EE 227A: Convex Optimization and Applications **March 1, 2012**

Lecture 13: **Optimality Conditions for Convex Problems**

Lecturer: Laurent El Ghaoui

Reading assignment: §5.5 of BV.

13.1 Optimality Conditions

13.1.1 Complementary slackness

We consider a so-called "primal" convex optimization problem (without equality constraints for simplicity):

$$p^* := \min_x \ f_0(x) \ : \ f_i(x) \le 0, \ \ i = 1, \ldots, m,$$

and its dual

$$p^* \ge d^* := \max_\lambda \ g(\lambda),$$

where g is the dual function

$$g(\lambda) := \min_x \ \mathcal{L}(x, \lambda) \left(:= f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) \right).$$

We assume that the duality gap is zero: $p^* = d^*$, and that both primal and dual values are attained, by a primal-dual pair (x^*, λ^*). We have

$$p^* = f_0(x^*) = d^* = g(\lambda^*) = \min_x \ \mathcal{L}(x, \lambda^*) \le \mathcal{L}(x^*, \lambda^*) \le f_0(x^*) = p^*. \tag{13.1}$$

Thus, equalities hold in the above.

This implies that x^* minimizes the function $\mathcal{L}(\cdot, \lambda^*)$:

$$x^* \in \arg\min_x \mathcal{L}(x, \lambda^*).$$

If the functions $f_0, \ldots, f_m$ are differentiable, the above implies

$$\nabla_x \mathcal{L}(x, \lambda^*)\big|_{x=x^*} := \nabla f_0(x^*) + \sum_{i=1}^m \lambda_i^* \nabla f_i(x^*) = 0.$$

In addition, the equalities in (13.1) imply

$$\sum_{i=1}^m \lambda_i^* f_i(x^*) = 0.$$

Since $\lambda^* \ge 0$, $f_i(x^*) \le 0$, $i = 1, \ldots, m$, the above is equivalent to the *complementary slackness condition:*

$$\lambda_i^* f_i(x^*) = 0, \ \ i = 1, \ldots, m.$$

13.1.2 KKT optimality conditions

Assume that the functions $f_0, \ldots, f_m$ are differentiable.

Consider the so-called KKT[1] conditions on a primal-dual pair (x^*, λ^*).

$$\begin{array}{cr} f_i(x^*) \leq 0, \;\; i = 1, \ldots, m & \text{(primal feasibility)}, \\ \lambda^* \geq 0 & \text{(dual feasibility)}, \\ \lambda_i^* f_i(x^*) = 0, \;\; i = 1, \ldots, m & \text{(complementary slackness)}, \\ \nabla_x \mathcal{L}(x, \lambda^*)|_{x=x^*} = 0 & \text{(Lagrangian stationarity)}. \end{array} \tag{13.2}$$

The previous development shows that for any problem (convex or not) for which strong duality holds, and primal and dual values are attained, the KKT conditions (13.2) are *necessary* for a primal-dual pair (x^*, λ^*) to be optimal.

If, in addition the problem is convex, then the conditions are also *sufficient.* To see this, note that the first two conditions imply that x^*, λ^* are feasible for the primal and dual problems, respectively. Since $\mathcal{L}(\cdot, \lambda^*)$ is convex, the fourth condition (which we called Lagrangian stationarity) states that x^* is a minimizer of $\mathcal{L}(\cdot, \lambda^*)$, hence

$$g(\lambda^*) = \min_x \; \mathcal{L}(x, \lambda^*) = \mathcal{L}(x^*, \lambda^*) = f_0(x^*),$$

where we used the third condition (complementary slackness) for the last equality. The above proves that the primal-dual pair (x^*, λ^*) is optimal, since the corresponding gap is zero.

13.1.3 Primal solutions from dual variables

Assume that the problem has a zero duality gap, withdual values attained. Now assume that λ^* is optimal for the dual problem, and assume that the minimization problem

$$\min_x \; \mathcal{L}(x, \lambda^*). \tag{13.3}$$

has a unique solution $x(\lambda^*)$ that is feasible for the primal problem. Then, $x(\lambda^*)$ is optimal. Indeed, the fourth KKT condition (Lagrange stationarity) states that any optimal primal point minimizes the partial Lagrangian $\mathcal{L}(\cdot, \lambda^*)$, so it must be equal to the unique minimizer $x(\lambda^*)$.

This allows to compute the primal solution when a dual solution is known, by solving the above problem. Note that the requirement that $x(\lambda^*)$ be unique is necessary in our proof. Often, it is easier to make sure that the solution to (13.3) be unique for every (feasible) λ. This sufficient condition holds when the (primal) objective function is strictly convex.

[1] The acronym comes from the names Karush, Kuhn and Tucker, researchers in optimization around 1940-1960.

Exercises

1. *Sum of the largest elements in a vector.* For given $x \in \mathbf{R}^n$, and $k \in \{1, \ldots, n\}$ we define
$$s_k(x) := \sum_{i=1}^{k} x_{[i]},$$
where $x_{[i]}$ is the i-th largest element in x.

   (a) Show that
$$s_k(x) = \max_{0 \le u \le \mathbf{1}} u^T x \ : \ \sum_{i=1}^{n} u_i = k.$$
Is the function s_k is convex?

   (b) Formulate the dual to the above problem, and express the optimality conditions.

   (c) Show that
$$s_k(x) = \min_t \ kt + \sum_{i=1}^{n} \max(0, x_i - t).$$
Find an optimal value for t.

   (d) Why is the above representation useful when dealing with a constraint $s_k(x) \le S$ (with S a constant) in an optimization problem? Discuss.

2. *Feature elimination.* return to the "square-root" LASSO problem seen in lecture 8 (§8.3.4). Show that if $\lambda > \|a_i\|_2$, then $w_i = 0$ at optimum.

3. *Risk parity.* In portfolio optimization the estimation of returns is hard, while that of the covariance matrix maybe easier. Recently some authors have proposed to design a portfolio based only on the covariance, and using the notion of *risk parity.* We are given a $n \times n$ covariance matrix Σ, and seek to find a vector $x \in \mathbf{R}^n_+$, with $\mathbf{1}^T x = 1$, such that the vector of *partial risks* $r(x)$ with elements
$$r_i(x) = x_i (\Sigma x)_i, \quad i = 1, \ldots, n$$
is such that all its elements are all close to the same number. If this is the case, we say (loosely) that risk parity is achieved. Note that the term "risk" used here is a misnomer, as the partial risks may be negative. Show that solving the problem
$$\max_x \sum_{i=1}^{L} \log x_i - x^T \Sigma x$$
results in a portfolio that achieves exact risk parity. Make sure to justify your derivations rigorously.