Apprentissage Statistique et grande dimension

Projet 6 : Estimation dans le modèle linéaire fonctionnel

Nathan Assouline & Bèdis Mzali

Mai 2017

Partie 1 : Introduction

Notre objectif est de proposer et comparer différents estimateurs dans le cadre de la statistique pour données fonctionnelles. Notre modèle est le suivant :

$$Y = \beta + \varepsilon$$

Cette dernière s'oppose à la statistique classique car nous faisons l'hypothèse que les observations ne sont pas des vecteurs de $\mathbb{R}^d$ mais des fonctions. La dimension de l'espace des observations est donc supposée infinie. On cherche donc à déterminer un estimateur de β tel que le risque $R(Y, \beta)$ soit le plus petit possible et nous minimiserons ce risque sur la classe des fonctions $\mathbb{L}^2([0, 1])$.
Par exemple, une application de ces modèles serait d'apprendre la relation entre la courbe spéctrométrique X d'un échantillon d'eau et la concentration en nitrites Y de manière à pouvoir déterminer automatiquement celle d'un nouvel échantillon.
Après avoir étudier le modèle, nous commencerons par une estimation par méthode de projections puis utiliser l'estimateur par ACP fonctionnelle, et enfin nous passerons la régression Ridge. Nous terminerons pas une étude numérique à l'aide du langage Python.

Partie 2 : Étude théorique du sujet

Pour alléger les notations nous supposons, sans perte de généralité, que X_1 est centré.

1. Étude du modèle

Question a)

Nous pouvons réecrire le modèle sous forme vectorielle :

$$Y = \beta + \xi$$

En posant : $Y = \Big[Y_1 \cdots Y_n\Big]^T$, $\beta = \Big[\langle \beta, X_1 \rangle \cdots \langle \beta, X_n \rangle\Big]^T$ et $\xi = \Big[\xi_1 \cdots \xi_n\Big]^T$.

On a bien $Y, \beta, \xi \in \mathbb{R}^n$ et β et ξ indépendants.

Question b)

Dans le cadre du modèle (1) on a :

$$\begin{aligned}
\mathbb{E}(X_1Y_1) &= \mathbb{E}[(\langle \beta, X_1\rangle + \xi_1)X_1] \\
&= \mathbb{E}(\langle \beta, X_1\rangle X_1) + \mathbb{E}(\xi_1)\mathbb{E}(X_1) \\
&= \mathbb{E}(\langle \beta, X_1\rangle X_1) \\
&= \Gamma\beta
\end{aligned}$$

En effet, X_1 et ξ_1 sont indépendant, et ξ_1 est centré, ce qui justifie le passage à la deuxième et troisième ligne.

Ensuite, par analogie avec le modèle linéaire multivarié, on remarque que Γ est un opérateur de covariance dans ce modèle linéaire fonctionnel.

Question d)

Soient $f, g \in \mathbb{L}^2([0,1])$, montrons que $\langle \Gamma f, g\rangle = \langle f, \Gamma g\rangle$, c'est-à-dire que Γ est auto-adjoint.

$$\begin{aligned}
\langle \Gamma f, g\rangle &= \langle \mathbb{E}(\langle f, X_1\rangle X_1), g\rangle \\
&= \int_0^1 \mathbb{E}[\langle f, X_1\rangle X_1(t) g(t)]dt \\
&= \mathbb{E}\left(\langle f, X_1\rangle \int_0^1 X_1(t)g(t)dt\right) (*) \\
&= \mathbb{E}(\langle f, X_1\rangle\langle X_1, g\rangle) \\
&= \mathbb{E}(\langle f, X_1\langle X_1, g\rangle\rangle) \\
&= \mathbb{E}\left(\int_0^1 f(t)X_1(t)\langle X_1, g\rangle dt\right) \\
&= \int_0^1 f(t)\mathbb{E}[X_1(t)\langle X_1, g\rangle]dt (*) \\
&= \langle f, \mathbb{E}(X_1\langle X_1, g\rangle)\rangle \\
&= \langle f, \Gamma g\rangle
\end{aligned}$$

Nous obtenons les lignes $(*)$ en appliquant le théorème de Fubini cas général.
Montrons que l'utilisation de ce théorème est licite (nous faisons le calcul pour la 1ère utilisation, la seconde est analogue).

$$
\begin{aligned}
\int_0^1 \mathbb{E}(|\langle f, X_1\rangle X_1(t)g(t)|)dt &= \mathbb{E}\left(\int_0^1 |\langle f, X_1\rangle X_1(t)g(t)|dt\right) &(1)\\
&= \mathbb{E}\left(|\langle f, X_1\rangle| \int_0^1 |X_1(t)g(t)|dt\right) &(2)\\
&\leq \|f\|_2 \|g\|_2 \mathbb{E}\left(||X_1||_2^2\right) &(3)\\
&< \infty &(4)
\end{aligned}
$$

On obtient la ligne (1) par le théorème de Fubini cas positif, la ligne (3) par l'inégalité de Cauchy-Schwarz et la dernière en supposant que $\mathbb{E}(\|X_1\|_2^2) < \infty$.
Conclusion : Γ est bien un opérateur auto-adjoint.

Γ étant un opérateur auto-adjoint ses valeurs propres sont réelles. De plus, d'après le théorème d'Hilbert-Schmitt, il existe une base de Hilbert de $\mathbb{L}^2([0,1])^2$ de fonctions propres de Γ.

Question e)

Nous souhaitons démontrer que pour tout $j \in \mathbb{N}$, il existe $\lambda_j \in \mathbb{R}$ tel que $\Gamma\varphi_j = \lambda_j\varphi_j$.

Tout d'abord, X_1 étant 1-périodique on note que la fonction C est 1-périodique. Aussi, C est continue sur $[-1,1]$. En effet, on applique le théorème de continuité sous l'intégrale, on rappelle que X_1 est continue p.s. et bornée (car périodique). Ainsi, C est un 1-périodique et appartient $\mathbb{L}^1([-1,1])$.
On remarque aussi que les $(\varphi_j)_{j\geq 1}$ sont 1-périodique et appartiennent $\mathbb{L}^1([-1,1])$.

Soient $j \in \mathbb{N}$ et $t \in [0,1]$,

$$
\begin{aligned}
\Gamma\varphi_j(t) &= \mathbb{E}(\langle X_1, \varphi_j\rangle X_1(t))\\
&= \mathbb{E}\left(\int_0^1 X_1(u)X_1(t)\varphi_j(u)du\right)\\
&= \int_0^1 \mathbb{E}(X_1(u)X_1(t))\varphi_j(u)du\\
&= \int_0^1 C(t-u)\varphi_j(u)du\\
&= C * \varphi_j(t)\\
&= \int_0^1 C(u)\varphi_j(t-u)du
\end{aligned}
$$

La troisième ligne s'obtient encore une fois avec le théorème de Fubini.

Ainsi on obtient $\Gamma\varphi_1 = \lambda_1\varphi_1$ simplement en posant $\lambda_1 = \int_0^1 C(u)du$.
De plus,

$$\begin{aligned}
\Gamma\varphi_{2j}(t) &= \int_0^1 C(u)\varphi_{2j}(t-u)du \\
&= \int_0^1 C(u)\sqrt{2}\cos(2\pi j(t-u))du \\
&= \int_0^1 C(u)\sqrt{2}[\cos(2\pi jt)\cos(2\pi ju) + \sin(2\pi jt)\sin(2\pi ju)]du \\
&= \varphi_{2j}(t)\int_0^1 C(u)\cos(2\pi ju)du + \sqrt{2}\sin(2\pi jt)\int_0^1 C(u)\sin(2\pi ju)du
\end{aligned}$$

On a ensuite :

$$\begin{aligned}
\int_0^1 C(u)\sin(2\pi ju)du &= \int_0^1 C(u-1)\sin(2\pi ju)du \quad \text{en utilisant la périodicité de C} \\
&= \int_{-1}^0 C(u)\sin(2\pi j(u+1))du \quad \text{en effectuant un changement de variable} \\
&= \int_{-1}^0 C(u)\sin(2\pi ju)du \\
&= \frac{1}{2}\int_{-1}^1 C(u)\sin(2\pi ju)du \\
&= 0
\end{aligned}$$

Ainsi on obtient $\Gamma\varphi_{2j} = \lambda_{2j}\varphi_{2j}$ simplement en posant $\lambda_{2j} = \int_0^1 C(u)\cos(2\pi ju)du$.

De façon analogue on montre que $\Gamma\varphi_{2j+1} = \lambda_{2j+1}\varphi_{2j+1}$ avec $\lambda_{2j+1} = \int_0^1 C(u)\cos(2\pi ju)du$.

Conclusion : la base trigonométrique $(\varphi_j)_{j\geq 1}$ est bien une base de fonctions propres de Γ.

2. Estimation par projection

Passons à une seconde partie où nous étudions l'estimation par projection.

Question a)

Pour définir un estimateur des moindres carrés pour ce modèle, on peut poser :

$$\hat{\beta}_M^{(EMC)} \in \underset{\beta\in Vect\{\varphi_1,\cdots,\varphi_M\}}{\arg\min} \frac{1}{n}\sum_{i=1}^n (Y_i - \langle\beta, X_i\rangle)^2 = \underset{\beta\in Vect\{\varphi_1,\cdots,\varphi_M\}}{\arg\min} \gamma_n(\beta)$$

Question b)

Etudions le risque $R(\hat{\beta}_M^{(EMC)}, \beta)$ de l'estimateur des moindres carrés.

Question (a)

$$\gamma_n(\beta_M) - \gamma_n\left(\hat{\beta}_M^{(EMC)}\right) = \frac{1}{n}\sum_{i=1}^{n}(\langle \beta - \beta_M, X_i\rangle + \xi_i)^2 - \frac{1}{n}\sum_{i=1}^{n}(\langle \beta - \hat{\beta}_M^{(EMC)}, X_i\rangle + \xi_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\langle \beta-\beta_M, X_i\rangle)^2 + 2\langle \beta-\beta_M, X_i\rangle\xi_i + \xi_i^2 - \frac{1}{n}\sum_{i=1}^{n}(\langle \beta-\hat{\beta}_M^{(EMC)}, X_i\rangle)^2 + 2\langle \beta-\hat{\beta}_M^{(EMC)}, X_i\rangle\xi_i + \xi_i^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(\langle \beta - \beta_M, X_i\rangle)^2 - \frac{1}{n}\sum_{i=1}^{n}(\langle \beta - \hat{\beta}_M^{(EMC)}, X_i\rangle)^2 + \frac{2}{n}\sum_{i=1}^{n}\langle \hat{\beta}_M^{(EMC)} - \beta_M, X_i\rangle\xi_i$$

$$= \|\beta_M - \beta\|_n^2 - \left\|\hat{\beta}_M^{(EMC)} - \beta\right\|_n^2 + 2\nu_n\left(\hat{\beta}_M^{(EMC)} - \beta_M\right)$$

Question (b)

Majorons $2\nu_n\left(\beta_M - \hat{\beta}_M^{(EMC)}\right)$: soit $\delta > 0$,

$$\begin{aligned} 2\nu_n\left(\beta_M - \hat{\beta}_M^{(EMC)}\right) &= 2\nu_n\left(\frac{\beta_M - \hat{\beta}_M^{(EMC)}}{\left\|\beta_M - \hat{\beta}_M^{(EMC)}\right\|_n}\right)\left\|\beta_M - \hat{\beta}_M^{(EMC)}\right\|_n \\ &\leq \delta\left\|\beta_M - \hat{\beta}_M^{(EMC)}\right\|_n^2 + \delta^{-1}\nu_n^2\left(\frac{\beta_M - \hat{\beta}_M^{(EMC)}}{\left\|\beta_M - \hat{\beta}_M^{(EMC)}\right\|_n}\right)^2 \\ &\leq \delta\left\|\beta_M - \hat{\beta}_M^{(EMC)}\right\|_n^2 + \delta^{-1}\sup_{t\in B_M}\nu_n^2(t) \end{aligned}$$

On obtient la deuxième ligne en utilisant l'inégalité suggérée par l'énoncé, et la troisième en utilisant le fait que $\frac{\beta_M-\hat{\beta}_M^{(EMC)}}{\left\|\beta_M-\hat{\beta}_M^{(EMC)}\right\|_n} \in B_M^{(n)}$ (car : $\beta_M - \hat{\beta}_M^{(EMC)} \in \mathrm{Vect}\{\varphi_1, \cdots, \varphi_M\}$ et $\left\|\frac{\beta_M-\hat{\beta}_M^{(EMC)}}{\left\|\beta_M-\hat{\beta}_M^{(EMC)}\right\|_n}\right\|_n^2 = 1$).

Question (c)

Premièrement, par le théorème de Fubini on a :

$$\forall i, j \in \mathbb{N}^*, \mathbb{E}(\langle X_1, \varphi_i\rangle\langle X_1, \varphi_j\rangle) = \mathbb{E}(\langle X_1\langle X_1, \varphi_j\rangle, \varphi_i\rangle) = \langle \Gamma\varphi_j, \varphi_i\rangle$$

Ensuite,

$$
\begin{aligned}
\mathbb{E}\left(\sup_{t\in B_{M,1/2}} \nu_n^2(t)\right) &= \mathbb{E}\left(\sup_{t\in B_{M,1/2}} \frac{1}{n^2}\left(\sum_{i=1}^n \xi_i\langle t, X_i\rangle\right)^2\right) \\
&= \frac{1}{n^2}\mathbb{E}\left(\sup_{t\in B_{M,1/2}} \sum_{i=1}^n \xi_i^2\langle t, X_i\rangle^2 + 2\sum_{1\le i<j\le n} \xi_i\langle t, X_i\rangle\xi_j\langle t, X_j\rangle\right) \\
&\le \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^n \xi_i^2 \sup_{t\in B_{M,1/2}}\langle t, X_i\rangle^2 + 2\sum_{1\le i<j\le n} \xi_i\xi_j \sup_{t\in B_{M,1/2}}\langle t, X_i\rangle\langle t, X_j\rangle\right) \\
&= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^n \xi_i^2 \sup_{t\in B_{M,1/2}}\langle t, X_i\rangle^2\right) + 2\sum_{1\le i<j\le n} \mathbb{E}(\xi_i)\mathbb{E}(\xi_j)\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\langle t, X_i\rangle\langle t, X_j\rangle\right) \\
&= \frac{\sigma^2}{n^2}\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\langle t, X_1\rangle^2\right) \text{ car } \xi_i \text{ indépendant de } X_i \\
&= \frac{\sigma^2}{n^2}\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\langle\sum_{j=1}^M\langle t, \varphi_j\rangle\varphi_j, X_1\rangle^2\right) \\
&= \frac{\sigma^2}{n^2}\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\sum_{j=1}^M\langle\langle t, \varphi_j\rangle\varphi_j, X_1\rangle^2 + 2\sum_{1\le i<j\le M}\langle\langle t, \varphi_i\rangle\varphi_i, X_1\rangle\langle\langle t, \varphi_j\rangle\varphi_j, X_1\rangle\right) \\
&= \frac{\sigma^2}{n^2}\left(\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\sum_{j=1}^M\langle\langle t, \varphi_j\rangle\varphi_j, X_1\rangle^2\right) + 2\sum_{1\le i<j\le M}\langle\Gamma\varphi_i, \varphi_j\rangle\langle t, \varphi_i\rangle\langle t, \varphi_j\rangle\right) \\
&= \frac{\sigma^2}{n^2}\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\sum_{j=1}^M\langle t, \varphi_j\rangle^2\langle\varphi_j, X_1\rangle^2\right) \text{car } \langle\Gamma\varphi_i, \varphi_j\rangle = \lambda_i\langle\varphi_i, \varphi_j\rangle = 0(*) \\
&\le \frac{\sigma^2}{n^2}\mathbb{E}\left(\sum_{j=1}^M \sup_{t\in B_{M,1/2}}\|t\|^2\|\varphi_j\|^2\langle\varphi_j, X_1\rangle^2\right) \text{Inégalité de Cauchy-Schwarz} \\
&\le \frac{\sigma^2}{2n}\sum_{j=1}^M \mathbb{E}\left(\langle\varphi_j, X_1\rangle^2\right) \\
&= \frac{\sigma^2}{2n}\sum_{j=1}^M\langle\Gamma\varphi_j, \varphi_j\rangle
\end{aligned}
$$

Pour (*) on suppose que $(\varphi_j)_{j\ge1}$ est une base de fonctions propres de Γ.

<u>Question (d)</u>

Soit $\delta\in]0,1/2[$,

$$\begin{aligned}\mathbb{E}\left(\left\|\hat{\beta}_M^{(EMC)}-\beta\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right) &\le \mathbb{E}\left(\left(\|\beta_M-\beta\|_n^2+2\nu_n\left(\hat{\beta}_M^{(EMC)}-\beta_M\right)\right)\mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)\\ &= \mathbb{E}\left(\|\beta_M-\beta\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)+\mathbb{E}\left(2\nu_n\left(\hat{\beta}_M^{(EMC)}-\beta_M\right)\mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)\\ &\le \mathbb{E}\left(\|\beta_M-\beta\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)+\delta\mathbb{E}\left(\left\|\beta_M-\hat{\beta}_M^{(EMC)}\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)\\ &+\delta^{-1}\mathbb{E}\left(\sup_{t\in B_{M,1/2}}\nu_n^2(t)\right) \text{ par la question 2.b}\end{aligned}$$

De plus,

$$\begin{aligned}\mathbb{E}\left(\left\|\beta_M-\hat{\beta}_M^{(EMC)}\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right) &= \mathbb{E}\left(\left\|\beta_M-\beta+\beta-\hat{\beta}_M^{(EMC)}\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)\\ &\le 2\mathbb{E}\left(\|\beta_M-\beta\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)+2\mathbb{E}\left(\left\|\hat{\beta}_M^{(EMC)}-\beta\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)\end{aligned}$$

Puis, en utilisant la question (c) on arrive à :

$$\begin{aligned}(1-2\delta)\mathbb{E}\left(\left\|\hat{\beta}_M^{(EMC)}-\beta\right\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right) &\le (1+2\delta)\mathbb{E}\left(\|\beta_M-\beta\|_n^2 \mathbb{1}_{B_M^{(n)}\subset B_{M,1/2}}\right)+\delta^{-1}\frac{\sigma^2}{2n}\sum_{j=1}^M\langle\Gamma\varphi_j,\varphi_j\rangle\\ &\le (1+2\delta)\mathbb{E}\left(\|\beta_M-\beta\|_n^2\right)+\delta^{-1}\frac{\sigma^2}{2n}\sum_{j=1}^M\langle\Gamma\varphi_j,\varphi_j\rangle\end{aligned}$$

3. ACP fonctionnelle

Passons maintenant l'étude de l'ACP fonctionnelle.

Question a)

Il nous suffit de reprendre les raisonnements de la question d) la partie 1. En effet, on montre de manière analogue que Γ_n est auto-adjoint, et Γ_n est aussi compact (car de rang fini). Donc comme précedemment ces deux résultats prouvent que la base des $(\hat{\varphi}_j)_{j\ge 1}$ de fonctions propres de Γ_n existe.

Question b)

La question 1) de la partie 2 nous donne :

$$\hat{\beta}_M^{(PCA,MC)} \in \underset{\beta\in Vect\{\hat{\varphi_1},\cdots,\hat{\varphi_M}\}}{\arg\min}\ \frac{1}{n}\sum_{i=1}^n(Y_i-\langle\beta,X_i\rangle)^2$$

Par conséquent, on pose la fonction :

$$f:(a_1,\cdots,a_M)\mapsto\frac{1}{n}\sum_{i=1}^n\left(Y_i-\sum_{k=1}^M a_k\langle\hat{\varphi}_k,X_i\rangle\right)^2$$

Ainsi, nous passons au problème équivalent : $\arg\min_{a \in \mathbb{R}^M} f(a_1, \cdots, a_M)$.

Soit $j \in \{1, \cdots, M\}$,

$$
\begin{aligned}
\frac{\partial f}{\partial a_j}(a) &= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i \rangle \left(Y_i - \sum_{k=1}^{M} a_k \langle \hat{\varphi}_k, X_i \rangle \right) \\
&= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + \frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i \rangle \sum_{k=1}^{M} a_k \langle \hat{\varphi}_k, X_i \rangle \\
&= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + 2 \sum_{k=1}^{M} a_k \frac{1}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_k, X_i \rangle \langle \hat{\varphi}_j, X_i \rangle \\
&= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + 2 \sum_{k=1}^{M} \langle \Gamma_n \hat{\varphi}_k, \hat{\varphi}_j \rangle a_k \\
&= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + 2 \sum_{k=1}^{M} \hat{\lambda}_k \langle \hat{\varphi}_k, \hat{\varphi}_j \rangle a_k \\
&= -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + 2 \hat{\lambda}_j a_j \text{ Car les } (\hat{\varphi}_j)_{j \geq 1} \text{ sont orthonormales.}
\end{aligned}
$$

Passons ensuite aux conditions du premier ordre :

$$
\begin{aligned}
\frac{\partial f}{\partial a_j}(a) = 0 &\Leftrightarrow -\frac{2}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle + 2 \hat{\lambda}_j a_j = 0 \\
&\Leftrightarrow a_j = \frac{\frac{1}{n} \sum_{i=1}^{n} \langle \hat{\varphi}_j, X_i Y_i \rangle}{\hat{\lambda}_j} \\
&\Leftrightarrow a_j = \frac{\langle \hat{\varphi}_j, \frac{1}{n} \sum_{i=1}^{n} X_i Y_i \rangle}{\hat{\lambda}_j} \\
&\Leftrightarrow a_j = \frac{\langle \hat{\varphi}_j, \widehat{\mathbb{E}(XY)} \rangle}{\hat{\lambda}_j}
\end{aligned}
$$

D'où le résultat souhaité.

4. Régression Ridge

Terminons l'étude théorique par la régression Ridge.

Question a)

On cherche à résoudre le problème :

$$
\underset{\beta \in \mathbb{L}^2([0,1])}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^{n} (Y_i - \langle \beta, X_i \rangle)^2 + \rho \|\beta\|_2^2 \right\}
$$

Pour cela on pose la fonction :

$$H : \beta \mapsto \frac{1}{n}\sum_{i=1}^{n}(Y_i - \langle \beta, X_i\rangle)^2 + \rho \left\|\beta\right\|_2^2$$

Puis on calcule son gradient :

$$\begin{aligned}\nabla H(\beta) &= -\frac{2}{n}\sum_{i=1}^{n} X_i(Y_i - \langle \beta, X_i\rangle) + 2\rho\beta \\ &= -2\widehat{\mathbb{E}(XY)} + 2\Gamma_n\beta + 2\rho\beta\end{aligned}$$

On en déduit la condition du premier ordre (condition suffisante par convexité) :

$$\begin{aligned}\nabla H(\beta) = 0 &\Leftrightarrow -2\widehat{\mathbb{E}(XY)} + 2\Gamma_n\beta + 2\rho\beta = 0 \\ &\Leftrightarrow (\Gamma_n + \rho I)\beta = \widehat{\mathbb{E}(XY)} \text{ on suppose alors que } \Gamma_n + \rho I \text{ est inversible} \\ &\Leftrightarrow \hat{\beta}^{(R)} = (\Gamma_n + \rho I)^{-1}\widehat{\mathbb{E}(XY)}\end{aligned}$$

Ainsi on a bien trouvé une expression de $\hat{\beta}^{(R)}$.

<u>Question b)</u>

Pour finir, pour le cas M grand, on réecrit le problème comme : $\underset{a\in\mathbb{R}^M}{\arg\min}$, avec :

$$g : (a_1, \cdots, a_M) \mapsto \frac{1}{n}\sum_{i=1}^{n}\left(Y_i - \sum_{k=1}^{M} a_k\langle \varphi_k, X_i\rangle\right)^2 + \rho\sum_{i=1}^{M} a_i^2$$

Soit $j \in \{1, \cdots, M\}$,

$$\begin{aligned}\frac{\partial g}{\partial a_j}(a) &= -\frac{2}{n}\sum_{i=1}^{n}\langle \varphi_j, X_iY_i\rangle + 2\sum_{k=1}^{M}\langle \Gamma_n\varphi_k, \varphi_j\rangle a_k + 2\rho a_j \\ &= -\frac{2}{n}\sum_{i=1}^{n}\langle \varphi_j, X_iY_i\rangle + 2\left(\sum_{k=1}^{M}\langle \Gamma_n\varphi_k, \varphi_j\rangle a_k\mathbb{1}_{k\neq j} + (\rho + \langle \Gamma_n\varphi_j, \varphi_j\rangle)a_j\right) \\ &= -2b_j^M + 2(A^M a)_j\end{aligned}$$

Avec

$$b^M = \left(\frac{1}{n}\sum_{i=1}^{n}\langle \varphi_1, X_iY_i\rangle, \cdots, \frac{1}{n}\sum_{i=1}^{n}\langle \varphi_M, X_iY_i\rangle\right)^T$$

$$A^M = \begin{pmatrix} \rho + \langle \Gamma_n\varphi_1, \varphi_1\rangle & \langle \Gamma_n\varphi_2, \varphi_1\rangle & \cdots & \langle \Gamma_n\varphi_M, \varphi_1\rangle \\ \langle \Gamma_n\varphi_1, \varphi_2\rangle & \rho + \langle \Gamma_n\varphi_2, \varphi_2\rangle & \ddots & \langle \Gamma_n\varphi_M, \varphi_2\rangle \\ \vdots & \cdots & \ddots & \vdots \\ \langle \Gamma_n\varphi_1, \varphi_M\rangle & \cdots & \cdots & \rho + \langle \Gamma_n\varphi_M, \varphi_M\rangle \end{pmatrix}$$

D'où :

$$\nabla g(a) = -2b^M + 2A^M a$$

Puis avec les conditions du premier ordre (en supposant A^M inversible) :

$$a^* = (A^M)^{-1} b^M$$

Ce qui nous amène à la conclusion :

$$\boxed{\hat{\beta}^{(R)} = \sum_{k=1}^{M} \left((A^M)^{-1} b^M\right)_k \varphi_k}$$