ELE 538B: Sparsity, Structure and Inference Spring 2017

Homework 1

Due date: Wednesday, Mar. 1, 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. Mutual coherence (*40 points*)

Recall that for an arbitrary pair of orthonormal bases $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \cdots, \boldsymbol{\psi}_n] \in \mathbb{R}^{n\times n}$ and $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_n] \in \mathbb{R}^{n\times n}$, the mutual coherence $\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi})$ of these two bases is defined by

$$\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) = \max_{1\le i,j\le n} \left| \boldsymbol{\psi}_i^\top \boldsymbol{\phi}_j \right| \tag{1}$$

(a) Show that

$$\frac{1}{\sqrt{n}} \le \mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) \le 1.$$

(b) Let $\boldsymbol{\Psi} = \boldsymbol{I}$, and suppose that $\boldsymbol{\Phi} = [\phi_{i,j}]_{1\le i,j\le n}$ is a Gaussian random matrix such that the $\phi_{i,j}$'s are i.i.d. random variables drawn from $\phi_{i,j} \sim \mathcal{N}(0, 1/n)$. Can you provide an upper estimate on $\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi})$ as defined in (1)? Since $\boldsymbol{\Phi}$ is a random matrix, we expect your answer to be a function $f(n)$ such that $\mathbb{P}\{\mu(\boldsymbol{\Psi}, \boldsymbol{\Phi}) > f(n)\} \to 0$ as n scales.

Hint: to simplify analysis, you are allowed to use the crude approximation $\mathbb{P}\{|z| > \tau\} \approx \exp(-\tau^2/2)$ for large $\tau > 0$, where $z \sim \mathcal{N}(0, 1)$.

(c) Set $n = 100$. Generate a random matrix $\boldsymbol{\Phi}$ as in Part (b), and compute $\mu(\boldsymbol{I}, \boldsymbol{\Phi})$. Report the empirical distribution (i.e. histogram) of $\mu(\boldsymbol{I}, \boldsymbol{\Phi})$ out of 1000 simulations. How does your simulation result compare to your estimate in Part (b)?

(d) We now generalize the mutual coherence measure to accommodate a more general set of vectors beyond two bases. Specifically, for any given matrix $\boldsymbol{A} = [\boldsymbol{a}_1, \cdots, \boldsymbol{a}_p] \in \mathbb{R}^{n\times p}$ obeying $n \le p$, define the mutual coherence of $\boldsymbol{A}$ as

$$\mu(\boldsymbol{A}) = \max_{1\le i,j\le p,\ i\ne j} \left| \frac{\boldsymbol{a}_i^\top \boldsymbol{a}_j}{\|\boldsymbol{a}_i\| \|\boldsymbol{a}_j\|} \right|.$$

Show that

$$\mu(\boldsymbol{A}) \ge \sqrt{\frac{p-n}{p-1} \cdot \frac{1}{n}}.$$

This is a special case of the Welch bound.

Hint: you may want to use the following inequality: for any positive semidefinite $\boldsymbol{M} \in \mathbb{R}^{n\times n}$, $\|\boldsymbol{M}\|_{\mathrm{F}}^2 \ge \frac{1}{n}\left(\sum_{i=1}^n \lambda_i(\boldsymbol{M})\right)^2$.

2. Picket-fence signal (*10 points*)

Suppose that $\sqrt{n}$ is an integer. Let $\boldsymbol{x} \in \mathbb{R}^n$ be a picket-fence signal with uniform spacing $\sqrt{n}$ such that

$$x_i = \begin{cases} 1, & \text{if } \frac{i-1}{\sqrt{n}} \text{ is an integer,} \\ 0, & \text{else,} \end{cases} \qquad i = 1, \cdots, n.$$

Compute

$$\|\boldsymbol{x}\|_0 \cdot \|\boldsymbol{F}\boldsymbol{x}\|_0 \quad \text{and} \quad \|\boldsymbol{x}\|_0 + \|\boldsymbol{F}\boldsymbol{x}\|_0,$$

where $\boldsymbol{F}$ is the Fourier matrix such that

$$(\boldsymbol{F})_{k,l} = \frac{1}{\sqrt{n}} \exp\left(-i\frac{2\pi(k-1)(l-1)}{n}\right), \qquad 1 \le k, l \le n.$$

How do they compare to the uncertainty principles we derive in class?

3. ℓ_1 minimization (*20 points*)

Suppose that $\boldsymbol{A}$ is an $n \times 2n$ dimensional matrix. Let $\boldsymbol{x} \in \mathbb{R}^{2n}$ be an unknown k-sparse vector, and $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ the observed system output. This problem is concerned with ℓ_1 minimization (or basis pursuit) in recovering $\boldsymbol{x}$, i.e.

$$\text{minimize}_{\boldsymbol{z}\in\mathbb{R}^{2n}} \quad \|\boldsymbol{z}\|_1 \quad \text{s.t. } \boldsymbol{A}\boldsymbol{z} = \boldsymbol{y}. \tag{2}$$

(a) An optimization problem is called a linear program (LP) if it has the form

$$\begin{aligned} \text{minimize}_{\boldsymbol{z}} \quad & \boldsymbol{c}^\top \boldsymbol{z} + \boldsymbol{d} \\ \text{s.t.} \quad & \boldsymbol{G}\boldsymbol{z} \le \boldsymbol{h} \\ & \boldsymbol{A}\boldsymbol{z} = \boldsymbol{b} \end{aligned}$$

where $\boldsymbol{c}, \boldsymbol{d}, \boldsymbol{G}, \boldsymbol{h}, \boldsymbol{A}$, and $\boldsymbol{b}$ are known. Here, for any two vectors $\boldsymbol{r}$ and $\boldsymbol{s}$, we say $\boldsymbol{r} \le \boldsymbol{s}$ if $r_i \le s_i$ for all i. Show that (2) can be converted to a linear program.

(b) Set $n = 256$, and let k range between 1 and 128. For each choice of k, run 10 independent numerical experiments: in each experiment, generate $\boldsymbol{A} = [a_{i,j}]_{1\le i\le n, 1\le j\le 2n}$ as a random matrix such that the $a_{i,j}$'s are i.i.d. standard Gaussian random variables, generate $\boldsymbol{x} \in \mathbb{R}^{2n}$ as a random k-sparse signal (e.g. you may generate the support of $\boldsymbol{x}$ uniformly at random, with each non-zero entry drawn from the standard Gaussian distribution), and solve (2) with $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$. An experiment is claimed successful if the solution $\boldsymbol{z}$ returned by (2) obeys $\|\boldsymbol{x} - \boldsymbol{z}\|_2 \le 0.001\|\boldsymbol{x}\|_2$. Report the empirical success rates (averaged over 10 experiments) for each choice of k.