E9 203: Homework - 5

Assigned on: 27 Mar. 2015, due Apr. 10, 2015.

1 Topics

- Coherence, ℓ_1 coherence
- Restricted Isometry Property
- Null space property
- Recovery guarantees via coherence and restricted isometry property

Notation: "Triple-bar" norms $\|\!|\cdot|\!\|$ denote (vector norm) induced matrix norms while "double-bar" norms $\|\cdot\|$ denote vector norms (possibly, on matrices). For example, $\|\!|A|\!\|_2$ is the same as $\|A\|_{2\to 2}$ we used in class. Also, $\|A\|_2$ or $\|A\|_F$ is the matrix Frobenius norm.

2 Problems

1. Given a matrix $A \in \mathbb{R}^{m\times n}$, show that spark$(A) \leqslant 2k$ if and only if $\exists$ some $\mathbf{y} \in \mathbb{R}^m$ for which there is more than one k-sparse vector $\mathbf{x}$ such that $\mathbf{y} = A\mathbf{x}$.

2. (Spiked Identity Model): Let $\mathbf{X} \in \mathbb{R}^{m\times n}$ be a random design matrix with each row, $\underline{x}_i \in \mathbb{R}^n \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$, i.i.d. For $\alpha \in (0\ 1)$, consider the following covariance matrix,

$$\mathbf{\Sigma} = (1-\alpha)\mathbb{I}_{n\times n} + \alpha\mathbf{1}\mathbf{1}^T$$

where $\mathbf{1}$ is a n-dimensional vector of all ones.

(a) Verify that for any $i \neq j$:

$$\mathbb{P}\left[\frac{1}{m}\underline{x}_i^T\underline{x}_j \geqslant \alpha - \epsilon\right] \geqslant 1 - c_1\exp(-c_2 m\epsilon^2). \tag{1}$$

for some positive constants c_1 and c_2. (Hint: Find the moment generating function of inner product of two Gaussian vectors; Use Chernoff bound with a suitable value of s to arrive at above form for the bound.)

(b) Argue that above result is equivalent to possible violation of mutual incoherence property (for large values of k for example)

(c) (Optional) Confirm, empirically, that these matrices do lead to exact recoveries using min-ℓ_1 optimization programs.

3. (Rauhut and Foucart, Ex. 4.17) Stable and robust recovery via dual certificate
Let $A \in \mathbb{C}^{m\times N}$ be a matrix with ℓ_2 normalized columns. Let $\mathbf{x} \in \mathbb{C}^N$ and let $S \subset [N]$ be an index set of s largest absolute entries of $\mathbf{x}$. Assume that

$$\|A_S^H A_S - \mathbf{I}\|_2 \leqslant \alpha$$

for some $\alpha \in (0,1)$ and that there exists a dual certificate $\mathbf{u} = A^H \mathbf{h} \in \mathbb{C}^N$ with $\mathbf{h} \in \mathbb{C}^m$ such that

$$\mathbf{u}_S = \operatorname{sign}(\mathbf{x}_S),\ \|\mathbf{u}_{\bar{S}}\|_\infty \leqslant \beta,\ \|\mathbf{h}\|_2 \leqslant \gamma\sqrt{s}$$

for some constants $0 < \beta < 1$ and $\gamma > 0$. Suppose that we are given corrupted measurements $\mathbf{y} = A\mathbf{x}+\mathbf{e}$ with $\|\mathbf{e}\|_2 \leqslant \eta$. Show that a solution $\mathbf{x}^\# \in \mathbb{C}^N$ of the ℓ_1 minimization problem

$$\min_{\mathbf{z}\in\mathbb{C}^N} \|\mathbf{z}\|_1 \text{ subject to } \|A\mathbf{z} - \mathbf{y}\|_2 \leqslant \eta$$

satisfies

$$\|\mathbf{x} - \mathbf{x}^\#\|_2 \leqslant C\sigma_s(\mathbf{x})_1 + D\sqrt{s}\eta$$

for appropriate constants $C, D > 0$ depending only on α, β and γ.

4. (Welch Bound) Let $A \in \mathbb{C}^{m\times n}$ be with normalized columns. Prove that $\mu(A) \geqslant \sqrt{\frac{n-m}{m(n-1)}}$, where $\mu(A)$ is the pairwise incoherence parameter for matrix A.

5. Given $A \in \mathbb{C}^{m\times N}$, show that, for $1 \leqslant s, t \leqslant N-1$, the ℓ_1 coherence function $\mu_1(\cdot)$ satisfies

$$\max\{\mu_1(s), \mu_1(t)\} \leqslant \mu_1(s+t) \leqslant \mu_1(s) + \mu_1(t).$$

6. (Thanks to Abhay Sharma for this question.)
Recall the definition of the *Null space property (NSP):*

Definition 1 *A matrix* $\mathbf{A} \in \mathbb{R}^{m\times n}$ *satisfies null space property of order* k *if, for all subsets* $S \subset \{1, 2, \ldots, n\}$ *with* $|S| = k$*, it holds that:*

$$\mathcal{C}(S) \cap \{\mathbf{v} \in \mathbb{R}^n \mid A\mathbf{v} = 0\} = \{0\} \quad \textit{where} \quad \mathcal{C}(S) \triangleq \{\mathbf{v} \in \mathbb{R}^n \mid \|\mathbf{v}_{S^c}\|_1 \leqslant \|\mathbf{v}_S\|_1\} \tag{2}$$

or,

$$\|\mathbf{v}_S\|_1 < \|\mathbf{v}_{S^c}\|_1 \quad \textit{for all} \quad \mathbf{v} \in \mathcal{N}(A)\setminus\{0\} \tag{3}$$

We will conduct a series of experiments to obtain better insight into the above definition and its consequences. For these experiments, we will work with 1-sparse vectors, $m = 2$ and $N = 3$. The routines used in the starter Matlab routine ("NSP_1") have been downloaded from:
"http://www.nbb.cornell.edu/neurobio/land/PROJECTS/Hierarchy/".

(a) Run "NSP_1". With regards to the definition above, what does the Fig. 1 represent? (*Hint:* Note that you can rotate the figure. What in the construction above represents a cone?)

(b) Generate a random 2×3 matrix A. Using the definition above and using Fig. 1, how can you verify whether or not A satisfies NSP of order 1. (*Hint:* See (2) above.) Using multiple trials, generate (and save for later use) two instances of "good" and "bad" matrices, i.e., one satisfying NSP and the other not satisfying NSP, and provide the corresponding plots to illustrate your assertion.

(c) The "bad" matrix is not guaranteed to lead to a successful recovery, by ℓ_1 minimization, of all 1-sparse vectors, since it does not satisfy NSP of order 1. Note that Fig. 2 generated by running "NSP_1", is a ℓ_1 ball with radius 1. How can we use Fig. 2 to demonstrate the consequence of not satisfying the NSP? (*Hint:* Plot all the possible solutions of $A\mathbf{x} = \mathbf{b}$ for 1-sparse vectors of norm 1 and see how the solution space interacts with the ℓ_1 ball.) How does the same experiment turn out in the case of the "good" matrix ?

(d) Using the "bad" matrix generated above, setup the actual recovery experiment where we keep the matrix fixed but we randomly generate a 1-sparse vector and empirically compute the probability of recovery error. Repeat the same experiment with the "good" matrix also. Can you guess what would be the approximate probability of recovery error from the second exercise? Why? (*Hint:* Remember the comment about uniform/non-uniform recovery from the class.)

(e) By using the ℓ_2 ball instead of ℓ_1 ball in the second experiment above, what can we infer about sparse recovery using ℓ_2-minimization programs?

7. Show that:

   (a) If $A \in \mathbb{C}^{m\times N}$ satisfies the NSP of order s, then the null space $\mathcal{N}(A)$ does not contain any $2s$-sparse vector other than the zero vector.

   (b) If A satisfies the NSP of order s, then so does $\tilde{A} \triangleq \mathbf{G}A$, where $\mathbf{G} \in \mathbb{C}^{m\times m}$ is an invertible matrix.

   (c) If A satisfies the NSP of order s, then so does $\tilde{A} \triangleq \begin{bmatrix} A \\ B \end{bmatrix}$ where $B \in \mathbb{C}^{m'\times N}$ is an arbitrary matrix.

8. In defining the restricted isometry property in class, we used bounds that are symmetric about 1, e.g., bounds of the form $1 - \delta_k$ and $1 + \delta_k$. Show that, if the bounds were of the form

$$\alpha\|\mathbf{x}\|_2^2 \leqslant \|A\mathbf{x}\|_2^2 \leqslant \beta\|\mathbf{x}\|_2^2,$$

   one can always scale A to get $\hat{A}$ satisfying

$$(1-\delta)\|\mathbf{x}\|_2^2 \leqslant \|\hat{A}\mathbf{x}\|_2^2 \leqslant (1+\delta)\|\mathbf{x}\|_2^2.$$

   Express δ in terms of α and β.

9. (Relationship between RIC and incoherence) Let δ_k be the restricted isometry constants for a given matrix A. Let μ be the pairwise incoherence parameter for A. Further, let $\mu_1(k)$ be the 1-coherence parameter (or Babel function) for the matrix A, i.e.,

$$\mu_1(k) = \max_{S\subset[n],|S|=k} \|\!|A_{S^c}^T A_S|\!\|_\infty \leqslant k\mu$$

   Verify:

   (a) $\delta_2 = \mu$.

   (b) $\delta_k \leqslant \mu_1(k-1) \leqslant (k-1)\mu$.

10. Show that $\delta_{2k} < 1$ is a sufficient condition for uniqueness of any k-sparse solution to $\mathbf{y} = A\mathbf{x}$, where $\mathbf{y} \in \mathbb{R}^m, \mathbf{x} \in \mathbb{R}^n, A \in \mathbb{R}^{m\times n}$ and $m \ll n$.

11. Prove that $\delta_3 \leqslant 3\delta_2$. (Hint: Use Gershgorin's Theorem and the RIP constants to bound the required quantities.) (Note: In fact a more general inequality is true: For c, k positive integers, $\delta_{ck} \leqslant c\delta_{2k}$. To prove this we have to use the block version of the Gershgorin's theorem. The key message here is that sometimes controlling δ_{2k} provides some control over higher RIP constants.)