10-801: Advanced Optimization and Randomized Methods

Homework 3: KKT conditions, optimality, concentration of measure

(Feb 12, 2014)

Instructor: *Suvrit Sra, Alex Smola* ***Due:*** *Feb 24, 2014*

1. For two scalars γ, d, with $d > 0$, we define
$$B(\gamma, d) := \min_{\lambda \geq 0} \lambda + \frac{\gamma^2}{\lambda + d}.$$
   (a) Show that
$$B(\gamma, d) = \begin{cases} \dfrac{\gamma^2}{d} & \text{if } |\gamma| \leq d, \\ 2|\gamma| - d & \text{if } |\gamma| \geq d. \end{cases}$$
   with minimizer $\lambda^* = \max(0, |\gamma| - d)$.
   (b) Is B convex? Justify your answer rigorously.
   (c) Find a sub-gradient of B at a given point (γ, d), with $d > 0$.
   (d) The function B, often referred to as the Huber function, is sometimes used as a penalty in classification or regression, in problems of the form
$$\min_x L(x) + B(x, d),$$
   where L is a loss function, and $d > 0$ is a (now fixed) parameter. Explain intuitively what effect the penalty B has on the solution, depending on the parameter d.

2. Consider the optimization problem
$$\min_x \sum_{i=1}^n \left(\frac{1}{2} d_i x_i^2 + r_i x_i \right) \ : \ a^T x = 1, \ \ x_i \in [-1, 1], \ \ i = 1, \ldots, n,$$
where $a, d \in \mathbb{R}^n$, with , and $d > 0$.
   (a) Show that the problem is strictly feasible if and only if $\|a\|_1 > 1$, which we will henceforth assume. *Hint:* consider the problem
$$\min_x \|x\|_\infty \ : \ a^T x = 1,$$
   and its dual.
   (b) Write a dual for the problem, expressing the last n constraints as $x_i^2 \leq 1$, $i = 1, \ldots, n$.
   (c) Does strong duality hold? Justify your answer.
   (d) Write down the KKT optimality conditions for the problem. Do these conditions characterize optimal points?
   (e) Show how to reduce the dual problem to a one-dimensional convex problem, of the form
$$\min_\mu \mu + \frac{1}{2} \sum_{i=1}^n B(r_i + \mu a_i, d_i),$$
   where B is the Huber function defined in part 1.
   (f) Suggest an algorithm to solve the dual problem. Analyze its running time complexity.
   (g) How can you recover an optimal primal point x, after solving the dual?

3. We examine the problem of fitting a polynomial of degree d through data points $(u_i, y_i) \in \mathbb{R}^2$, $i = 1, \ldots, m$. Without loss of generality, we assume that the input satisfies $|u_i| \leq 1$, $i = 1, \ldots, m$. We parametrize a polynomial of degree d via its coefficients:
$$p_w(u) = w_0 + w_1 u + \ldots + w_d u^d,$$
where $w \in \mathbb{R}^{d+1}$. Our problem is to minimize, over the vector w, the error norm
$$\sum_{i=1}^m (p_w(u_i) - y_i)^2.$$

(a) Show that the problem can be written as

$$\min_{w} \|\Phi^T w - y\|_2^2,$$

where the matrix Φ has columns $\phi_i = (1, u_i, \ldots, u_i^d)$, $i = 1, \ldots, m$.

(b) In practice it is desirable to encourage polynomials that are not too rapidly varying over the interval of interest. To that end, we modify the above problem as follows:

$$\min_{w} \|\Phi^T w - y\|_2^2 + \lambda b(w), \tag{3.1}$$

where $\lambda > 0$ is a regularization parameter, and $b(w)$ is a bound on the size of the derivative of the polynomial over $[-1, 1]$:

$$b(w) = \max_{u \,:\, |u| \le 1} \left| \frac{d}{du} p_w(u) \right|.$$

Is the penalty function b convex? Is it a norm?

(c) Explain how to compute a subgradient of b at a point w.

(d) Write KKT optimality conditions for (3.1).

4. [Self-bounding concentration of measure inequalities for random partitioning]

Denote by b the number of bins and by n the number of elements. Now assume that we use a uniformly random assignment of elements to bins. Denote by n_i the number of elements per bin. Prove concentration of measure inequalities for the following quantities:

(a) Largest bin size, $\max_i n_i$.

(b) Variance of bin sizes, $\frac{1}{b} \sum_i (n_i - \bar{n_i})^2$.

Hint: Use the McDiarmid and Reed self-bounding inequality.

5. [Expected quantile rank of sample minimum]

Assume that we draw k random variables x_i *iid* from a distribution p. We pick the smallest value

$$z_k := \min(x_1, ...x_k).$$

Given the cumulative distribution function F, what is $\mathbb{E}[F(z_k)]$ – the expected quantile rank of z_k?

6. [*Bonus*] Let $\mathcal{D}_n$ be the set of matrices with entries in $[0, 1]$.

(a) For odd n, prove that for every nonsingular A in $\mathcal{D}_n$, it holds that $\|A^{-1}\|_{\mathrm{F}} \ge \frac{2n}{n+1}$, where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm.

(b) Determine the class of matrices for which equality holds. Partial credit will be given for examples of matrices for which equality holds.

(c) Prove the claim for all n.