ELE 538B: Sparsity, Structure and Inference Spring 2017

Homework 3

Due date: Wednesday, Apr. 26, 2017 (at the beginning of class)

You are allowed to drop 1 subproblem without penalty. In addition, up to 1 bonus point will be awarded to each subproblem for clean, well-organized, and elegant solutions.

1. Proximal minimization (*40 points*)

Recall that the proximal operator of a convex function h is defined as

$$\mathsf{prox}_h(\boldsymbol{x}) \;:=\; \arg\min_{\boldsymbol{z}} \left\{ \frac{1}{2} \|\boldsymbol{x} - \boldsymbol{z}\|^2 + h(\boldsymbol{z}) \right\}$$

(a) Suppose that $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_2$. Show that

$$\mathsf{prox}_{\lambda f}(\boldsymbol{x}) \;:=\; \left(1 - \frac{\lambda}{\|\boldsymbol{x}\|_2}\right)_+ \boldsymbol{x},$$

where $(a)_+ := \max\{a, 0\}$.

(b) Suppose that $f(\boldsymbol{x}) = h(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{a}\|^2$. Show that

$$\mathsf{prox}_{\lambda f}(\boldsymbol{x}) \;:=\; \mathsf{prox}_{\frac{\lambda}{1+\lambda\rho} h}\left(\frac{1}{1+\lambda\rho}\boldsymbol{x} + \frac{\lambda\rho}{1+\lambda\rho}\boldsymbol{a}\right).$$

(c) Suppose that $f(\boldsymbol{x}) = h(\boldsymbol{x}) + \boldsymbol{a}^\top \boldsymbol{x} + \boldsymbol{b}$. Show that

$$\mathsf{prox}_{\lambda f}(\boldsymbol{x}) \;:=\; \mathsf{prox}_{\lambda h}(\boldsymbol{x} - \lambda \boldsymbol{a}).$$

(d) Show that a point $\boldsymbol{x}^*$ is the minimizer of $h(\cdot)$ if and only if

$$\boldsymbol{x}^* = \mathsf{prox}_h(\boldsymbol{x}^*).$$

This simple observation is the motivation of the so-called *proximal minimization algorithm*, which finds the optimizer of h by the iterative procedure

$$\boldsymbol{x}^{t+1} = \mathsf{prox}_{\lambda h}(\boldsymbol{x}^t).$$

2. Restricted isometry properties (*30 points*)

Recall that the restricted isometry constant $\delta_s \geq 0$ of $\boldsymbol{A}$ is the smallest constant such that

$$(1 - \delta_s)\|\boldsymbol{x}\|_2^2 \leq \|\boldsymbol{A}\boldsymbol{x}\|_2^2 \leq (1 + \delta_s)\|\boldsymbol{x}\|_2^2 \tag{1}$$

holds for all s-sparse vector $\boldsymbol{x} \in \mathbb{R}^p$.

(a) Show that

$$|\langle \boldsymbol{A}\boldsymbol{x}_1, \boldsymbol{A}\boldsymbol{x}_2\rangle| \le \delta_{s_1+s_2}\|\boldsymbol{x}_1\|_2\|\boldsymbol{x}_2\|_2$$

for all pairs of $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ that are supported on disjoint subsets $S_1, S_2 \subset \{1, \cdots, n\}$ with $|S_1| \le s_1$ and $|S_2| \le s_2$.

(b) For any $\boldsymbol{u}$ and $\boldsymbol{v}$, show that

$$|\langle \boldsymbol{u},\ (\boldsymbol{I} - \boldsymbol{A}^\top \boldsymbol{A})\boldsymbol{v}\rangle| \le \delta_s\|\boldsymbol{u}\| \cdot \|\boldsymbol{v}\|,$$

where s is the cardinality of support $(\boldsymbol{u}) \cup$ support $(\boldsymbol{v})$.

(c) Suppose that each column of $\boldsymbol{A}$ has unit norm. Show that $\delta_2 = \mu(\boldsymbol{A})$, where $\mu(\boldsymbol{A})$ is the mutual coherence of $\boldsymbol{A}$.

3. Statistical dimension (*10 points*) Recall that for any convex cone $\mathcal{K}$, its statistical dimension and Gaussian width are defined respectively as

$$\text{stat-dim}(\mathcal{K}) := \mathbb{E}\big[\|\mathcal{P}_{\mathcal{K}}(\boldsymbol{g})\|^2\big]$$

and

$$w(\mathcal{K}) := \mathbb{E}\left[\sup_{\boldsymbol{z}\in\mathcal{K}, \|\boldsymbol{z}\|=1} \langle \boldsymbol{z}, \boldsymbol{g}\rangle\right],$$

where $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$ and $\mathcal{P}_{\mathcal{K}}$ denotes the projection to $\mathcal{K}$ as

$$\mathcal{P}_{\mathcal{K}}(\boldsymbol{g}) = \arg\min_{\boldsymbol{z}\in\mathcal{K}} \|\boldsymbol{g} - \boldsymbol{z}\|.$$

(a) Prove that $w^2(\mathcal{K}) \le \text{stat-dim}(\mathcal{K})$.

(b) (Optional (10 bonus points)) Prove the reverse inequality $\text{stat-dim}(\mathcal{K}) \le w^2(\mathcal{K}) + 1$.
hint: Let $f(\cdot)$ be a function that is Lipschitz with respect to the Euclidean norm:

$$|f(\boldsymbol{u}) - f(\boldsymbol{v})| \le M\|\boldsymbol{u} - \boldsymbol{v}\| \qquad \forall \boldsymbol{u}, \boldsymbol{v}.$$

Then, $\text{Var}(f(\boldsymbol{g})) \le M^2$.