36-755, Fall 2017
Homework 4

Due Wed Nov 1 by 5:00pm in Jisu's mailbox

1. In earlier works on the lasso, people have used a even stronger assumptions than the restricted eigenvalue property. Here is one. Suppose that the design matrix X is such that, for some integer $k > 0$,
$$\max_{i,j} \left| \frac{X_i^\top X_j}{n} - 1(i = j) \right| \leq \frac{1}{23k} \tag{1}$$
where X_i is the ith column of X, $i = 1, \ldots, d$. Think about what that means.

   (a) Show that this condition implies that, for any subset S of $\{1, \ldots, d\}$ of cardinality no larger than $k < d$ and any $\Delta \in \mathbb{R}^d$ with $\|\Delta_{S^c}\|_1 \leq 3\|\Delta_S\|_1$,
   $$\|\Delta\|^2 \leq \frac{2}{n}\|X\Delta\|^2.$$
   That is, show that this condition implies the $RE(3, 1/2)$ condition given in class for all non-empty subsets S of $\{1, \ldots, d\}$ of size no larger than k. *Instead of the constant* 23 *you may take a larger one if it simplifies your calculations.*

   (b) Suppose that the entries of X are now populated by independent Rademacher variables (a Ra detacher variable is one that that takes the values $+1$ and -1 with equal probability). Show that, for any $\delta \in (0, 1)$, if
   $$n \geq Ck^2(\log(d) + \log(1/\delta)),$$
   for some constant $C > 0$, then X satisfies the condition (1), with probability at least $1 - \delta$. *Again, instead of* 23 *feel free to show the result for a different constant if it helps with the calculations.*

2. Read the paper "p-Values for High-Dimensional Regression" by Nicolai Meinshausen, Lukas Meier and Peter Bühlmann, JASA 2009, volume 104, issue 448, pages 1671-1681. Reproduce the proof of Theorem 3.2.

3. **Performance of the best selection procedure.** Consider the regression framework $Y = X\beta^* + \epsilon$, where X is a $n \times d$ deterministic design matrix, and ϵ a vector in $\mathbb{R}^n$ of independent $SG(\sigma^2)$ error variables. Assume that the true regression coefficient β^* belongs to the set $S_0(k) = \{x \in \mathbb{R}^d \colon, \|x\|_0 = k\}$ of k-sparse vectors, where $k \leq d$. Consider the estimator
$$\hat{\beta} = \mathrm{argmin}_{\beta \in B_0(k)} \|Y - X\beta\|^2.$$
This the best least squares solution computed over all subsets of the coordinates of size k. Computationally, it requires evaluating $\binom{d}{k}$ least squares. Analyze the performance of $\hat{\beta}$ by showing that, with probability at lest $1 - \delta$
$$\|X(\hat{\beta} - \beta^*)\|^2 \leq C(\delta)\frac{\sigma^2 k}{n} \log\left(\frac{ed}{2k}\right),$$
where $C(\delta)$ is a constant that depend on δ. Notice that, up to a logarithmic term, this is the (optimal) performance of the least squares estimator *if the support of* β^* *were known.* This is something that

is quite typical: the statistical price for not knowing the support of of β^* is only logarithmic (and therefore rather minimal). However, at least for the estimator $\hat{\beta}$, the computational price is huge. The trade-off between computational and statistical guarantees in a very important topic in the theoretical literature on high-dimensional statistics.

Hint: follow the proof of the performance of the least squares estimator. You may want to use the fact that $\binom{d}{k} \le \left(\frac{ed}{k}\right)^k$.

4. **Matrix Algebra Problems.**

   (a) Problem 8.3 (You may assume the result of Problem 8.1 as given).

   (b) Recall the spiked covariance model: $\Sigma = \theta v v^\top + I_d$, where $\theta > 0$ and $v \in \mathbb{S}^{d-1}$. Let $\hat{v}$ be another unit vector in $\mathbb{S}^{d-1}$. Show that

   $$v^\top \Sigma v - \hat{v}^\top \Sigma \hat{v} = \theta \sin^2(\angle(v, \hat{v}))$$

   where $\angle(v, \hat{v}) = \cos^{-1}(|v^\top \hat{v}|)$

   (c) Show that

   $$\left\| \hat{v}\hat{v}^\top - vv^\top \right\|_F^2 = 2\sin^2(\angle(v, \hat{v})),$$

   where, for a matrix $A = (A_{i,j})$, $\|A\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$.