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

Homework 2

Due date: Wednesday, Feb. 21, 2017 (in class)

1. Restricted isometry properties (*20 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

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

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 $|S_1| \leq s_1$ and $|S_2| \leq s_2$.

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

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

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

2. Sparsity for model selection (*25 points*)

Suppose that the observation $\boldsymbol{y} \in \mathbb{R}^n$ obeys

$$\boldsymbol{y} = \boldsymbol{\beta} + \boldsymbol{\eta}, \qquad \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}),$$

i.e. each entry of $\boldsymbol{y}$ is corrupted by i.i.d. Gaussian noise.

(a) Suppose we want to estimate β by projecting it to a sparse vector $\hat{\boldsymbol{\beta}}_S$ that is supported on a fixed subset $S \subset \{1, \ldots, n\}$, where the estimate is given as

$$(\hat{\boldsymbol{\beta}}_S)_i = \begin{cases} y_i, & \text{if } i \in S \\ 0, & \text{else} \end{cases}$$

What is the mean squared error (MSE) $\mathbb{E}[\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_S\|_2^2]$ for a fixed S?

(b) If we fix the model size $|S| = k$, what is the subset S that achieves the minimum MSE?

(c) Suppose we want to further minimize the MSE over all possible model size k, what is the minimum MSE? Explain when we prefer small model size k.

3. Lasso with a single parameter (*25 points*)

Consider the single parameter setting $\boldsymbol{y} = \beta \boldsymbol{z} + \boldsymbol{\eta}$ with $\beta \in \mathbb{R}$. In this case, the Lasso estimator is given by

$$\text{minimize}_{\hat{\beta} \in \mathbb{R}} \quad \frac{1}{2}\|\boldsymbol{y} - \hat{\beta}\boldsymbol{z}\|^2 + \lambda|\hat{\beta}|.$$

Show that $\hat{\beta} = \psi_{\mathsf{st}}\left(\frac{\boldsymbol{z}^\top \boldsymbol{y}}{\|\boldsymbol{z}\|^2}; \frac{\lambda}{\|\boldsymbol{z}\|^2}\right)$ is a closed-form solution to the above program, where $\psi_{\mathsf{st}}(x;\lambda) = \operatorname{sign}(x)\max\{|x| - \lambda, 0\}$ is the soft-thresholding operator. You should use the optimality condition based on subgradients.

4. Convexity of the SLOPE estimator (*30 points*)

The SLOPE (Sorted L-One Penalized Estimation) estimator is recently proposed and shown to have the remarkable property of being adaptive to unknown sparsity and asymptotically minimax. (W. Su and E. J. Candès, Annals of Statistics 44(3), pp. 1038–1068.) In this problem, we study the formulation of SLOPE and show it is a convex program.

For any $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_p]^\top \in \mathbb{R}^p$, let $|\beta|_{(1)} \geq |\beta|_{(2)} \geq \cdots \geq |\beta|_{(p)}$ denote the order statistics of $\{|\beta_1|, \cdots, |\beta_p|\}$, i.e. $|\beta|_{(i)}$ is the ith largest in $\{|\beta_1|, \cdots, |\beta_p|\}$.

(a) Suppose that $p = 2$. Show that the function

$$f(\boldsymbol{\beta}) := \lambda_1 |\beta|_{(1)} + \lambda_2 |\beta|_{(2)}$$

is convex if $\lambda_1 \geq \lambda_2 \geq 0$.

(b) Show that the function

$$g_k(\boldsymbol{\beta}) = \sum_{i=1}^{k} |\beta|_{(i)}$$

is convex for any $1 \leq k \leq p$.

(c) Show that the function

$$f(\boldsymbol{\beta}) := \sum_{i=1}^{p} \lambda_i |\beta|_{(i)}$$

is convex for any $p \geq 3$, as long as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. This justifies that SLOPE

$$\text{minimize}_{\boldsymbol{\beta}} \quad \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}\|^2 + \sum_{i=1}^{p} \lambda_i |\beta|_{(i)}$$

is a convex program.