ECE 18-898G: Sparsity, Structure and Inference Spring 2018

Homework 3

Due date: Wednesday, Mar. 7, 2018 (in class)

1. Proximal methods (*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. Iterative Hard Thresholding (*30 points*) (Foucart and Rauhut, Problem 6.21)

Let $\boldsymbol{x} \in \mathbb{R}^p$ be a s-sparse vector, and given $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ for some measurement matrix $\boldsymbol{A}$. Denote 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 $\boldsymbol{x} \in \mathbb{R}^p$.

Assume we are given a sequence of iterates $\boldsymbol{x}_n$, as

$$\boldsymbol{x}_{n+1} = H_s(\boldsymbol{x}_n + \mu \boldsymbol{A}^\top(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}_n)) \tag{2}$$

where $\boldsymbol{x}_0$ is an initial s-sparse vector, and the hard thresholding operator H_s keeps the s largest absolute entries of a vector. This is the iterative hard thresholding algorithm discussed in class. We will determine μ later.

(a) Establish the identity

$$\|\boldsymbol{A}(\boldsymbol{x}_{n+1}-\boldsymbol{x})\|_2^2 - \|\boldsymbol{A}(\boldsymbol{x}_n-\boldsymbol{x})\|_2^2 = \|\boldsymbol{A}(\boldsymbol{x}_{n+1}-\boldsymbol{x}_n)\|_2^2 + 2\langle \boldsymbol{x}_n-\boldsymbol{x}_{n+1}, \boldsymbol{A}^\top \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_n)\rangle$$

(b) Establish the inequality

$$2\mu\langle \boldsymbol{x}_n-\boldsymbol{x}_{n+1}, \boldsymbol{A}^\top \boldsymbol{A}(\boldsymbol{x}-\boldsymbol{x}_n)\rangle \le \|\boldsymbol{x}_n-\boldsymbol{x}\|_2^2 - 2\mu\|\boldsymbol{A}(\boldsymbol{x}_n-\boldsymbol{x})\|_2^2 - \|\boldsymbol{x}_{n+1}-\boldsymbol{x}_n\|_2^2.$$

(c) Derive the inequality

$$\|\boldsymbol{A}(\boldsymbol{x}_{n+1}-\boldsymbol{x})\|_2^2 \le \left(1-\frac{1}{\mu(1+\delta_{2s})}\right)\|\boldsymbol{A}(\boldsymbol{x}_{n+1}-\boldsymbol{x}_n)\|_2^2 + \left(\frac{1}{\mu(1-\delta_{2s})}-1\right)\|\boldsymbol{A}(\boldsymbol{x}_n-\boldsymbol{x})\|_2^2.$$

Deduce that the sequence $\boldsymbol{x}_n$ converges to $\boldsymbol{x}$ when $1+\delta_{2s} < \frac{1}{\mu} < 2(1-\delta_{2s})$. Conclude by justifying the choice $\mu = 3/4$ under the condition $\delta_{2s} < 1/3$.

3. Subgradient of nuclear norm (*30 points*)

The nuclear norm to low-rank matrix recovery plays a similar role as the ℓ_1 norm to sparse recovery.

(a) Find the subgradient of the nuclear norm.

(b) Use (a) to find the optimality condition of a nuclear-norm regularized optimization problem:

$$\min_{\boldsymbol{X}\in\mathbb{R}^{n\times n}} \|\boldsymbol{y}-\mathcal{A}(\boldsymbol{X})\|_2^2 + \lambda\|\boldsymbol{X}\|_*$$

where $\mathcal{A}() : \mathbb{R}^{n\times n} \mapsto \mathbb{R}^m$ is a linear operator, $\boldsymbol{y}\in\mathbb{R}^m$, and $\lambda > 0$ is a regularization parameter.