EXERCISE SHEET: (MODEL-BASED) COMPRESSED SENSING

JEAN-LUC BOUCHOT

1. Day 1: Introduction, hands-on

1.1. Mathematical topics.

Exercise 1 (Regularization). Let $A \in \mathbb{R}^{m\times N}$ and $\lambda > 0$. What is the solution to

$$\operatorname{argmin}_{\mathbf{z}\in\mathbb{R}^N} \|A\mathbf{z}-\mathbf{y}\|_2^2 + \lambda\|\mathbf{z}\|_2^2.$$

Does your result hold in complex spaces?

Exercise 2. Prove the following equivalences from class.

Let A be an $m \times N$ sensing matrix in $\mathbb{R}^N$. The following statements are equivalent

- Every s-sparse $\mathbf{x}$ is the unique s-sparse solution of $A\mathbf{z} = A\mathbf{x}$.
- $\ker(A) \cap \{\mathbf{x} \in \mathbb{R}^N : \|\mathbf{x}\|_0 \leq 2s\} = \{0\}$.
- Let $S \subset \{1, \cdots, N\}$ be any set with $|S| \leq 2s$. The submatrix A_S is injective.
- Every set of $2s$ columns of A is linearly independent.

Exercise 3. We want to understand the importance of rotations of the space on the null space property. Throughout this exercise, let $A \in \mathbb{R}^{m\times N}$ be a matrix satisfying the null space property of order s and let D_1 and D_2 be two non-singular matrices respectively in $\mathbb{R}^{m\times m}$ and $\mathbb{R}^{N\times N}$.

(1) Show that the matrix D_1A satisfies the null space property of order s.
(2) Find a counter example such that AD_2 does not satisfy the null space property. ($m = 2$, $N = 3$, $s = 1$)

Exercise 4. Let $A \in \mathbb{R}^{m\times N}$ satisfy the null-space property of order s. Prove directly without using any other theorems from the course, that $\mathbf{x} \in \ker(A)$ and $\|\mathbf{x}\|_0 \leq 2s \Rightarrow \mathbf{x} = 0$.

Exercise 5. Prove Stechkin's (tail) estimate: For any $q > p$ and any $\mathbf{x} \in \mathbb{R}^N$,

$$\sigma_s(\mathbf{x})_q \leq s^{1/q-1/p}\|\mathbf{x}\|_p.$$

1.2. Programming (Python) experiments.

Exercise 6. Program (and test!) the following functions: (adapt the inputs/ outputs to your needs...)

(1) `l2normalize`: a function that takes a matrix as an input, and returns the same matrix with its columns having ℓ_2 norm 1.
(2) `coherence`: which takes a matrix as an input and returns its coherence
(3) `random_matrix`: which takes at least three inputs: `nb_rows`, the number of rows, `nb_cols`, the number of columns, and `rdm_type`, the type of random generation used. Check the documentation for `random.random()` from `NumPy`. This function returns an `nb_rows x nb_cols` matrix whose entries are generated at random according to the distribution `rdm_type`. You may pass an extra parameter to normalize the columns of your matrix or not. `rdm_type` should at least contain *Gaussian*, *Rademacher*, and *Exponential*, but you may (should?) include more options.
(4) `subsample_DFT`: which takes two inputs, `N` the size of the signals, and `Omega` the subset of rows selected from the DFT matrix. It returns a $|\text{Omega}| \times N$ matrix constructed by extracting the rows of an $N \times N$ DFT matrix which are supported on the index set `Omega`.

Exercise 7. (1) For various random matrices, generate graphs of the their coherences with respect to m, the number of rows (fixing the number of columns N). To this end, generate 100 random matrices and compute the average, min, and max of the coherence. (*Hint:* Remember to normalize the columns of the matrix!)

Date: August 8, 2019.

(2) Using the basis pursuit procedure provided, test various settings (in terms of s, m, and N, as well as various random matrices). For 100 s-sparse random vectors, compute the average recovery rate. Let s vary and plot the percentage of recovery with respect to the sparsity s.
(3) Repeat the procedures above with random subsampled Fourier matrices, in which the support `Omega` is generated at random with a given size.

2. Day 2: More on coherence; RIP

2.1. Mathematical topics.

Exercise 8. Let $A \in \mathbb{R}^{m\times N}$ be a matrix whose columns are ℓ_2 normalized, and let $U \in \mathbb{R}^{m\times m}$ be a unitary matrix. Prove the following statements:

(1) $\mu(A) \leq 1$,
(2) $\mu(UA) = \mu(A)$.

Exercise 9 (Difficult). Show that $\|A^\dagger A\|_{1\to 1} < 1$ implies the null space property. Here $A^\dagger$ denotes the Moore-Penrose pseudo-inverse.

Exercise 10. Let $A \in \mathbb{R}^{m\times N}$ be a full rank matrix and $\mathbf{y} \in \mathbb{R}^m$ be a given vector. Let $S \subset \{1, \cdots, N\}$ be *small* ($|S| \leq m$). Compute

$$\mathbf{x}^\# := \text{argmin}_{\mathbf{z}\in\mathbb{R}^N}\{\|\mathbf{y} - A\mathbf{z}\|_2, \quad \text{subject to } \operatorname{supp}(\mathbf{z}) \subset S\}.$$

Exercise 11. Prove the Exact Recovery Condition for OMP:
Let $A \in \mathbb{R}^{m\times N}$ and $S \subset \{1, \cdots, N\}$, $|S| = s$. Every vector $\mathbf{x} \neq 0$ supported on S is recovered after at most s iterations of OMP provided

(1) A_S is injective and
(2) the (ERC) holds.

Exercise 12. Prove the following:
Given an index set $S \subset \{1, \cdots, N, \}$, if

$$\mathbf{v} := \text{argmin}_{\mathbf{z}\in\mathbb{R}^N}\{\|\mathbf{y} - A\mathbf{z}\|, \operatorname{supp}(\mathbf{z}) \subset S\}$$

then

$$\langle \mathbf{y} - A\mathbf{v}, a_i\rangle = 0,$$

for all $i \in S$.

Exercise 13. Prove the following
Let $A \in \mathbb{C}^{m\times N}$ be a matrix with ℓ_2-normalized columns. Given $S \subset \{1, \cdots, N\}$ and $\mathbf{v}$ supported on S, and $1 \leq j \leq N$, if

$$\mathbf{w} := \text{argmin}_{\mathbf{z}\in\mathbb{C}^N}\{\|A\mathbf{z} - \mathbf{y}\|_2, \operatorname{supp}(\mathbf{z}) \subset S \cup \{j\}\}$$

then

$$\|\mathbf{y} - A\mathbf{w}\|_2^2 \leq \|\mathbf{y} - A\mathbf{v}\|_2^2 - |(A^*(\mathbf{y} - A\mathbf{v}))_j|^2.$$

Exercise 14. Prove the alternate form for the restricted isometry constant:

$$\delta_s := \max_{|S|\leq s} \|A_S^* A_S - I\|_{2\to 2}.$$

2.2. Programming.

Exercise 15. Consider two (or more) natural images (say the traditional `pepper` and `barbara` images). Compute their wavelet coefficients. Compute the best s-term approximation for varying value of s a plot everything on a log-log plot. Let the value p of the norm consider vary too and plot on a single graph the results for various values of p.

Exercise 16. (1) Implement the Orthogonal Matching Pursuit algorithm.
(2) Set the ambient dimension $N = 300$ and the number of observations to be $m = 120$. For 100 experiments per level, let the sparsity evolve from 1 to 60 and plot the average number of correct recovery. For this example, consider a Gaussian sensing matrix (appropriately normalized).

Exercise 17. (1) Verify numerically that s iterations are sufficient for the recovery of s-sparse vectors via OMP
(2) What happens to the number of iterations as you add a little bit of noise to the measurements?

3. Day 3: more on RIP, iterative methods

3.1. Mathematical topics.

Exercise 18. Let $A \in \mathbb{R}^{m\times N}$ be a matrix satisfying the $RIP(s, \delta_s)$ for parameters $s > 0$ and $\delta_s > 0$. Prove that

$$\|A\mathbf{x}\|_2 \leq \sqrt{1+\delta_2}\left(\|\mathbf{x}\|_2 + \frac{\|\mathbf{x}\|_1}{\sqrt{s}}\right).$$

Exercise 19. Establish the following stability result of BPDN using the RIP.

Let $A \in \mathbb{R}^{m\times N}$ be a matrix satisfying $RIP(2s, \delta_{2s})$ with $\delta_{2s} < 1/3$. Let $\mathbf{x} \in \mathbb{R}^{m\times N}$ and $\mathbf{y} = A\mathbf{x} + \mathbf{e}$ with $\|\mathbf{e}\|_2 \leq \varepsilon$. The solution $\mathbf{x}^\#$ of the quadratically constrained ℓ_1 minimization (i.e. BPDN) approximates $\mathbf{x}$ in the following sense:

$$\|\mathbf{x}^\# - \mathbf{x}\|_2 \leq \frac{C}{\sqrt{s}}\sigma_s(\mathbf{x})_1 + D\varepsilon,$$

where C is a universal constant.

Exercise 20. Let $A \in \mathbb{R}^{m\times N}$ satisfy the $RIP(2s, \delta_{2s})$ and let $\mathbf{v} \in \mathrm{Ker}(A)$. Let S_0, S_1, $\cdots$ be index sets of size s containing the indices of largest magnitudes of $\mathbf{v}$ in decreasing order. Show that

$$\|\mathbf{v}_{S_0}\|_2^2 + \|\mathbf{v}_{S_1}\|_2^2 \leq \frac{\delta_{2s}}{1-\delta_{2s}}\sum_{k\geq 2}\|\mathbf{v}_{S_k}\|_2\left(\|\mathbf{v}_{S_0}\|_2 + \|\mathbf{v}_{S_1}\|_2\right).$$

Furthermore, by completing the squares, prove that the matrix A satisfies the stable null space property of order S and constant $\rho = \frac{1+\sqrt{2}}{2}\frac{\delta_{2s}}{1-\delta_{2s}}$.

Exercise 21. Consider the following variant of iterative hard thresholding in which the iterates are as follows:

$$\mathbf{x}^{n+1} = H_s\left(\mathbf{x}^n + \mu A^T(\mathbf{y} - A\mathbf{x}^n)\right)$$

defined for a constant (the so-called step size) $\mu > 0$. Prove the following inequalities

$$\|A(\mathbf{x}^{n+1} - \mathbf{x})\|_2^2 - \|A(\mathbf{x}^n - \mathbf{x})\|_2^2 = \|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2 + 2\langle \mathbf{x}^n - \mathbf{x}^{n+1}, A^TA(\mathbf{x} - \mathbf{x}^{n+1})\rangle$$

$$2\mu\langle \mathbf{x}^n - \mathbf{x}^{n+1}, A^TA(\mathbf{x} - \mathbf{x}^n)\rangle \leq \|\mathbf{x}^n - \mathbf{x}\|_2^2 - 2\mu\|A(\mathbf{x}^n - \mathbf{x})\|_2^2 - \|\mathbf{x}^{n+1} - \mathbf{x}^n\|_2^2.$$

Denoting by δ_{2s} the $2s$ isometry constant of A, prove the following inequality

$$\|A(\mathbf{x}^{n+1} - \mathbf{x})\|_2^2 \leq \left(1 - \frac{1}{\mu(1+\delta_{2s})}\right)\|A(\mathbf{x}^{n+1} - \mathbf{x}^n)\|_2^2 + \left(\frac{1}{\mu(1-\delta_{2s})} - 1\right)\|A(\mathbf{x} - \mathbf{x}^n)\|_2^2.$$

Conclude that the sequence of iterates $(\mathbf{x}^n)_n$ converges to $\mathbf{x}$ when $1 + \delta_{2s} < 1/\mu < 2(1 - \delta_{2s})$.

3.2. Programming exercises.

Exercise 22. Implement both iterative hard thresholding and hard thresholding pursuit and compare their numerical behaviours to that of OMP (see Day 2).

School of Mathematics and Statistics, Beijing Institute of Technology
Email address: jlbouchot@bit.edu.cn