EXAMEN

Duration : 2h. All the documents can be used. The quality of writing is one of the evaluation criteria.

Problem 1 [12 points]

Let X_i, $i = 1, \ldots, n$ dbe iid random variables with a square integrable density f^* on $[0, 1]$. We denote by $\langle g, g' \rangle$ the usual scalar product in $L_2([0, 1])$. Let $\{\varphi_k : k \in \mathbb{N}\}$ the sequence of functions defined by

$$\varphi_{2k}(x) = \sqrt{2}\cos(2\pi kx), \qquad \varphi_{2k+1}(x) = \sqrt{2}\cos(2\pi kx), \qquad k = 0, 1, 2, \ldots.$$

We denote by $\ell_2(\mathbb{N})$ the set of all square summable sequences $(u_k)_{k\in\mathbb{N}}$. The usual scalar product of two sequences $u, u' \in \ell_2(\mathbb{N})$ is denoted by $\langle u, u' \rangle = \sum_{k\in\mathbb{N}} u_k u'_k$.

Let $\vartheta^*_k = \langle f^*, \varphi_k \rangle$ stand for the Fourier coefficient of the unknown density function f^*, for every $k \in \mathbb{N}$. Furthermore, let

$$\widehat{\vartheta}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(X_i)$$

the empirical Fourier coefficients. The goal of this problem is to provide an adaptive estimator of f^* based on hard thresholding of the empirical Fourier coefficients.

1. For a threshold $\lambda > 0$, we set $\mathcal{A}_k = \big\{|\widehat{\vartheta}_k - \vartheta^*_k| \le \lambda\big\}$ and $\mathcal{A} = \big\{\max_{k=0,\ldots,n-1} |\widehat{\vartheta}_k - \vartheta^*_k| \le \lambda\big\}$.
   (a) Prove that for all $\lambda > 0$,
   $$\mathbf{P}(\mathcal{A}_k) \ge 1 - 2\exp\Big\{-\frac{n\lambda^2}{16}\Big\}, \qquad \forall k \in \mathbb{N}.$$
   and $\mathcal{A}^c$ is the contrary of $\mathcal{A}$, is exponentially close to zero.
   (b) Deduce from the previous result that for all $\lambda > 0$,
   $$\mathbf{P}(\mathcal{A}) \ge 1 - 2n\exp\Big\{-\frac{n\lambda^2}{16}\Big\}.$$
   (c) For a prescribed tolerance level $\delta \in (0, 1)$ (close to zero), determine a value $\lambda > 0$ such that $\mathbf{P}(\mathcal{A}) \ge 1 - \delta$. What is the rate of convergence of this value $\lambda = \lambda_n$ to zero as $n \to +\infty$.
2. Let $\widehat{T} = (\widehat{T}_j)_{j\in\mathbb{N}}$ be the thresholded version of the empirical Fourier coefficients :
   $$\widehat{T}_k = \begin{cases} \widehat{\vartheta}_k; & \text{if } |\widehat{\vartheta}_k| \ge 2\lambda \text{ and } k < n,\\ 0, & \text{otherwise.}\end{cases}$$
   Prove that on the event $\mathcal{A}$, we have
   $$(\widehat{T}_k - \vartheta^*_k)^2 \le 9\vartheta^{*2}_k \wedge \lambda^2, \qquad \forall k = 0, 1, \ldots, n-1.$$
   Deduce from the above inequality that
   $$\|\widehat{T} - \vartheta^*\|_2^2 := \sum_{k\in\mathbb{N}} (\widehat{T}_k - \vartheta^*_k)^2 \le 9 \min_{K=1,\ldots,n-1} \Big(K\lambda^2 + \sum\nolimits_{k\ge K} \vartheta^{*2}_k\Big).$$
3. Let $\beta \in \mathbb{N}$ and $L > 0$ be constants. We assume now that f^* is β-times continuously differentiable and belongs to the Sobolev ball $\Sigma(\beta, L)$ defined by
   $$\Sigma(\beta, L) = \Big\{g : [0, 1] \to \mathbb{R} : \int_0^1 g^{(\beta)}(x)^2\,dx \le L; g^{(\ell)}(0) = g^{(\ell)}(1), \forall \ell < \beta\Big\}.$$
   We recall that this implies the following constraint on the Fourier coefficients of f^* :
   $$\sum_{k\in\mathbb{N}} k^{2\beta}\vartheta^{*2}_k \le L. \tag{1}$$
   Prove that the estimator $\widehat{f}(x) = \sum_k \widehat{T}_k \varphi_k(x)$ satisfies, for some $C > 0$, the inequality
   $$\mathbf{P}\Big(\|\widehat{f} - f^*\|_2^2 \le C\Big\{\frac{\log(2n/\delta)}{n}\Big\}^{2\beta/(2\beta+1)}\Big) \ge 1 - \delta.$$

Problème 2 [14 points]

We consider the multiple linear regression model

$$\mathbf{y} = \Phi\boldsymbol{\beta}^* + \boldsymbol{\xi},$$

where $\mathbf{y} \in \mathbb{R}^n$ and $\Phi \in \mathbb{R}^{n\times p}$ are the observed signal and the given design matrix, $\boldsymbol{\beta}^* \in \mathbb{R}^p$ is the unknown regression vector and $\boldsymbol{\xi} \in \mathbb{R}^n$ is the random noise vector assumed to be drawn from $\mathcal{N}_n(0, \sigma^2\mathbf{I}_n)$ distribution [1]. We consider the case of large p and small ℓ_0-norm $s = \|\boldsymbol{\beta}^*\|_0$. In such a situation, one can estimate $\boldsymbol{\beta}^*$ by the Dantzig selector

$$\widehat{\boldsymbol{\beta}} \in \arg\min_{\boldsymbol{\beta}\in\mathcal{C}} \|\boldsymbol{\beta}\|_1, \tag{2}$$

where $\mathcal{C} = \{\boldsymbol{\beta} \in \mathbb{R}^p : \frac{1}{n}\|\Phi^\top(\mathbf{y} - \Phi\boldsymbol{\beta})\|_\infty \leq \lambda\}$.

1. Prove that the set of all possible solutions of (2) is a convex subset of $\mathbb{R}^p$. Furthermore, check that any two vectors belonging to this set have equal ℓ_1 norms.
2. <u>In what follows, we assume that</u> $\max_j \frac{1}{n}\|\Phi^j\|_2^2 \leq 1$. We also admit the following fact : if Z is a Gaussian random variable with zero mean and unit variance, then $\mathbf{P}(|Z| \geq x) \leq e^{-x^2/2}$ for all $x > 0$. Prove that the event
$$\mathcal{B} = \Big\{\frac{1}{n}\|\Phi^\top\boldsymbol{\xi}\|_\infty \leq \sigma\sqrt{\frac{2}{n}\log(p/\delta)}\Big\}$$
has a probability $\geq 1 - \delta$.
3. <u>In the sequel, we assume that λ is chosen to be equal to</u> $\sigma\sqrt{\frac{2}{n}\log(p/\delta)}$.
   (a) Prove that on the event $\mathcal{B}$, we have $\boldsymbol{\beta}^* \in \mathcal{C}$ and $\|\widehat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}^*\|_1$.
   (b) Check that $\mathbf{P}(\frac{1}{n}\|\Phi^\top\Phi(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_\infty \leq 2\lambda) \geq 1 - \delta$.
4. Let $J = \{j : \beta_j^* \neq 0\}$ be the support of $\boldsymbol{\beta}^*$ and set $\boldsymbol{u} = \widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*$. We recall that J^c is the complementary set of J in $\{1, \ldots, p\}$ and $\boldsymbol{u}_J$ is the vector obtained from $\boldsymbol{u}$ by removing all the coordinates with indices in J^c. Prove that on $\mathcal{B}$, it holds that
$$\|\boldsymbol{u}_{J^c}\|_1 \leq \|\boldsymbol{u}_J\|_1.$$
Hint : use the fact that $\|\widehat{\boldsymbol{\beta}}\|_1 \leq \|\boldsymbol{\beta}^*\|_1$.
5. Using the fact that $\|\Phi\boldsymbol{u}\|_2^2 = \boldsymbol{u}^\top\Phi^\top\Phi\boldsymbol{u}$, prove that
$$\mathbf{P}\Big(\|\boldsymbol{u}_{J^c}\|_1 \leq \|\boldsymbol{u}_J\|_1 \text{ and } \frac{1}{n}\|\Phi\boldsymbol{u}\|_2^2 \leq 4\lambda\|\boldsymbol{u}_J\|_1\Big) \geq 1 - \delta.$$
6. Remaind the condition of Restricted Eigenvalues (RE).
7. Prove that if Φ satisfies the condition RE$(1, s)$ with constant κ, then
$$\mathbf{P}\Bigg(\frac{1}{n}\|\Phi(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_2^2 \leq \underbrace{\frac{16\lambda^2 s}{\kappa}}_{=32\frac{\sigma^2 s\log(p/\delta)}{n\kappa}}\Bigg) \geq 1 - \delta.$$
8. Compare this bound (in terms of conditions and sharpness) with the one for the Lasso seen in the classroom.

1. We recall that this means that $\xi_1, \ldots, \xi_n$ are iid Gaussian random variables with $\mathbf{E}(\xi_i) = 0$ and $\mathbf{Var}(\xi_i) = \sigma^2$.