Problem set 4

Veniamin Morgenshtern, Helmut Bölcskei, Erwin Riegler

Due date: 4. July 2019

Problem 1: Orthogonal matching pursuit

This is a computer exercise.

Let $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in \mathbb{R}^{M,N}$ be a measurement matrix, with columns normalized as $\|\mathbf{a}_i\| = 1$, and $\mathbf{x} \in \mathbb{R}^N$ an s-sparse vector, i.e., $\mathbf{x}$ has at most s nonzero entries. Define $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{R}^M$.

Implement, preferably in Python, the orthogonal matching pursuit (OMP) algorithm, which allows to recover the vector $\mathbf{x}$ from the measurement vector $\mathbf{y}$, knowing the measurement matrix $\mathbf{A}$ and the sparsity level s only.

For concreteness, take $N = 512$ and $M = 128$. Generate a random sparse vector $\mathbf{x} \in \mathbb{R}^N$ with a sparsity level $s = 20$. Generate a random measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ with entries identically and independently distributed according to a Gaussian distribution of zero mean and variance $1/M$. *Renormalize the columns of* $\mathbf{A}$ *to have l2-norm equal to one.* Compute $\mathbf{y} = \mathbf{A}\mathbf{x}$ and reconstruct $\mathbf{x}$ from $\mathbf{A}$ and $\mathbf{y}$ using OMP algorithm. Repeat the procedure with other values of s.

Recall: The key to OMP is to determine which columns of $\mathbf{A}$ participate in the measurement vector $\mathbf{y}$. The idea behind this is to pick columns of $\mathbf{A}$ in a greedy fashion. At each iteration, we choose the column of $\mathbf{A}$ which is most correlated with the residual, i.e., the remaining part of $\mathbf{y}$ which has not yet been approximated. This contribution is then subtracted from $\mathbf{y}$ and the algorithm iterates on the residual.

Problem 2: Recovery of approximately sparse signal using l_1-minimization

Let $\mathbf{D} \in \mathbb{C}^{K \times N}$ be a matrix ($K < N$). For an index set $\Lambda \subset [1, N]$, define the quantity

$$C(\mathbf{D}, \Lambda) = \max_{\mathbf{h} \in \mathcal{N}(\mathbf{D}),\ \mathbf{h} \neq \mathbf{0}} \frac{\|\mathbf{h}_\Lambda\|_1}{\|\mathbf{h}\|_1} = \max_{\mathbf{h} \in \mathcal{N}(\mathbf{D}),\ \mathbf{h} \neq \mathbf{0}} \frac{\sum_{k \in \Lambda} |h_k|}{\sum_{k=1}^{N} |h_k|},$$

where $\mathbf{h}_\Lambda$ is the vector constructed from $\mathbf{h}$ by setting to zero all but the entries indexed by Λ. Let $\mathbf{x} \in \mathbb{C}^N$ and consider the l_1-minimization problem (P1):

$$\text{minimize}_{\hat{\mathbf{x}} \in \mathbb{C}^N} \text{ subject to } \mathbf{D}\hat{\mathbf{x}} = \mathbf{D}\mathbf{x}.$$

In the lecture, we have seen that if $\mathbf{x}$ is s-sparse with support $\mathcal{S} = \operatorname{supp}(\mathbf{x}) = \{k \in [1, N] : x_k \neq 0\}$, $|\mathcal{S}| \leq s$, and if $C(\mathbf{D}, \mathcal{S}) < 1/2$, then $\mathbf{x}$ is the unique solution to (P1).

Now, assume that $\mathbf{x}$ is *approximately* sparse, meaning that many of the entries of $\mathbf{x}$ are close to zero. Show that if $C(\mathbf{D}, \mathcal{S}) < 1/2$, the solution $\mathbf{x}^*$ to (P1) satisfies

$$\|\mathbf{x}^* - \mathbf{x}\|_1 \leq \frac{2\|\mathbf{x} - \mathbf{x}_\mathcal{S}\|}{1 - 2C(\mathbf{D}, \mathcal{S})},$$

where $\mathcal{S}$ denotes the set containing the indices of the s largest (in magnitude) components of $\mathbf{x}$, the vector $\mathbf{x}_{\mathcal{S}}$ thus representing the best s-sparse approximation to $\mathbf{x}$. Do you recover the condition derived in the lecture for the case of exactly sparse signals?

Problem 3: Restricted isometry property and coherence

Let $\mathbf{A} \in \mathbb{C}^{M\times N}$ be a matrix having normalized columns, i.e., each column $\mathbf{a}_l$, $l \in [1, N]$, satisfies $\|a_l\|_2 = 1$. For $s \in [1, N]$, we say that $\mathbf{A}$ satisfies the restricted isometry property (RIP) of order s if there exists $\delta \in (0, 1)$ such that

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

for all $\mathbf{x} \in \mathbb{C}^N$ such that $\left|\operatorname{supp}(\mathbf{x})\right| \leq s$ and we define $\delta_s(\mathbf{A})$ to be the smallest such δ.

1. Show that if $\mathbf{A}$ satisfies the RIP of order s, then it also satisfies the RIP or order $s' \leq s$.
2. Show that
$$\delta_s(\mathbf{A}) = \max_{\mathcal{S}\subset[1,N],|\mathcal{S}|\leq s} \|\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}} - \mathbf{I}_s\|_{2\to 2},$$
where $\mathbf{A}_{\mathcal{S}}$ denotes the matrix consisting of the columns of $\mathbf{A}$ indexed by the set $\mathcal{S} \subset [1, N]$ and $\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}$ denotes the conjugate transpose of $\mathbf{A}_{\mathcal{S}}$. *Recall:* The matrix norm $\|\cdot\|_{2\to 2}$ is defined as
$$\|\mathbf{A}\|_{2\to 2} = \sup_{\mathbf{x}\in\mathbb{C}^N,\mathbf{x}\neq\mathbf{0}} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$$
for $\mathbf{A} \in \mathbb{C}^{N\times N}$.
3. What does $\delta_s(\mathbf{A})$ say about the eigenvalues of the matrix $\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}}$, where $\mathcal{S}$ is a subset of $[1, N]$ of size s? Find an upper bound for the condition number of $\mathbf{A}_{\mathcal{S}}^{\mathsf{H}}\mathbf{A}_{\mathcal{S}}$ expressed in terms of $\delta_s(\mathbf{A})$.
4. Prove that
$$\delta_s(\mathbf{A}) \leq \mu(\mathbf{A})(s-1)$$
where
$$\mu(\mathbf{A}) = \max_{k,l\in[1,N],k\neq l} |\langle \mathbf{a}_k, \mathbf{a}_l\rangle|$$
is the coherence of the matrix $\mathbf{A}$.

Problem 4: Restricted isometry property: a counterexample

Let $\mathbf{A} \in \mathbb{C}^{M\times N}$ be the matrix with constant elements

$$a_{k,l} = \frac{1}{\sqrt{M}}, k \in [1, M],\ \ l \in [1, N].$$

Show that $\mathbf{A}$ does not satisfy the RIP of order $s \geq 2$.

Problem 5: Normal vector via PCA

This is a computer exercise. The goal of the exercise is to use PCA to estimate the normal vector to a 2D curve and a 3D surface.

1. In the file `circle.txt` your are given a list of 2D points that are sampled from a 1D curve: an arc of a circle.

   - Read the data from `circle.txt` and visualize the points using the `normal_estimation_pca.ipynb` notebook.
   - Explain how you can use PCA to estimate the tangent line to the curve and the normal vector to the curve. *Hint:* The tangent direction is the direction in which the variance of the data is the largest; the normal direction is the direction in which the variance of the data is the smallest.
   - Implement your idea, visualize the normal vector and the tangent line.

2. In the file `sphere.txt` your are given a list of 3D points that are sampled from a 2D manifold: a sector of a sphere.

   - Read the data from `sphere.txt` and visualize the points using the `normal_estimation_pca.ipynb` notebook.
   - Explain how you can use PCA to estimate the tangent plane to the manifold and the normal vector to the manifold. *Hint:* The normal direction is the direction in which the variance of the data is the smallest.
   - Implement your idea, visualize the normal vector.

Problem 6: Dual norm of the spectral norm is the nuclear norm

Consider $\mathbf{Z} \in \mathbb{R}^{n\times n}$. The spectral norm is defined

$$\|\mathbf{Z}\| = \max_i \sigma_i(\mathbf{X}).$$

The nuclear norm is defined as

$$\|\mathbf{Z}\|_* = \sum_i \sigma_i(\mathbf{X}).$$

Let $\|\cdot\|$ be a norm. The associated dual norm, is defined as $\|\mathbf{Z}\|_{dual} = \sup\{\langle \mathbf{Z}, \mathbf{X}\rangle \ : \ \|\mathbf{X}\| \leq 1\}$.

The inner product between two matrices is given by $\langle \mathbf{X}, \mathbf{Z}\rangle = \mathrm{trace}(\mathbf{X}^\mathsf{T}\mathbf{Z})$.

Show that the nuclear norm is the dual norm of the spectral norm with respect to this inner product.

Problem 7: Subgradient of a norm

A vector $\mathbf{g} \in \mathbb{R}^n$ is a subgradient of $f : \mathbb{R}^n \to \mathbb{R}$ at $\mathbf{x}$ if for all $\mathbf{z}$, $f(\mathbf{z}) \geq f(\mathbf{x}) + \langle \mathbf{g}, \mathbf{z} - \mathbf{x}\rangle$. The set of all subgradients is called the subdifferential, denoted $\partial f(x)$.

The subdifferential of a norm $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x}\rangle}$ in a inner product space can be characterized as follows:

$$\partial\|\mathbf{x}\| = \underbrace{\{\mathbf{g} : \langle \mathbf{g}, \mathbf{x}\rangle = \|\mathbf{x}\|, \|\mathbf{g}\|_* \leq 1\}}_{\mathcal{G}}$$

where $\|\cdot\|_*$ is the dual norm of $\|\cdot\|$.

Show one inclusion in this statement: if $\mathbf{g} \in \mathcal{G}$, then $\mathbf{g} \in \partial\|\mathbf{x}\|$.

Problem 8: Subgradient of a nuclear norm

Let $\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^\mathsf{T}$ denote the SVD of $\mathbf{M}$. Then, the subdifferential of $\mathbf{M}$ is given by

$$\partial\|\mathbf{M}\|_* = \underbrace{\{\mathbf{U}\mathbf{V}^\mathsf{T} + \mathbf{W} : \mathcal{P}_\mathbf{U}\mathbf{W} = 0, \mathbf{W}\mathcal{P}_\mathbf{V} = 0, \|\mathbf{W}\| \le 1\}}_{\mathcal{G}}.$$

Here $\mathcal{P}_\mathbf{U} = \sum_i \mathbf{u}_i\mathbf{u}_i^\mathsf{T}$ is the orthogonal projector onto the columns of $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ and similarly $\mathcal{P}_\mathbf{V} = \sum_i \mathbf{v}_i\mathbf{v}_i^\mathsf{T}$ is the orthogonal projector onto the columns of $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$.

Prove one inclusion of this statement: if $\mathbf{G} \in \mathcal{G}$, then $\mathbf{G} \in \partial\|\mathbf{M}\|_*$.