $\nabla_x L(x, \lambda) = \nabla f(x) + A^T \lambda$, and thus a minimizer of $L(x, \lambda)$ gives us a solution of the first equation in (3), with the particular choice of λ (note however that the solution $x^*(\lambda)$ is not guaranteed to satisfy the linear constraints!). This leads us to define the dual function

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

(Note that this is an unconstrained minimization.) An easy, yet important, observation here is the following:

$$\text{Weak duality:} \;\; g(\lambda) \leq f^* \qquad \forall \lambda \in \mathbb{R}^m.$$

Indeed

$$g(\lambda) = \min_{x \in \mathbb{R}^n} L(x, \lambda) \leq \min_{x : Ax = b} L(x, \lambda) = \min_{x : Ax = b} f(x)$$

where in the last equality we used the fact that $L(x, \lambda) = f(x)$ when $Ax = b$. Since $g(\lambda) \leq f^*$ for all λ, this means that $\max_\lambda g(\lambda) \leq f^*$. The problem of maximizing $g(\lambda)$ over λ is precisely the *dual problem*.

Primal problem	Dual problem
$\min_{Ax=b} f(x)$	$\max_{\lambda \in \mathbb{R}^m} g(\lambda) = \max_{\lambda \in \mathbb{R}^m} \min_{x \in \mathbb{R}^n} L(x, \lambda)$

Remark 1. *Note that the primal problem can be written equivalently as*

$$\min_{x \in \mathbb{R}^n} \max_{\lambda \in \mathbb{R}^m} L(x, \lambda).$$

This is because the inner maximization is equal to $f(x)$ if $Ax = b$ and $+\infty$ otherwise. Thus we see that the only difference between the primal and dual problems is the order of the min/max.

The equations (3) actually tell us that the primal and dual problems have the same optimal value; this is known as *strong duality*.

$$\text{Strong duality:} \;\; \max_{\lambda \in \mathbb{R}^m} g(\lambda) = f^*.$$

Indeed let x^* be the optimal solution of the primal problem, and let λ^* be the corresponding dual variable obtained from (3). Then note that $g(\lambda^*) = L(x^*, \lambda^*) = f(x^*)$ where the first equality follows from the fact that $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + A^T \lambda^* = 0$ (first equation in (3)) and the second equality follows from $Ax^* = b$ (second equation in (3)).