Final Exam

Inverse problems and sparsity

Parcimonie et grande dimension

Monday 24th of March, 2025

- This subject has three exercises, the exercises are independent and can be treated separately.
- **Non-master students** (option MIR students non-affiliated to master MeA or GRAF) will be evaluated on **all the exercise 1 and 2**.
- **Master students** will be evaluated on all the subject. They are asked to treat exercise 1 and 2 **before** exercise 3. The subject is not meant to be finished in $2h$ and the last exercise concerns only Master students which have treated, at least partially, the first two exercises and have enough time. Based on our experience of the previous years, it is likely that the third exercise will be seen as a bonus exercise to assess top grades on this exam.
- Questions with a coffee cup ☕ are **bonus questions, they are optional, their results can be admitted**, and these questions would be granted extra-points (the more coffee, the higher) in case of good answers.
- This subject has **3 pages and you have two hours**.

⚠ You are kindly asked to write your answers on **two separate sheets**: one for exercise 1 and a new one for exercises 2 and 3 in French and/or in English. For each group of sheets, your are asked to write your surname and first name on the first page.

⚠ On vous demande d'écrire vos réponses sur **deux jeux de copies séparés**: un pour l'exercice 1 et un autre pour les exercices 2 et 3. Pour chaque jeu de copies vous dervrez écire votre nom et votre prénom sur la première page.

Exercise 1: Concatenation of dictionaries

Consider the matrix $\mathbf{A} = [\mathbf{I}, \mathbf{F}] = [\mathbf{a}_1, \ldots, \mathbf{a}_{2m}]$ resulting from the concatenation of the following two matrices: the identity matrix of size m, $\mathbf{I}$, and the matrix $\mathbf{F}$ associated to the discrete Fourier transform in dimension m. The columns $\mathbf{e}_k$ of $\mathbf{F}$ are given by

$$\mathbf{e}_k(t) = \frac{1}{\sqrt{m}} e^{2\jmath\pi kt/m}, \quad 0 \leq k, t < m.$$

1. Give a detailed pseudo-code (Matlab or python-like) of Orthogonal Matching Pursuit with the following syntax: $(\texttt{x}, \texttt{res}) = \texttt{omp}(\texttt{b}, \texttt{A}, \texttt{k})$ where the input consists of: `b`, an m-dimensional vector, `A`, an $m \times N$ matrix, and `k`, the number of iterations; and the output consists of: `x`, the estimated N-dimensional coefficient vector, and `res`, the residual. Comment your code.
2. Consider $m = 1024$, and the m-dimensional vector $\mathbf{b} = 2\mathbf{a}_1 + 12\mathbf{a}_{497} - 18\mathbf{a}_{570} + 97\mathbf{a}_{1002}$.
   (a) Give the numerical value of the coherence of $\mathbf{A}$. Explain.
   (b) We apply 4 iterations of Orthonormal Matching Pursuit $((\texttt{x}, \texttt{res}) = \texttt{omp}(\texttt{b}, \texttt{A}, 2))$. Describe the resulting vectors `x` and `res`. Explain.
   (c) Assume we compute the vector x with smallest ℓ^1 norm such that $\mathbf{b} = \mathbf{A}x$. What can we say about the resulting vector ? Justify.
3. Consider $m = 2$.
   (a) Using the coherence of $\mathbf{A}$, what is the numerical value k of the sparsity level for which we have recovery guarantees for the main algorithms studied during the course ?
   (b) Find an example of two distinct $(k+1)$-sparse vectors x_1, x_2 such that $\mathbf{A}x_1 = \mathbf{A}x_2$.
   (c) What do you conclude ?
4. Consider $m = 4$.
   (a) Same questions as for $m = 2$.
   (b) What can you say about the RIP constant $\delta_{2s}(\mathbf{A})$ for $s = 1$? For $s = 2$? For larger s?

Exercise 2: Σ-OMP

OMP is defined for dictionaries with columns normalized in the usual Euclidean metric associated to the usual inner product $\langle x, y \rangle = \sum_i x_i y_i = x^\top y$ in $\mathbb{R}^m$ via the equality $\|x\|_2 := \sqrt{\langle x, x \rangle}$. In certain settings such as image processing, it is useful to consider weighted inner products $\langle x, y \rangle_\Sigma := x^\top \Sigma y$ where $\Sigma \in \mathbb{R}^{m \times m}$ is a prescribed symmetric positive definite matrix. The goal of this exercise is to define and study the corresponding variant of OMP, that will be called Σ-OMP.

1. Show that if Σ is symmetric positive definite then there exists a unique symmetric positive definite matrix M of the same size such that $\Sigma = M^2$. Such a matrix is denoted $\Sigma^{1/2}$.
2. Show that a vector $x \in \mathbb{R}^m$ has unit $\|\cdot\|_\Sigma$ norm if, and only if, $x' := \Sigma^{1/2}x$ has unit Euclidean norm
3. Give a simple characterization of the fact that a dictionary $D = [d_1, \ldots, d_K] \in \mathbb{R}^{m \times K}$ has all its columns d_i with unit $\|\cdot\|_\Sigma$ norm
4. Consider a vector $r \in \mathbb{R}^m$ and a dictionary $D = [d_1, \ldots, d_K] \in \mathbb{R}^{m \times K}$, and denote $r' := \Sigma^{1/2}r$, $D' = \Sigma^{1/2}D$. Characterize $\arg\max_i |\langle r, d_i \rangle_\Sigma|$ in terms of r' and D'.
5. Given a vector $y \in \mathbb{R}^m$ and a full-rank matrix $A \in \mathbb{R}^{m \times s}$ with $s \leq m$, the orthogonal projection $P_A^\Sigma y$ of y onto the range of A *in the sense of the weighted norm* $\|\cdot\|_\Sigma$ is defined as $P_A^\Sigma y = \arg\min_{z \in \texttt{range}(A)} \|y - z\|_\Sigma$. Characterize this projection in terms of $y' := \Sigma^{1/2}y$ and $A' := \Sigma^{1/2}A$.
6. Write the pseudo-code of Σ-OMP, where every step of OMP involving an inner product (resp. a Euclidean norm) is replaced by the corresponding step with their weighted equivalent. Comment your pseudo-code and specify the requirements on the input of the algorithm
7. What is the equivalent for Σ-OMP of the Exact Recovery Condition for OMP? Justify.

⚠ MASTER STUDENTS ONLY AFTER EXERCISES 1 AND 2 ⚠

(☕ probably) Exercise 3: About the LARS algorithm

Least-angle regression (LARS) is an algorithm for fitting linear regression models to high-dimensional data, developed by Bradley Efron, Trevor Hastie, Iain Johnstone and Robert Tibshirani in 2004. The goal of this exercise is to study two equivalent formulations of LARS and investigate their differences with Orthogonal Matching Pursuit (OMP) and LASSO algorithms.

Assume that you observe $y \in \mathbb{R}^m$ and you wan to fit a sparse regression model

$$y = Ax + \varepsilon$$

where $A \in \mathbb{R}^{m\times n}$ is a known matrix, $x \in \mathbb{R}^n$ is an (unknown) sparse target vector, and $\varepsilon \in \mathbb{R}^m$ is some noise. We will denote by

$$z := A^\top y \quad \text{and} \quad \Sigma := A^\top A$$

the so-called residual vector z and covariance matrix Σ. We assume that Σ has rank $r \geq 2$.

Notation: The (i, j) entry of Σ is denoted by $\Sigma_{i,j}$. Given p reals $a_1, \ldots, a_p$, we denote by $(a_1 \cdots a_p) \in \mathbb{R}^{1\times p}$ a row vector and by $(a_1, \ldots, a_p) \in \mathbb{R}^{p\times 1}$ a (column) vector. We denote by $A_i \in \mathbb{R}^m$ the i-th column of A.

Given $\imath_1, \ldots, \imath_k \in [n] := \{1, \ldots, n\}$, we denote by $\Sigma_{(\imath_1,\ldots,\imath_k)} \in \mathbb{R}^{k\times k}$ the sub-matrix of Σ keeping the columns and the rows indexed by $\{\imath_1, \ldots, \imath_k\}$.

For the sake of simplicity, and without loss of generality, we assume that $x \in \mathbb{R}^n$ has non-negative entries, namely $x_i \geq 0$ for all $i \in [n]$. The expressions of the LARS algorithm of this exercise are tailored under this assumption. We begin with the standard formulation below.

Algorithm 1 LARS algorithm (standard formulation)

Data: Residual vector z and covariance matrix Σ.

Result: Sequence $((\lambda_k, \imath_k))_{k\geq 1}$ where $\lambda_1 \geq \lambda_2 \geq \ldots > 0$ are the so-called knots, and $\imath_1, \imath_2, \ldots$ are the variables that enter the model.

`/* Initialize computing` $(\lambda_1, \imath_1)$ `and residual` $z^{(1)}$. `*/`

1 Set $k = 1$, $\lambda_1 := \max_{i\in[n]} z_i$, $\imath_1 := \arg\max_{i\in[n]} z_i$ and $z^{(1)} := z$.

`/* Note that` $((\lambda_\ell, \imath_\ell))_{1\leq \ell\leq k-1}$ `and` $z^{(k-1)}$ `have been defined at the previous iteration. */`

2 Set $k \leftarrow k + 1$ and compute the so-called *least-squares fit*

$$\eta_j := \big(\Sigma_{j,\imath_1} \cdots \Sigma_{j,\imath_{k-1}}\big)\Sigma^{-1}_{(\imath_1,\ldots,\imath_{k-1})}(1, \ldots, 1)\,, \quad \forall j \in [n]\,. \tag{1}$$

3 For $0 < \lambda \leq \lambda_{k-1}$ define the intermediate residual $z^{(k-\frac{1}{2})}(\lambda) = (z_1^{(k-\frac{1}{2})}(\lambda), \ldots, z_n^{(k-\frac{1}{2})}(\lambda))$ given by

$$z_j^{(k-\frac{1}{2})}(\lambda) := z_j^{(k-1)} - (\lambda_{k-1} - \lambda)\eta_j\,, \quad j = 1, \ldots, n\,,$$

and pick

$$\begin{aligned}
\lambda_k &:= \max\big\{\beta \geq 0 \;:\; \exists\, j \notin \{\imath_1, \ldots, \imath_{k-1}\},\ \text{s.t. } z_j^{(k-\frac{1}{2})}(\beta) = \beta\big\}\,,\\
\imath_k &:= \underset{j\notin\{\imath_1,\ldots,\imath_{k-1}\}}{\arg\max}\; z_j^{(k-\frac{1}{2})}(\lambda_k)\,,\\
z^{(k)} &:= z^{(k-\frac{1}{2})}(\lambda_k)\,.
\end{aligned}$$

Then, iterate from **2**.

The first questions might help you to understand the rationale behind this algorithm. The first step ($k = 1$) of Algorithm 1 finds the maximum of the residual z and it holds that

$$\lambda_1 = z_{\imath_1} \quad \text{and} \quad \lambda_1 \geq z_i \,,\ \forall i \in [n]\,.$$

We assume that **this argument maximum is unique**. When $A^\top \varepsilon$ is distributed according to a law that is absolutely continuous with respect to the Lebesgue measure, one can prove that the maximum is almost surely unique (prove it assuming that ε as standard Gaussian density). Note that $z^{(1)} = z$ and $z^{(1)}_{\imath_1} = \lambda_1$.
For $k = 2$ one compute $\eta \in \mathbb{R}^n$ by (1).

1. Show that
$$\eta = \frac{A^\top A_{\imath_1}}{\|A_{\imath_1}\|_2^2} \quad \text{and} \quad \eta_{\imath_1} = 1$$
in (1) for $k = 2$.
2. Show that, for every $0 < \lambda \leq \lambda_1$
$$z^{(\frac{3}{2})}(\lambda) = z - (\lambda_1 - \lambda)\eta = A^\top \Big(y - \frac{(\lambda_1 - \lambda)}{\|A_{\imath_1}\|_2^2} A_{\imath_1} \Big) \quad \text{and} \quad z^{(\frac{3}{2})}_{\imath_1}(\lambda) = \lambda\,.$$
3. Show that $0 \leq \lambda_2 < \lambda_1$.
4. Show that $z^{(2)}_{\imath_1} = z^{(2)}_{\imath_2} = \lambda_2$.
5. Assume that $\imath_1$ is the unique argument maximum of $\max_{i \in [n]} |z_i|$. We recall that $\imath_1$ is defined as the argument maximum of $\max_{i \in [n]} z_i$. Assume that $\imath_2$ is the unique argument maximum of $\max_{j \neq \imath_1} |z^{(\frac{3}{2})}_j(\lambda_2)|$. We recall that $\imath_2$ is defined as the argument maximum of $\max_{j \neq \imath_1} z^{(\frac{3}{2})}_j(\lambda_2)$. Show that for all λ such that $\lambda_2 < \lambda < \lambda_1$, the LASSO
$$\min_{x^\# \in \mathbb{R}^n} \Big\{ \frac{1}{2} \|y - Ax^\#\|_2^2 + \lambda \|x^\#\|_1 \Big\}$$
has a unique solution $x^{(\lambda)}$ given by
$$x^{(\lambda)}_{\imath_1} = \frac{(\lambda_1 - \lambda)}{\|A_{\imath_1}\|_2^2} \quad \text{and} \quad x^{(\lambda)}_j = 0\,,\ \text{for } j \neq \imath_1.$$
(Hint: use the KKT condition and $z^{(\frac{3}{2})}(\lambda)$)

We have shown that the first iteration of the LARS algorithm is related to ℓ_1-minimization. Consider the next iteration ($k = 3$).

6. Consider η in (1) for $k = 3$:
$$\eta = \big[\Sigma_{\imath_1} \Sigma_{\imath_2}\big] \Sigma^{-1}_{(\imath_1, \imath_2)}(1, 1)\,,$$
where Σ_i is the i-th column of Σ and $\big[\Sigma_{\imath_1} \Sigma_{\imath_2}\big] \in \mathbb{R}^{n \times 2}$. Show that $\eta = A^\top u$ where $u \in \mathbb{R}^m$ is the solution to
$$u = \operatorname*{arg\,min}_{v \in \mathbb{R}^m} \Big\{ \|v\|_2^2 \ :\ A^\top_{\imath_1} v = A^\top_{\imath_2} v = 1 \Big\}\,.$$
We understand now why η is called the least-squares fit in (1).
(Hint: Consider the Lagrangian $\mathcal{L}(v; w_1, w_2) := \|v\|_2^2 - 2(A^\top_{\imath_1} v - 1)w_1 - 2(A^\top_{\imath_2} v - 1)w_2$)
7. Show that, for every $0 < \lambda \leq \lambda_2$
$$z^{(\frac{5}{2})}(\lambda) = A^\top \Big(y - \frac{(\lambda_1 - \lambda_2)}{\|A_{\imath_1}\|_2^2} A_{\imath_1} - (\lambda_2 - \lambda) \big[A_{\imath_1} A_{\imath_2}\big] \Sigma^{-1}_{(\imath_1, \imath_2)}(1, 1) \Big) \quad \text{and} \quad z^{(\frac{5}{2})}_{\imath_1}(\lambda) = z^{(\frac{5}{2})}_{\imath_2}(\lambda) = \lambda$$
8. Assume that $\imath_2$ is the unique argument maximum of $\max_{j \neq \imath_1} z^{(\frac{3}{2})}_j(\lambda_2)$ and show that $0 \leq \lambda_3 < \lambda_2 < \lambda_1$.
9. Show that $z^{(3)}_{\imath_1} = z^{(3)}_{\imath_2} = z^{(3)}_{\imath_3} = \lambda_3$.

Remark: In the context of linear regression in high dimensions, the LARS algorithm can be used to identify a subset of potential covariates. The LARS outputs a piecewise affine solutions path, and the *knots* $\lambda_1 > \lambda_2 > \cdots > 0$ are the change points of the LARS path that are built by tracking the ℓ_∞ of the residual (in our case, the maximum of $z^{(k)}$). At each knot, the LARS algorithm adds to the active set of variables the covariate the most correlated with the actual residual. In that way, the descent direction is always equiangular to all variables present in the current active set (see the definition of η). This sequence of knots is closely related to the sequence of knots of LASSO, as they differ by only one rule: "*Only in the LASSO case, if a nonzero coefficient crosses zero before the next variable enters, drop it from the active set and recompute the current joint least-squares direction*", as mentioned in standard books on the subject.