Mathematical methods for machine learning and signal processing SS 19

Solutions to problem set 3

Prof. Veniamin Morgenshtern

Solver: Erwin Riegler, Veniamin Morgenshtern

Problem 1: Gram matrix

Let $\mathbf{A} \in \mathbb{C}^{M \times N}$ be the matrix whose kth column is $\mathbf{a}_k$. Since the entry in the kth row and in the lth column is $\langle \mathbf{a}_k, \mathbf{a}_l \rangle = \mathbf{a}_l^{\mathsf{H}} \mathbf{a}_k$, the Gram matrix can be written as $\mathbf{G} = \mathbf{A}^{\mathsf{H}} \mathbf{A}$. Therefore, it holds that

$$\mathbf{G}^{\mathsf{H}} = (\mathbf{A}^{\mathsf{H}} \mathbf{A})^{\mathsf{H}} = \mathbf{A}^{\mathsf{H}} (\mathbf{A}^{\mathsf{H}})^{\mathsf{H}} = \mathbf{A}^{\mathsf{H}} \mathbf{A} = \mathbf{G},$$

which shows that $\mathbf{G}$ is Hermitian. Moreover, for all $\mathbf{x} \in \mathbb{C}^M$, we have

$$\mathbf{x}^{\mathsf{H}} \mathbf{G} \mathbf{x} = \mathbf{x}^{\mathsf{H}} \mathbf{A}^{\mathsf{H}} \mathbf{A} \mathbf{x} = (\mathbf{A}\mathbf{x})^{\mathsf{H}} \mathbf{A} \mathbf{x} = \|\mathbf{A}\mathbf{x}\|^2 \geq 0.$$

Therefore, $\mathbf{G}$ is positive-semidefinite. Since $\|\cdot\|$ is a norm, $\mathbf{x}^{\mathsf{H}} \mathbf{G} \mathbf{x} = 0$ implies that $\mathbf{A}\mathbf{x} = 0$. So, if $\mathbf{A}$ has linearly independent columns, $\mathbf{A}\mathbf{x} = 0$ implies that $\mathbf{x} = 0$, and $\mathbf{G}$ is positive-definite.

Problem 2: "l_0-norm"

For a vector $\mathbf{x} \in \mathbb{C}^N$, we have that

$$\|2\mathbf{x}\|_0 = \|\mathbf{x}\|_0,$$

which shows that the homogeneity property of a norm is not satisfied. Therefore, $\|\cdot\|_0$ is not a norm for $\mathbb{C}^N$. Its name and notation come from an abuse of terminology, taking $p = 0$ in the definition of the l_p-norm $\|\cdot\|_p$ which is commonly defined for $p \in [1, \infty]$ as

$$\|\mathbf{x}\|_p^p = \sum_{k=1}^{N} |x_k|^p.$$

Problem 3: Fat matrix inversion

1. The equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ has infinitely many solutions if $\mathbf{y}$ is in the column range space of $\mathbf{A}$. This is guaranteed when rank$\mathbf{A} = S$, i.e., when all rows of $\mathbf{A}$ are linearly independent (or equivalently, S columns of $\mathbf{A}$ are linearly independent). The equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ has no solution if $\mathbf{y}$ is not in the column range space of $\mathbf{A}$. This can only happen when rank $\mathbf{A} < S$, i.e., when $\mathbf{A}$ has linearly dependent rows (or equivalently, less than S linearly independent columns).

2. Solving the equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ under the constraint that $x_j = 0$ for $j \notin \mathcal{S}$ amounts to solving the equation $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$, where $\tilde{\mathbf{A}}$ is the $S \times S$ matrix obtained by removing the $N - S$ columns of $\mathbf{A}$ that are indexed by S^c and where the unknown $\tilde{\mathbf{x}}$ is an S-dimensional vector. The equation $\tilde{\mathbf{A}}\tilde{\mathbf{x}} = \mathbf{y}$ has exactly one solution if det $\tilde{\mathbf{A}} \neq 0$, i.e., if $\tilde{\mathbf{A}}$ has full rank. Therefore, the equation $\mathbf{A}\mathbf{x} = \mathbf{y}$ has exactly one solution if the columns $\{\mathbf{a}_j\}_{j \in \mathcal{S}}$ indexed by $\mathcal{S}$ are linearly independent.

Problem 4: Compressed sensing

Intuitively, we can write $\mathbf{x}$ in terms of its discrete derivative. We can write $\mathbf{x} = \mathbf{U}\mathbf{x}'$, where

$$\mathbf{U} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 1 & \dots & 0 \\ & & & \ddots & \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \in \mathbb{R}^{N \times N}$$

is the discrete integration matrix and $\mathbf{x}' = [\underbrace{\alpha_1 0 \dots 0}_{\text{block1}} \underbrace{\alpha_2 - \alpha_1 0 \dots 0}_{\text{block2}} \dots \underbrace{\alpha_s - \alpha_{s-1} 0 \dots 0}_{\text{block}s}]^\mathsf{T} \in \mathbb{R}^N$ is an s-sparse vector. We can then recover $\mathbf{x}'$ from $\mathbf{y} = \mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{U}\mathbf{x}'$ by solving

$$\text{minimize}_{\hat{\mathbf{x}}} \ \|\hat{\mathbf{x}}\|_1 \text{ subject to } \mathbf{y} = \mathbf{A}\mathbf{U}\hat{\mathbf{x}}.$$

Problem 5: P0 recovery algorithm

See `pzero.ipynb`

Problem 6: Coherence in sines and spikes

By definition:

$$\mu(\mathbf{D}) = \max_{k,l} |\langle \mathbf{d}_k, \mathbf{d}_l \rangle|.$$

Observe that

$$\langle \mathbf{d}_k, \mathbf{d}_l \rangle = \begin{cases} 0, & k, l \le M \text{ or } k, l > M \\ \frac{1}{\sqrt{M}} \left| e^{2\pi i k(l-M)/M} \right| = \frac{1}{\sqrt{M}}, & k \le M \text{ and } l > M. \end{cases}$$

Therefore, $\mu(\mathbf{D}) = \frac{1}{\sqrt{M}}$.

From the lecture we know that if

$$s < \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{D})}\right) = \frac{1}{2}(1 + \sqrt{M})$$

successful recovery via P0 or basis pursuit is guaranteed for every s-sparse signal.

Problem 7: Coherence in super-resolution

Let's compute the absolute value of the inner product between the columns l_1 and l_2 of $\mathbf{F}_{\text{lo}}$:

$$|\langle \mathbf{f}_{l_1}, \mathbf{f}_{l_2} \rangle| = \frac{1}{N} \left| \sum_{k=-f_c}^{f_c} e^{i2\pi k(l_1 - l_2)/N} \right| = \frac{1}{N} \left| \sum_{k=0}^{M-1} e^{i2\pi k l/N} \right|$$

where $l = l_1 - l_2$ and $M = 2f_c + 1$.

Using the formula for the sum of geometric series:

$$\sum_{k=0}^{M-1} e^{ikx} = \frac{1 - e^{iMx}}{1 - e^{ix}} = \frac{\sin(Mx/2)}{\sin(x/2)} e^{ix(M-1)/2}$$

we find:

$$|\langle \mathbf{f}_{l_1}, \mathbf{f}_{l_2} \rangle| = \frac{1}{N} \left| \frac{\sin(\pi l M/N)}{\sin(\pi l/N)} \right|.$$

Therefore,

$$\mu(\mathbf{F}_{\text{lo}}) = \max_{l_1, l_2} |\langle \mathbf{f}_{l_1}, \mathbf{f}_{l_2} \rangle| = \max_{l} \frac{1}{N} \left| \frac{\sin(\pi l M/N)}{\sin(\pi l/N)} \right| \geq \frac{1}{N} \left| \frac{\sin(\pi M/N)}{\sin(\pi/N)} \right|.$$

For $M = N/2$ and large N we can approximate the right hand side of the equation above as

$$\frac{1}{N} \left| \frac{\sin(\pi M/N)}{\sin(\pi/N)} \right| \approx \frac{1}{N} \frac{\sin(\pi/2)}{\pi/N} = \frac{1}{\pi}.$$

From the lecture we know that if

$$s < \frac{1}{2} \left(1 + \frac{1}{\mu(\mathbf{F}_{\text{lo}})} \right)$$

successful recovery via P0 or basis pursuit is guaranteed for every s-sparse signal. Since $\mu(\mathbf{F}_{\text{lo}}) \geq \frac{1}{\pi}$ the bound we get for s is no better than

$$s < \frac{1}{2}(1 + \pi) \approx 2.$$

This is very pessimistic: for large N and $M = N/2$ the coherence-based bound only guarantees the successful recovery of only 2 spikes. It turns out, as we will see in the lecture, a much more optimistic bound may be obtained for low-frequency Fourier measurements when the signal is nonnegative.

Problem 8: Super-resolution experiment

See `superres.ipynb`