Exercice 3 :

1) . $l_D \in S_D$ donc $l_D = \sum_{j=1}^{D} a_j \varphi_j$ $a_j \in \mathbb{R}$

. $l - l_D \perp S_D$ donc $\forall j = 1, \dots, D$ $\langle l - l_D, \varphi_j \rangle = 0$

donc $\forall j = 1, \dots, D$ $\langle l, \varphi_j \rangle = \langle l_D, \varphi_j \rangle$

or $\langle l_D, \varphi_{j_0} \rangle = \sum_{j=1}^{D} a_j \underbrace{\langle \varphi_{j_0}, \varphi_j \rangle}_{\begin{cases} 1 \text{ si } j = j_0 \\ 0 \text{ sinon} \end{cases}} = a_{j_0}$

donc $\forall j = 1, \dots, D$ $a_j = \langle l_D, \varphi_j \rangle = \langle l, \varphi_j \rangle$.

et finalement, $\boxed{l_D = \sum_{j=1}^{D} \langle l, \varphi_j \rangle \varphi_j}$

2) $\mathbb{E}[Y_1 t(X_1)] = \mathbb{E}[(m(X_1) + \varepsilon_1) t(X_1)]$

$= \mathbb{E}[m(X_1) t(X_1) + \varepsilon_1 t(X_1)]$

$= \mathbb{E}[m(X_1) t(X_1)] + 0$ car $\mathbb{E}[\varepsilon_1] = 0$

$= \int_0^1 m(x) t(x) f(x) \, dx$ déf. de l'espérance

$= \langle l, t \rangle$ car $l := mf$

3) $\mathbb{E}[\gamma_n(t)] = \|t\|_2^2 - \frac{2}{n} \sum_{i=1}^{n} \mathbb{E}[Y_i t(X_i)]$

$= \|t\|_2^2 - 2 \langle l, t \rangle$ (X_i, Y_i) iid et question 2)

$= \|l - t\|^2 - \|l\|^2$

Donc minimiser $\mathbb{E}[\gamma_n(t)]$ revient à minimiser $\|l - t\|^2$ ie à chercher $t \in S_D$ le plus proche de l au sens L^2.

4) On cherche la solution $\hat{\ell}_D = \underset{t \in S_D}{\operatorname{argmin}} \gamma_m(t)$.

On calcule le gradient puis on cherche les points critiques :

$\forall t \in S_D \quad \exists (a_j)_j \quad tq \qquad t = \sum_{j=1}^{D} a_j \varphi_j$

et $\gamma_m(t) = \|t\|_2^2 - \frac{2}{m} \sum_{i=1}^{m} Y_i\, t(X_i)$

$$= \sum_{j=1}^{D} a_j^2 - \frac{2}{m} \sum_{i=1}^{m} Y_i \sum_{j=1}^{D} a_j \varphi_j(X_i)$$

$$= \sum_{j=1}^{D} a_j^2 - \frac{2}{m} \sum_{j=1}^{D} a_j \left(\sum_{i=1}^{m} Y_i \varphi_j(X_i) \right)$$

donc $\forall k \in [\![1, D]\!] \qquad \frac{\partial \gamma_m}{\partial a_k} = 2a_k - \frac{2}{m} \sum_{i=1}^{m} Y_i \varphi_k(X_i)$

donc $\begin{pmatrix} \frac{\partial \gamma_m}{\partial a_1} \\ \vdots \\ \frac{\partial \gamma_m}{\partial a_D} \end{pmatrix} = \begin{pmatrix} 0 \\ | \\ 0 \end{pmatrix} \iff \boxed{\forall k \in [\![1, D]\!] \quad a_k = \frac{1}{m} \sum_{i=1}^{m} Y_i \varphi_k(X_i)}$

Ce point critique est bien un minimum car $\left(\frac{\partial^2 \gamma_m}{\partial a_k \partial a_l} \right)_{k,l} = 2\,Id \gg 0$

Donc, $\hat{\ell}_D = \underset{t \in S_D}{\operatorname{argmin}} \gamma_m(t) = \sum_{j=1}^{D} \hat{a}_j \varphi_j \qquad$ où $\hat{a}_j = \frac{1}{m} \sum_{i=1}^{m} Y_i \varphi_k(X_i)$

5) $\mathbb{E}\left[\|\ell_D - \hat{\ell}_D\|_2^2\right] = \mathbb{E}\left[\left\|\sum_{j=1}^{D} (\langle \ell, \varphi_j \rangle - \hat{a}_j)\, \varphi_j\right\|_2^2\right]$ d'après 1) et 4)

$$= \mathbb{E}\left[\sum_{j=1}^{D} (\langle \ell, \varphi_j \rangle - \hat{a}_j)^2\right]$$ base orthonormée

$$= \sum_{j=1}^{D} \mathbb{E}\left[(\langle \ell, \varphi_j \rangle - \hat{a}_j)^2\right]$$ linéarité de $\mathbb{E}$

or $\hat{a}_j = \frac{1}{m} \sum_{i=1}^{m} Y_i \varphi_j(X_i)$ donc $\mathbb{E}[\hat{a}_j] = \langle \ell, \varphi_j \rangle$ (question 2) avec $t = \varphi_j$)

donc $\mathbb{E}\left[(\langle \ell, \varphi_j \rangle - \hat{a}_j)^2\right] = \mathbb{E}\left[(\mathbb{E}(\hat{a}_j) - \hat{a}_j)^2\right] = \operatorname{Var}(\hat{a}_j)$

Et finalement, on obtient :

$$\mathbb{E}(\|\ell_D - \hat{\ell}_D\|_2^2) = \sum_{j=1}^{D} \mathrm{Var}(\hat{a}_j) = \sum_{j=1}^{D} \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m} Y_i \varphi_j(X_i)\right).$$

6) c'est du cours.

7) aussi

8) $\mathbb{E}[\|\ell - \hat{\ell}_D\|_2^2] = \mathbb{E}[\|\ell - \ell_D + \ell_D - \hat{\ell}_D\|_2^2]$

$$= \|\ell - \ell_D\|_2^2 + \mathbb{E}[\|\ell_D - \hat{\ell}_D\|_2^2]$$

Pythagore car $\ell - \ell_D \perp S_D$ et $\ell_D - \hat{\ell}_D \in S_D$

il suffit donc de montrer que $\mathbb{E}[\|\ell_D - \hat{\ell}_D\|_2^2] \leq \frac{1}{m}(M + \sigma^2)D$.

D'après la question 5),

$$\mathbb{E}(\|\ell_D - \hat{\ell}_D\|_2^2) = \sum_{j=1}^{D} \mathrm{Var}\left(\frac{1}{m}\sum_{i=1}^{m} Y_i \varphi_j(X_i)\right)$$

$$= \frac{1}{m^2}\sum_{j=1}^{D} \mathrm{Var}\left(\sum_{i=1}^{m} Y_i \varphi_j(X_i)\right)$$

$$\overset{(X_i, Y_i) \perp\!\!\!\perp}{=} \frac{1}{m^2}\sum_{j=1}^{D}\sum_{i=1}^{m} \mathrm{Var}(Y_i \varphi_j(X_i))$$

$$\overset{(X_i, Y_i)\text{ identiquement distribués}}{=} \frac{1}{m^2}\, m \sum_{j=1}^{D} \mathrm{Var}(Y_1 \varphi_j(X_1))$$

$$\overset{\mathrm{Var}(Z) \leq \mathbb{E}[Z^2]}{\leq} \frac{1}{m}\sum_{j=1}^{D} \mathbb{E}[Y_1^2 \varphi_j^2(X_1)]$$

$$\overset{\text{linéarité de l'}\mathbb{E}.}{=} \frac{1}{m}\mathbb{E}\left[Y_1^2 \sum_{j=1}^{D} \varphi_j^2(X_1)\right]$$

$$\leq \frac{1}{m}\left\|\sum_{j=1}^{D} \varphi_j^2\right\|_\infty \mathbb{E}[Y_1^2]$$

$$\overset{7)\text{ et } \varepsilon_1 \perp\!\!\!\perp X_1}{\leq} \frac{D}{m}\left(\mathbb{E}[m^2(X_1)] + \mathbb{E}[\varepsilon_1^2] + 2\,\mathbb{E}[\varepsilon_1]\mathbb{E}[m(X_1)]\right)$$

$$\leq \frac{D}{m}(M + \sigma^2 + 0)$$

D'où le résultat annoncé.

9) Avec la majoration proposée, on obtient pour le risque quadratique :

$$\mathbb{E}\left[\|\ell - \hat{\ell}_D\|_2^2\right] \leq \underbrace{C D^{-2\beta} + \frac{1}{m}(\Pi + \sigma^2) D}_{=: \Psi(D)} \qquad \text{(question 8)}$$

on cherche D_{opt} qui minimise $\Psi(D)$.

$$\Psi'(D) = -2\beta C D^{-2\beta-1} + \frac{1}{m}(\Pi+\sigma^2)$$

donc $\Psi'(D) = 0 \iff D^{-2\beta-1}(2\beta C) = \frac{1}{m}(\Pi+\sigma^2)$

$$\iff D^{2\beta+1} = \frac{2m\beta C}{\Pi+\sigma^2}$$

$$\iff D = \left(\frac{2m\beta C}{\Pi+\sigma^2}\right)^{\frac{1}{2\beta+1}}$$

c'est bien un min car $\Psi''(D) = -2\beta(-2\beta-1) C D^{-2\beta-2} \geq 0$

donc $\boxed{D_{opt} = \left(\frac{2\beta C}{\Pi+\sigma^2}\right)^{\frac{1}{2\beta+1}} m^{\frac{1}{2\beta+1}}}$

On obtient $\mathbb{E}\left[\|\ell - \hat{\ell}_{D_{opt}}\|_2^2\right] \leq C D_{opt}^{-2\beta} + \frac{1}{m}(n+\sigma^2) D_{opt}$

$$= O\left(m^{-\frac{2\beta}{2\beta+1}}\right)$$

10) Pour déduire de $\hat{\ell}_D$ un estimateur de $m = \frac{\ell}{f}$, il faut estimer f la densité inconnue des (X_i) : $\rightsquigarrow$ projection par exemple $\hat{f}_D$

ou $\rightsquigarrow$ estimateur à noyaux ...

puis $\hat{m}_D := \frac{\hat{\ell}_D}{\hat{f}_D}$