Final Exam
Inverse problems and sparsity
Parcimonie et grande dimension
The LASSO (the good), the IRLS (the bad) and the Universal Distortion Property (the ugly)

Wednesday 25th of March, 2026
9.00 am to 11.00 am

- This subject has two exercises, the exercises are independent and can be treated separately.
- **Master students** will be evaluated on all the subject. They are asked to treat exercise 1 and 2 **before** the extra part of exercise 2. The subject is not meant to be finished in $2h$ and the last part of the exercise 2 concerns only Master students who 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 this 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 **4 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 exercise 2 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 l'exercice 2. Pour chaque jeu de copies vous devrez écrire votre nom et votre prénom sur la première page.

Exercise 1: Iteratively Reweighted Least Squares (IRLS)

The Iteratively Reweighted Least Squares (IRLS) algorithm is a computationally efficient method to approximate the solution of the ℓ_1-minimization problem. Consider the noiseless high-dimensional linear model $y = X\beta^*$ where $X \in \mathbb{R}^{n\times p}$ is the known design matrix and $y \in \mathbb{R}^n$ are the observations. We assume that the target vector $\beta^* \in \mathbb{R}^p$ is exactly s-sparse with support S, and remember that $\beta^*_j = 0$ for $j \notin S$.

To find a sparse solution, the IRLS algorithm iteratively solves *weighted* ℓ_2-minimization problems. Let $\epsilon_m > 0$ be a regularizing parameter at iteration m. Given a previous iterate $\beta^{(m-1)} \in \mathbb{R}^p$, we define the weights $w_j^{(m)}$ for $j = 1, \dots, p$ as:

$$w_j^{(m)} = \left((\beta_j^{(m-1)})^2 + \epsilon_m^2\right)^{-1/2}$$

The next iterate $\beta^{(m)}$ is then computed as the solution to the quadratically constrained weighted least squares problem:

$$\beta^{(m)} = \arg\min_{\beta\in\mathbb{R}^p} \sum_{j=1}^{p} w_j^{(m)}\beta_j^2 \quad \text{subject to} \quad X\beta = y.$$

The goal of this exercise is to investigate the *local convergence rate* of this algorithm, which means how fast the iterative scheme goes towards the solution when well initialized. Let

$$W_m = \operatorname{diag}(w_1^{(m)}, \dots, w_p^{(m)}),$$

be the diagonal weight matrix which is a function of the previous iterate $\beta_j^{(m-1)}$. Let $h^{(m)} = \beta^{(m)} - \beta^*$ be the approximation error at iteration m.

1. **Optimality condition.** By using the method of Lagrange multipliers or basic projection principles, show that $\beta^{(m)}$ satisfies the orthogonality condition $\langle W_m\beta^{(m)}, v\rangle = 0$ for all $v \in \ker X$.

2. **Error decomposition.** Note that $h^{(m)} \in \ker X$. Using the orthogonality condition from Question 1, prove that:

$$\sum_{j=1}^{p} w_j^{(m)}(h_j^{(m)})^2 = -\sum_{j\in S} w_j^{(m)} h_j^{(m)}\beta_j^*$$

3. **Bounding the weighted error.** Using Question 2, deduce (with details) that:

$$\|W_m^{1/2}h^{(m)}\|_2^2 \le \Big(\sum_{j\in S} w_j^{(m)}(h_j^{(m)})^2\Big)^{1/2}\Big(\sum_{j\in S} w_j^{(m)}(\beta_j^*)^2\Big)^{1/2},$$

   where $W_m^{1/2} := \operatorname{diag}(\sqrt{w_1^{(m)}}, \dots, \sqrt{w_p^{(m)}})$

4. **Exploiting the Null Space Property.** Assume that the matrix X satisfies the Null Space Property (NSP) of order s with constant $0 < \rho < 1$, which implies that for all $v \in \ker X$, $\|v_S\|_1 \le \rho\|v_{S^c}\|_1$. Prove that the ℓ_1 norm of the error on the complement S^c satisfies the following relation:

$$\|h_{S^c}^{(m)}\|_1 \le \left(\sum_{j\in S^c}(w_j^{(m)})^{-1}\right)^{1/2}\left(\sum_{j\in S^c} w_j^{(m)}(h_j^{(m)})^2\right)^{1/2}$$

   Then, using the NSP, deduce an upper bound for the total error $\|h^{(m)}\|_1$ as a function of $\|W_m^{1/2}h^{(m)}\|_2$ and ρ.

☕ 5. **Local linear convergence with regularizer.** We maintain the regularization parameter $\varepsilon_m > 0$ such that the weights are $w_j^{(m)} = \left((\beta_j^{(m-1)})^2 + \varepsilon_m^2\right)^{-1/2}$. Let $\beta^*_{\min} = \min_{j\in S}|\beta_j^*| > 0$.

   Assume the matrix satisfies the NSP with constant $\rho < 1$, and that the initialization satisfies $\|h^{(m-1)}\|_1 \le \gamma\beta^*_{\min}$ for a given constant $\gamma \in (0, 1)$. Assume ρ is sufficiently small such that the constant μ defined below satisfies $\mu < 1$.

   (a) Show that on the complement S^c:

$$\sum_{j\in S^c}(w_j^{(m)})^{-1} \le \|h^{(m-1)}\|_1 + p\varepsilon_m$$

(b) Using the initialization condition, show that for $j \in S$, $w_j^{(m)} \leq \frac{1}{(1-\gamma)|\beta_j^*|}$. Deduce explicit upper bounds for $\max_{j\in S} w_j^{(m)}$ and $\sum_{j\in S} w_j^{(m)}(\beta_j^*)^2$.

(c) By bounding $\left(\sum_{j\in S} w_j^{(m)}(h_j^{(m)})^2\right)^{1/2} \leq \left(\max_{j\in S} w_j^{(m)}\right)^{1/2} \|h_S^{(m)}\|_1$, combine the inequalities established in Questions 3 and 4 to prove that the total error satisfies:

$$\|h^{(m)}\|_1 \leq \mu \|h^{(m-1)}\|_1 + \mu p \varepsilon_m$$

where $\mu = \frac{\rho(1+\rho)}{1-\gamma}\left(\frac{\|\beta_S^*\|_1}{\beta_{\min}^*}\right)^{1/2}$. Conclude on the convergence rate of the IRLS algorithm as $\varepsilon_m \to 0$.

Exercise 2: The Universal Distortion Property (UDP) and the Lasso

Consider the high-dimensional linear model where one observes a vector $y \in \mathbb{R}^n$ such that $y = X\beta^* + \epsilon$, where $X \in \mathbb{R}^{n\times p}$ is the known design matrix, $\beta^* \in \mathbb{R}^p$ is an unknown target vector one would like to recover, and $\epsilon \in \mathbb{R}^n$ is a stochastic error term. We recall the definition of the lasso estimator introduced by R. Tibshirani:

$$\hat{\beta} \in \underset{\beta\in\mathbb{R}^p}{\arg\min} \left\{ \frac{1}{2}\|X\beta - y\|_2^2 + \lambda\|\beta\|_1 \right\}$$

where $\lambda > 0$ denotes a tuning parameter. A standard hypothesis in high-dimensional regression requires that one can provide a constant $\lambda_0 \in \mathbb{R}$ such that $\|X^\top \epsilon\|_\infty \leq \lambda_0$ with an overwhelming probability. In this exercise, we introduce a new condition called the Universal Distortion Property (UDP).

Definition (UDP): Given $1 \leq S_0 \leq p$ and $0 < \kappa_0 < 1/2$, we say that a matrix $X \in \mathbb{R}^{n\times p}$ satisfies the universal distortion condition $UDP(S_0, \kappa_0, \Delta)$ if and only if for all $\gamma \in \mathbb{R}^p$, for all integers $s \in \{1, ..., S_0\}$, and for all subsets $S \subseteq \{1, ..., p\}$ such that $|S| = s$ it holds:

$$\|\gamma_S\|_1 \leq \Delta\sqrt{s}\|X\gamma\|_2 + \kappa_0\|\gamma\|_1$$

Part I: Oracle Inequality in Estimation

Let $h = \hat{\beta} - \beta^*$ be the estimation error. Throughout this section, assume that the tuning parameter satisfies $\lambda > \frac{\lambda_0}{1-2\kappa_0}$ (which guarantees $\frac{1}{2}(1 - \frac{\lambda_0}{\lambda}) - \kappa_0 > 0$). Let $S \subseteq \{1, ..., p\}$ be a fixed subset of indices of size $s \leq S_0$ corresponding to the s largest coefficients of β^* in absolute value.

1. By using the very definition of $\hat{\beta}$ as the minimizer of the lasso objective function, write down the basic inequality relating the objective evaluated at $\hat{\beta}$ and at β^*.

2. Expand the squared ℓ_2 norms to show the following intermediate inequality:

$$\frac{1}{2}\|Xh\|_2^2 - \langle X^\top\epsilon, h\rangle + \lambda\|\hat{\beta}\|_1 \leq \lambda\|\beta^*\|_1$$

3. Using the assumption $\|X^\top\epsilon\|_\infty \leq \lambda_0$, prove that:

$$|\langle X^\top\epsilon, h\rangle| \leq \lambda_0\|h\|_1$$

4. Justify that:

$$\|\hat{\beta}\|_1 = \|\beta^* + h\|_1 \geq \|\beta_S^*\|_1 - \|h_S\|_1 + \|h_{S^c}\|_1 - \|\beta_{S^c}^*\|_1$$

5. By rearranging the terms (explain which ones), deduce that:

$$\frac{1}{2\lambda}\left[\frac{1}{2}\|Xh\|_2^2 + (\lambda - \lambda_0)\|h\|_1\right] \leq \|h_S\|_1 + \|\beta_{S^c}^*\|_1$$

6. Assume the design matrix X satisfies the $UDP(S_0, \kappa_0, \Delta)$ condition and show that:

$$\left[\frac{1}{2}\left(1 - \frac{\lambda_0}{\lambda}\right) - \kappa_0\right]\|h\|_1 \leq \left(-\frac{1}{4\lambda}\|Xh\|_2^2 + \Delta\sqrt{s}\|Xh\|_2\right) + \|\beta_{S^c}^*\|_1$$

7. Conclude that the lasso satisfies the following so-called oracle inequality in estimation:

$$\|\hat{\beta} - \beta^*\|_1 \leq \frac{2}{(1 - \frac{\lambda_0}{\lambda}) - 2\kappa_0} \left(\lambda \Delta^2 s + \|\beta^*_{S^c}\|_1\right)$$

Hint (astuce): Consider the real function $f(x) = -\frac{1}{4\lambda}x^2 + \Delta\sqrt{s}x$.

8. Explain how each term in the bound can be interpreted, possibly in connexion with the course.

⚠ MASTER STUDENTS ONLY AFTER EXERCISES 1 AND 2 ⚠

Exercise 2 continued

Part II: Oracle Inequality in Prediction

We now wish to bound the prediction error $\|X\hat{\beta} - X\beta^*\|_2$.

9. Assume the result in Question 6 (if not yet proven), and show that:

$$\|Xh\|_2^2 - 4\lambda\Delta\sqrt{s}\|Xh\|_2 \leq 4\lambda\|\beta^*_{S^c}\|_1$$

10. Deduce the oracle inequality for the prediction error:

$$\|X\hat{\beta} - X\beta^*\|_2 \leq 4\lambda\Delta\sqrt{s} + \frac{\|\beta^*_{S^c}\|_1}{\Delta\sqrt{s}}$$

Part III: Bonus - UDP and the Restricted Isometry Property (RIP)

We recall the Restricted Isometry Property: A matrix X satisfies $RIP(k, \theta)$ if $(1-\theta)\|\gamma\|_2^2 \leq \|X\gamma\|_2^2 \leq (1+\theta)\|\gamma\|_2^2$ for all k-sparse vectors γ.

11. Let $T_0 \subseteq \{1, ..., p\}$ be a support of size s. Assume that the condition $\|\gamma_{T_0}\|_1 > \kappa_0\|\gamma\|_1$ holds. Show that:

$$\|\gamma_{T_0^c}\|_1 < \frac{1 - \kappa_0}{\kappa_0}\|\gamma_{T_0}\|_1$$

12. Prove that if X satisfies the $RIP(5s, \theta)$ condition with $\theta < \sqrt{2} - 1$, then X satisfies the $UDP(s, \kappa_0, \Delta)$ condition for properly chosen parameters κ_0 and Δ. *Hint: If* $\|\gamma_{T_0}\|_1 \leq \kappa_0\|\gamma\|_1$ *the result is immediate. Otherwise, use Question 11. Denote by* T_1 *the set of the indices of the* $4s$ *largest coefficients in* T_0^c, T_2 *the next* $4s$ *largest, etc. Decompose* $T_0^c = T_1 \cup T_2 \cup ...$ *and use the properties of RIP to bound* $\|X\gamma\|_2$ *from below.*