# EXAMEN

**Durée : 2h. Tous les documents sont autorisés. Les 2 problèmes sont indépendants. La qualité de rédaction fait partie des critères d'évaluation.**

## Problème 1 [12 points]

Soient $X_i$, $i = 1, \ldots, n$ des variables iid de densité $f$ sur $[0,1]$. On suppose que $f \in L_2([0,1])$ et désigne par $\langle g, g' \rangle$ le produit scalaire usuel de deux fonctions de $L_2$. Soit $\{\varphi_k : k \in \mathbb{N}\}$ une base orthonormée de $L_2([0,1])$. On note $\ell_2(\mathbb{N})$ l'ensemble des suites $(u_k)_{k\in\mathbb{N}}$ de carré sommables. Pour deux suites $u, u' \in \ell_2(\mathbb{N})$, on utilise le produit scalaire usuel $\langle u, u' \rangle = \sum_{k\in\mathbb{N}} u_k u'_k$.

1. On pose $\vartheta_k = \langle f, \varphi_k \rangle$ pour tout $k \in \mathbb{N}$. Montrer que l'estimation de $f$ équivaut à l'estimation de la suite $\vartheta = (\vartheta_k)_{k\in\mathbb{N}}$ pourvu que l'erreur d'estimation soit mesurée dans les deux cas par l'espérance du carré de la distance entre l'estimateur et la vraie valeur.
2. Montrer que
$$\widehat{\vartheta}_k = \frac{1}{n}\sum_{i=1}^{n} \varphi_k(X_i)$$
est un estimateur sans biais de $\vartheta_k$. Vérifier que la variance de $\widehat{\vartheta}_k$ est donnée par
$$\frac{1}{n}\bigg(\int_0^1 \varphi_k^2(x) f(x)\,dx - \vartheta_k^2\bigg).$$
En supposant que $\sup_{x\in[0,1]} f(x) \le L$, montrer que la variance de $\widehat{\vartheta}_k$ est bornée par $L/n$.
3. On suppose maintenant qu'il existe une constante $\kappa > 0$ telle que $(\varphi_k)$ vérifient $\varphi_{2p}(x)^2 + \varphi_{2p+1}(x)^2 \ge \kappa$ pour tout $x$ et pour tout $p \in \mathbb{N}$. Montrer que le risque de l'estimateur $\widehat{\vartheta} = (\widehat{\vartheta}_k)_{k\in\mathbb{N}}$ est égal à $+\infty$. Quelle est la raison de cette explosion du risque : le sur-lissage (underfitting) ou le sous-lissage (overfitting) ?
4. Pour pallier ce défaut, on utilise un estimateur de seuillage dur ($\lambda > 0$ est un paramètre de réglage) :
$$\widehat{\vartheta}_k^\lambda = \begin{cases} \widehat{\vartheta}_k; & \text{si } |\widehat{\vartheta}_k| \ge 2\lambda \text{ et } k \le n, \\ 0, & \text{sinon.} \end{cases}$$
On considère dans la suite que les fonctions $\varphi_k$ sont toutes bornées par une constante $C_\varphi$. Donner une majoration exponentielle de la probabilité de l'événement $\mathcal{A}^c$ où $\mathcal{A} = \big\{ \sup_{k=1,\ldots,n} |\widehat{\vartheta}_k - \vartheta_k| \le \lambda \big\}$ et $\mathcal{A}^c$ désigne le contraire de $\mathcal{A}$.
5. Prouver que sur l'événement $\mathcal{A}$, on a
$$\sum_{k=0}^{n} (\widehat{\vartheta}_k^\lambda - \vartheta_k)^2 \le \sum_{k=0}^{n} \vartheta_k^2 \wedge \lambda^2.$$
En déduire que
$$\|\widehat{\vartheta}^\lambda - \vartheta\|_2^2 \le \min_{K=1,\ldots,n} \Big(K\lambda^2 + \textstyle\sum_{k>K} \vartheta_k^2\Big).$$
6. On suppose maintenant que les coefficients $\vartheta_k$ de la vraie densité, en plus d'être de carré sommables, vérifient la propriété d'ellipsoïde suivante :
$$\sum_{k\in\mathbb{N}} k^{2\beta} \vartheta_k^2 \le L$$
où $\beta$ et $L$ sont deux constantes positives. Proposer un estimateur $\widehat{f}^\lambda$ de $f$ et une borne supérieure (aussi petite que possible) de l'erreur d'estimation $\|\widehat{f}^\lambda - f\|_2^2$ sur l'événement $\mathcal{A}$.
7. Soit $\delta \in (0,1)$ un niveau de tolérance. Déduire des questions précédentes le choix du paramètre $\lambda$ tel que
$$\mathbf{P}\bigg(\|\widehat{f}^\lambda - f\|_2^2 \le C\Big\{\frac{\log(n/\delta)}{n}\Big\}^{2\beta/(2\beta+1)}\bigg) \ge 1 - \delta,$$
pour une constante $C$.
8. Discuter les propriétés d'optimalité de la vitesse de convergence de $\widehat{f}^\lambda$ et son adaptativité à la régularité de la fonction inconnue.

## Problème 2 [10 points]

On considère le modèle de régression à design fixe :

$$Y_i = f(x_i) + \sigma\xi_i, \qquad i = 1, \dots, n.$$

Soit $\{\phi_1, \dots, \phi_M\}$ un dictionnaire dans lequel $f$ admet une représentation parcimonieuse :

$$f(x) = \sum_{m=1}^{M} \vartheta_m^* \phi_m(x), \qquad \text{avec} \qquad s = \|\boldsymbol{\vartheta}^*\|_0 \ll M.$$

Dans ce qui suit, on utilise la notation $f_{\boldsymbol{\vartheta}}(x) = \sum_{m=1}^{M} \vartheta_m \phi_m(x)$. Soit $\widehat{\boldsymbol{\vartheta}}$ l'estimateur Lasso défini par

$$\widehat{\boldsymbol{\vartheta}} \in \arg\min_{\boldsymbol{\vartheta} \in \mathbb{R}^M} \left\{ \frac{1}{n} \sum_{i=1}^{n} \Big( Y_i - f_{\boldsymbol{\vartheta}}(x_i) \Big)^2 + \lambda \sum_{m=1}^{M} w_m |\vartheta_m| \right\}, \tag{1}$$

où $\lambda$ est un paramètre positif et $w_m = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \phi_m(x_i)^2}$.

1. On introduit la matrice
$$\mathbf{X} = \begin{pmatrix} \phi_1(x_1) & \dots & \phi_M(x_1) \\ \vdots & \ddots & \vdots \\ \phi_1(x_n) & \dots & \phi_M(x_n) \end{pmatrix}$$
Montrer que
$$\widehat{\boldsymbol{\vartheta}} \in \arg\min_{\boldsymbol{\vartheta} \in \mathbb{R}^M} \left\{ \frac{1}{n} \|\mathbf{Y} - \mathbf{X}\boldsymbol{\vartheta}\|_2^2 + \lambda \sum_{m=1}^{M} w_m |\vartheta_m| \right\}. \tag{2}$$

2. Posons $Z_m = \frac{1}{n} \sum_{i=1}^{n} Y_i \phi_m(x_i)$. Donner la valeur explicite des coordonnées de $\widehat{\boldsymbol{\vartheta}}$ dans le cas où la matrice $\frac{1}{n}\mathbf{X}^\top \mathbf{X}$ est diagonale.

3. On considère dans la suite le cas d'une matrice $\mathbf{X}$ générale. Montrer que si $\widehat{\boldsymbol{\vartheta}}$ et $\widehat{\boldsymbol{\vartheta}}'$ sont deux solutions de (2), alors pour tout $\alpha \in (0,1)$, $\alpha\widehat{\boldsymbol{\vartheta}} + (1-\alpha)\widehat{\boldsymbol{\vartheta}}'$ est aussi solution de (2). Prouver également que
$$\sum_{m=1}^{M} w_m |\widehat{\vartheta}_m| = \sum_{m=1}^{M} w_m |\widehat{\vartheta}'_m| \qquad \text{et} \qquad \mathbf{X}\widehat{\boldsymbol{\vartheta}} = \mathbf{X}\widehat{\boldsymbol{\vartheta}}'.$$

4. Soit $\mathcal{B}$ l'événement $\{4\sigma|X_m^\top \boldsymbol{\xi}| \leq n w_m \lambda; \forall m = 1, \dots, M\}$ où $X_m$ est la $m^{\text{ème}}$ colonne de $\mathbf{X}$. Montrer que sur cet événement, on a
$$\frac{1}{n} \|\mathbf{X}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^*)\|_2^2 \leq \lambda \sum_{m=1}^{M} w_m \Big( \frac{1}{2} |\widehat{\vartheta}_m - \vartheta_m^*| + |\vartheta_m^*| - |\widehat{\vartheta}_m| \Big).$$

5. Soit $J^* = \{m : \vartheta_m^* \neq 0\}$ et $\boldsymbol{\Delta} = \boldsymbol{\vartheta}^* - \widehat{\boldsymbol{\vartheta}}$. Déduire de la question précédente que sur $\mathcal{B}$, on a
$$\sum_{m \notin J^*} w_m |\Delta_m| \leq 3 \sum_{m \in J^*} w_m |\Delta_m|.$$

6. On suppose maintenant que $\mathbf{X}$ vérifie la condition des valeurs propres restreintes : il existe $\kappa > 0$ tel que pour tout ensemble $J$ et tout vecteur $\mathbf{v} \in \mathbb{R}^M$ vérifiant
$$|J| \leq s, \qquad \text{et} \qquad \sum_{m \in J^c} w_m |v_m| \leq 3 \sum_{m \in J} w_m |v_m|$$
on a
$$\frac{1}{n} \|\mathbf{X}\mathbf{v}\|_2^2 \geq \kappa \sum_{m \in J} w_m^2 v_m^2.$$
Montrer que pour une constante $C > 0$, sur l'événement $\mathcal{B}$, on a
$$\frac{1}{n} \|\mathbf{X}(\widehat{\boldsymbol{\vartheta}} - \boldsymbol{\vartheta}^*)\|_2^2 \leq C\lambda s.$$