Exercise 1 (Refresher on matrices)

1. Let $\mathbf{A}$ be a $n \times d$ matrix with real entries. Show that $\text{range}(\mathbf{A}) = \text{range}(\mathbf{A}\mathbf{A}^\top)$.

$n \updownarrow \overset{d}{\overleftrightarrow{A}}$? $\text{range}(A) = \text{range}(AA^T)$.

$\text{Ker}(AA^T) \overset{?}{=} \text{Ker}(A^T)$.

si $x \in \text{Ker}(A^T)$. alors: $A \underbrace{A^T x}_{=0} = 0$ donc $x \in \text{Ker}(AA^T)$.

Récip: si $x \in \text{Ker}(AA^T)$, alors $x^T \underbrace{AA^T x}_{0} = 0 = \|[A^T x]\|^2$

$\Rightarrow A^T x = 0 \Rightarrow x \in \text{Ker}\, A^T$.

Donc: $\text{Ker}(AA^T) = \text{Ker}(A^T)$.

$\boxed{\text{Rq: } [\text{Ker}(A^T)]^\perp = \text{Range}(A).}$ En effet: $\text{Range}(A)^\perp = \text{Ker}(A^T)$.

$A = [A_1, \dots, A_d]$

$x \in \text{Range}(A)^\perp \Leftrightarrow \begin{bmatrix} A_1^T x \\ A_2^T x \\ \vdots \\ A_d^T x \end{bmatrix} = 0 = A^T x$

$\Leftrightarrow x \in \text{Ker}(A^T)$.

Ainsi: $\text{Range}(\underline{A}) = \{\text{Ker}(\underline{A^T})\}^\perp = \{\text{Ker}(\underline{AA^T})\}^\perp = \text{Range}(\underline{AA^T})$.

2. Let $\{U_k\}_{1 \le k \le r}$ be a family of r orthonormal vectors of $\mathbb{R}^n$. Show that $\sum_{k=1}^r U_k U_k^\top$ is the matrix associated with the orthogonal projection onto $\mathbf{H} = \{\sum_{k=1}^r \alpha_k U_k \,;\, \alpha_1, \dots, \alpha_r \in \mathbb{R}\}$. Deduce that if $\mathbf{A}$ is a $n \times d$ matrix with real entries such that each column of $\mathbf{A}$ is in $\mathbf{H}$, then,

$$\left(\sum_{k=1}^r U_k U_k^\top\right) \mathbf{A} = \mathbf{A}\,.$$

$P_H^\perp(X) = \sum_{k=1}^{n} \underbrace{\langle X, u_k \rangle}_{u_k^T X} u_k = \underbrace{\left(\sum_{k=1}^{n} u_k u_k^T\right)}\, X$.

$A = [A_1, \dots, A_d]$.

$\left(\sum_{k=1}^{d} u_k u_k^T\right)\left([A_1, \dots, A_d]\right) = \left(\underbrace{\sum_{k=1}^{d} u_k u_k^T A_1}_{A_1}, \dots, \underbrace{\sum_{k=1}^{d} u_k u_k^T A_d}_{A_d}\right)$

$\boxed{\sum_{k=1}^{d} u_k u_k^T A = A.}$

3. Let $p < d$ and $\mathbf{B} \in \mathbb{R}^{d\times p}$ such that $\mathbf{B}^\top \mathbf{B} = I_p$. Let us denote $\mathbf{B} = (b_{ij})_{\substack{1\le i\le d\\1\le j\le p}}$ the components of $\mathbf{B}$ and for all $i \in [\![1,d]\!]$, $\alpha_i = \sum_{j=1}^p b_{ij}^2$. Show that $\sum_{i=1}^d \alpha_i = p$ and $\alpha_i \le 1$.

$$B \in M_{d,p} \iff B \in \mathbb{R}^{d\times p}. \qquad B = [B_1, \dots, B_p] \updownarrow d, \quad B^T B = I_p.$$

$$\tilde{B} = [B_{p+1}, \dots, B_d]$$

$$\text{tq}: [B, \tilde{B}] = \bar{B} : \text{BON de } \mathbb{R}^d!$$

$$\sum_{i=1}^d \alpha_i = \sum_{i=1}^d \sum_{j=1}^p b_{ij}^2 = \operatorname{tr}(\underbrace{B^T B}_{I_p}) = p.$$

$$= \sum_{j=1}^p \Big(\underbrace{\sum_{i=1}^d b_{ij}^2}_{\|B_j\|^2 = 1} \Big) = p. \qquad B = (B_1, \dots, B_p)$$

$$\alpha_i = \sum_{j=1}^p b_{ij}^2 \le \sum_{j=1}^d b_{ij}^2 = 1. \qquad \bar{B}\bar{B}^T = I \iff \bar{B}^T = \bar{B}^{-1} \iff \bar{B}^T \bar{B} = I.$$

Exercise 2 (Principal Component Analysis) Principal component analysis is a multivariate technique which aims at analyzing the statistical structure of high dimensional dependent observations by representing data using orthogonal variables called *principal components*. Reducing the dimensionality of the data is motivated by several practical reasons such as improving computational complexity. Let $(X_i)_{1\leqslant i\leqslant n}$ be i.i.d. random variables in $\mathbb{R}^d$ and consider the matrix $\mathbf{X}\in\mathbb{R}^{n\times d}$ such that the i-th row of $\mathbf{X}$ is the observation $X_i^\top$. In this exercise, it is assumed that data are preprocessed so that the columns of $\mathbf{X}$ are centered. This means that for all $1\leqslant k\leqslant d$, $\sum_{i=1}^n X_{i,k}=0$. Let $\mathbf{\Sigma}_n$ be the empirical covariance matrix:

$$\mathbf{\Sigma}_n = n^{-1}\sum_{i=1}^n X_i X_i^\top .$$

Principal Component Analysis aims at reducing the dimensionality of the observations $(X_i)_{1\leqslant i\leqslant n}$ using a *compression* matrix $\mathbf{U}\in\mathbb{R}^{d\times p}$ with orthonormal columns with $p\leqslant d$ so that for each $1\leqslant i\leqslant n$, $\mathbf{U}^\top X_i$ ia a low dimensional representation of X_i. The original observation may then be partially recovered using $\mathbf{U}$. Principal Component Analysis computes $\mathbf{U}$ using the least squares approach:

$$\mathbf{U}_\star \in \underset{\substack{U\in\mathbb{R}^{d\times p}\\ U^\top U=I_p}}{\operatorname{argmin}} \sum_{i=1}^n \|X_i - \underbrace{\mathbf{U}\mathbf{U}^\top}_{P_H^{\perp}} X_i\|^2 ,$$

$U = [u_1, \dots, u_p]$

H : Range(U)

1. Prove that for all matrix $\mathbf{A}\in\mathbb{R}^{n\times d}$ with rank r, there exist $\sigma_1\geqslant\ldots\geqslant\sigma_r>0$ such that

$$\mathbf{A} = \sum_{k=1}^r \sigma_k u_k v_k^\top ,$$

$v_k \in \mathbb{R}^d$. $u_k \in \mathbb{R}^n$.

where $\{u_1,\ldots,u_r\}\subset\mathbb{R}^n$ and $\{v_1,\ldots,v_r\}\subset\mathbb{R}^d$ are two families of orthonormal vectors. The vectors $\{u_1,\ldots,u_r\}$ (resp. $\{v_1,\ldots,v_r\}$) are the left-singular (resp. right-singular) vectors associated with $\{\sigma_1,\ldots,\sigma_r\}$, the singular values of $\mathbf{A}$. *If $\mathbf{U}$ denotes the $\mathbb{R}^{n\times r}$ matrix with columns given by $\{u_1,\ldots,u_r\}$ and $\mathbf{V}$ denotes the $\mathbb{R}^{d\times r}$ matrix with columns given by $\{v_1,\ldots,v_r\}$, then the singular value decomposition of $\mathbf{A}$ may also be written as*

$$\mathbf{A} = \mathbf{U}\mathbf{D}_r\mathbf{V}^\top ,$$

where $\mathbf{D}_r = \mathrm{diag}(\sigma_1,\ldots,\sigma_r)$. Then, $\mathbf{A}^\top\mathbf{A}$ and $\mathbf{A}\mathbf{A}^\top$ are positive semidefinite such that

$$\mathbf{A}^\top\mathbf{A} = \mathbf{V}\mathbf{D}_r^2\mathbf{V}^\top \quad\text{and}\quad \mathbf{A}\mathbf{A}^\top = \mathbf{U}\mathbf{D}_r^2\mathbf{U}^\top .$$

In the framwework of this exercise, $n\mathbf{\Sigma}_n = \mathbf{X}^\top\mathbf{X}$ so that diagonalizing $n\mathbf{\Sigma}_n$ is equivalent to computing the singular value decomposition of $\mathbf{X}$.

$$A = \sum_{k=1}^r \sigma_k u_k v_k^\top = \underbrace{\underbrace{[u_1,\dots,u_r]}_{U}\underbrace{\begin{pmatrix}\sigma_1 & & 0\\ & \ddots & \\ 0 & & \sigma_r\end{pmatrix}}_{D_r}}\underbrace{\begin{pmatrix}v_1^T\\ \vdots \\ v_r^T\end{pmatrix}}_{V^T}.$$

$$= [\sigma_1 u_1, \sigma_2 u_2, \dots, \sigma_r u_r] \begin{pmatrix} v_1^T \\ \vdots \\ v_r^T\end{pmatrix} = \sum_{i=1}^r \sigma_i u_i v_i^T$$

Soit $A \in M_{n,d}(\mathbb{R})$, AA^T est symétrique, réel, ⊕ (car : $u^T AA^T u = \|A^T u\|^2 \geq 0$)

$$AA^T = U \underbrace{\begin{pmatrix}\lambda_1 & & 0\\ & \ddots & \\ 0 & & \lambda_n\end{pmatrix}} U^T = \sum_{i=1}^n \lambda_i u_i u_i^T$$

où : $U = (u_1, \dots, u_n)$.

$\lambda_1 \geq \dots \geq \lambda_n$.

$(u_i)_{1\leq i\leq n}$ BON de vect propres de AA^T.

or : rang$(A) = r =$ rang(AA^T). donc : $\lambda_1 \geq \dots \geq \lambda_r > \lambda_{r+1} = 0 = \dots = \lambda_n$.

$AA^T = \sum_{i=1}^n \lambda_i u_i u_i^T$, $\underline{AA^T u_i = \lambda_i u_i}$, $\forall i \in \{1, .., n\}$.

$$A^T A \, \overline{A^T u_i} = \lambda_i \, \overline{A^T u_i} \qquad v_i = \gamma_i A^T u_i$$

En posant $\boxed{v_i = \frac{A^T u_i}{\sqrt{\lambda_i}}}$ ✓

$$v_i^T v_i = \gamma_i^2 \, u_i^T \underbrace{A A^T u_i}_{\lambda_i u_i} = \gamma_i^2 \lambda_i = 1$$

$$\gamma_i = 1/\sqrt{\lambda_i}$$

On a : $\boxed{A^T A v_i = \lambda_i v_i}$, v_i BON de vect. propres associé à (λ_i) pour $A^T A$.

Posons : $\boxed{\sigma_i = \sqrt{\lambda_i}}$ ✓

$$\sum_{i=1}^{r} \sigma_i u_i v_i^T = \sum_{i=1}^{r} \sqrt{\lambda_i} \, u_i \left(\frac{A^T u_i}{\sqrt{\lambda_i}}\right)^T = \Big(\underbrace{\sum_{i=1}^{r} u_i u_i^T}_{P^{\perp}_{\mathrm{Range}(AA^T)} \,=\, \mathrm{Range}(A)}\Big) A = A.$$

$$P^{\perp}_{\mathrm{Im}(A)}(A) = A.$$

2. Prove that solving the PCA least squares optimization problem boils down to computing

$$\mathbf{U}_\star \in \underset{\mathbf{U}\in\mathbb{R}^{d\times p},\, \mathbf{U}^\top\mathbf{U}=\mathbf{I}_p}{\operatorname{argmax}} \{\mathrm{Trace}(\mathbf{U}^\top \mathbf{\Sigma}_n \mathbf{U})\}.$$

$$U_\star \in \underset{\substack{U\in\mathbb{R}^{d\times p}\\ U^TU=I_p}}{\operatorname{argmin}} \sum_{i=1}^{n} \underbrace{\| X_i - UU^T X_i \|^2}_{\|X_i\|^2 - 2\underbrace{\langle X_i, UU^TX_i\rangle}_{X_i^T UU^T X_i} + \underbrace{X_i^T U \overbrace{U^T U}^{I_p} U^T X_i}_{X_i^T U U^T X_i}}$$

$$= \underset{\substack{U^TU=I_p\\ U\in\mathbb{R}^{d\times p}}}{\operatorname{argmax}} \sum_{i=1}^{n} \underbrace{X_i^T U U^T X_i}_{= \mathrm{tr}(X_i^T U U^T X_i) \,=\, \mathrm{tr}(U^T X_i X_i^T U)}$$

$$= \underset{\substack{U^TU=I_p\\ U\in\mathbb{R}^{d\times p}}}{\operatorname{argmax}} \mathrm{tr}\Big(U^T \big(\underbrace{\sum_{i=1}^{n} X_i X_i^T}_{n\Sigma_n}\big) U\Big) = \underset{\substack{U\in\mathbb{R}^{d\times p}\\ U^TU=I_p}}{\operatorname{argmax}} \mathrm{tr}(U^T \Sigma_n U).$$

3. Let $\lambda_1 \geqslant \ldots \geqslant \lambda_d$ be real numbers and denote $f : \alpha \in \mathbb{R}^d \mapsto \sum_{i=1}^d \alpha_i \lambda_i$. Show that

$$\sup \left\{ f(\alpha) : \alpha \in [0,1]^d, \sum_{i=1}^d \alpha_i = p \right\}$$

is attained for $\alpha^\star = (\mathbb{1}_{i \le p})_{1 \le i \le d}$.

4. Let $\{\vartheta_1, \ldots, \vartheta_d\}$ be orthonormal eigenvectors associated with the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_d$ of $\mathbf{\Sigma}_n$. Prove that a solution to the PCA least squares optimization problem is given by the matrix $\mathbf{U}_\star$ with columns $\underbrace{\{\vartheta_1, \ldots, \vartheta_p\}}$.

3). $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_d$.

$$\sum_{i=1}^d \alpha_i \lambda_i \leqslant \sum_{i=1}^p \alpha_i \lambda_i + \lambda_p \underbrace{\Big(\sum_{i=p+1} \alpha_i\Big)}_{p - \sum_{i=1}^p \alpha_i \,=\, \sum_{i=1}^p (1-\alpha_i)} = \sum_{i=1}^p \underbrace{\Big(\alpha_i \lambda_i + \underbrace{\lambda_p}_{\leqslant \lambda_i} (1-\alpha_i)\Big)}_{\leqslant \lambda_i} \leqslant \sum_{i=1}^p \lambda_i$$

4) Mq : $\underset{(v_1, \ldots, v_p)}{U^* \,\Vert} \in \underset{\substack{U \in \mathbb{R}^{d \times p} \\ U^T U = I_p}}{\mathrm{Argmax}} \; \mathrm{tr}(U^T \Sigma_n U)$.

où : v_i : vect propre de Σ_n associé à λ_i (avec $\lambda_1 \geqslant \cdots \geqslant \lambda_d$).

$$\Sigma_n = \underbrace{V D_n V^T}_{} = \sum_{i=1}^d \lambda_i V_i V_i^T$$

où : $D_n = \mathrm{diag}(\lambda_1, \ldots, \lambda_d)$.
$V = (\underbrace{V_1}, \ldots, \underbrace{V_d})$ où : $(V_i)_{1 \leqslant i \leqslant d}$: BON de vecteurs propres de Σ_n.

$$\mathrm{tr}\,(\underbrace{U^T \Sigma_n U}_{\in \mathbb{R}^{p \times p}}) = \mathrm{tr}\,(\underbrace{U^T V}_{B^T} \underbrace{D_n}_{D_n} \underbrace{V^T U}_{B}) = \sum_{i=1}^p \sum_{k=1}^d \underbrace{(B^T)_{ik}}_{B_{ki}} \lambda_k \qquad B_{ki}$$

$\overset{p}{\updownarrow} U \overset{d}{\updownarrow}$ $\quad p \updownarrow U^T$

$$= \sum_{i=1}^p \underbrace{\sum_{k=1}^d B_{ki}^2}_{} \lambda_k$$

$$\underbrace{B^T B} = U^T \underbrace{V V^T}_{I_d} U = U^T U = I_p$$

$$= \sum_{k=1}^d \underbrace{\sum_{i=1}^p B_{ki}^2}_{\alpha_k} \lambda_k.$$

Par Ex 1, Q3. $\sum_{i=1}^d \alpha_i = p$, $\alpha_i \in [0,1]$.

Par Ex 2, $\underline{Q3}$. $\sum_{k=1}^d \alpha_k \lambda_k \leqslant \underbrace{\sum_{k=1}^p \lambda_k} = \sum_{k=1}^d \alpha_k^* \lambda_k$ où $\alpha_k^* = \begin{cases} 1 & \text{si } k \in [\![1,p]\!] \\ 0 & \text{sinon.} \end{cases}$

En prenant $U_* = (V_1, \ldots, V_p)$. $\qquad V = (U_*, \tilde{V})$ où $\tilde{V} = (V_{p+1}, \ldots, V_d)$.

$$\mathrm{tr}(U_*^T V D_n V^T U_*) = \mathrm{tr}\Big(\underbrace{U_*^T (U_*, \tilde{V})} D_n \begin{pmatrix} U_*^T \\ \tilde{V} \end{pmatrix} U_* \Big)$$

$$= \mathrm{tr}\left((I_p, 0) D_n \begin{pmatrix} I_p \\ 0 \end{pmatrix} \right) = \underbrace{\sum_{i=1}^p \lambda_i}$$

5. For any dimension $1 \leqslant p \leqslant d$, let $\mathcal{F}_d^p$ be the set of all vector subpaces of $\mathbb{R}^d$ with dimension p. Consider the linear span V_p defined as

$$V_p \in \underset{V\in\mathcal{F}_d^p}{\operatorname{argmin}} \sum_{i=1}^n \|X_i - \pi_V(X_i)\|^2 ,$$

where π_V is the orthogonal projection onto the linear span V. Prove that $V_1 = \operatorname{span}\{v_1\}$ where

$$v_1 \in \underset{v\in\mathbb{R}^d\,;\,\|v\|=1}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 .$$

6. For all $2 \leqslant p \leqslant d$, following the same steps, prove that a solution to the optimization problem is given by $V_p = \operatorname{span}\{v_1,\dots,v_p\}$ where

$v^T\Sigma_n v.$

$$v_1 \in \underset{v\in\mathbb{R}^d\,;\,\|v\|=1}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 \quad\text{and for all } 2\leqslant k\leqslant p\,,\quad v_k \in \underset{\substack{v\in\mathbb{R}^d\,;\,\|v\|=1\,;\\ v\perp v_1,\dots,v\perp v_{k-1}}}{\operatorname{argmax}} \sum_{i=1}^n \langle X_i, v\rangle^2 . \qquad (1)$$

7. Prove that the vectors $\{v_1,\dots,v_k\}$ defined by (1) can be chosen as the orthonormal eigenvectors associated with the k largest eigenvalues of the empirical covariance matrix $\boldsymbol{\Sigma}_n$.

(5,6,7). $(v_1,\dots,v_p)$ p vect propres associés à $\lambda_1,\dots,\lambda_p$. et à la matrice Σ_n.

$(v_1,\dots,v_d)$: BON de vect propres de Σ_n.

$v=\sum_{i=1}^d \beta_i v_i$

$$\Rightarrow \quad v^T\Sigma_n v = v^T\Big[\sum_{i=1}^d \beta_i \underbrace{\Sigma_n v_i}_{\lambda_i v_i}\Big] = \Big(\sum_{h=1}^d \beta_h v_h\Big)^T\Big(\sum_{i=1}^d \beta_i \lambda_i v_i\Big).$$

$$= \sum_{i=1}^d \beta_k^2 \lambda_i \leq \lambda_1 \underbrace{\sum_{k=1}^d \beta_k^2}_{\|v\|^2=1} = \lambda_1 .$$

or: $v_1^T\Sigma_n v_1 = \lambda_1 v_1^T v_1 = \lambda_1$

Donc: $\boxed{v_1 \in \underset{\substack{v\in\mathbb{R}^d\\ \|v\|=1}}{\operatorname{Argmax}}\ v^T\Sigma_n v .}$

$\forall v \in \operatorname{Span}(v_1,\dots,v_{\ell-1})^\perp$

$v = \sum_{i=\ell}^d \beta_i v_i$

$$v^T\Sigma_n v = \sum_{i=\ell}^d \beta_i^2 \lambda_i \leq \lambda_\ell \Big(\underbrace{\sum_{i=\ell}^d \beta_i^2}_{=1}\Big) = \lambda_\ell .$$

$$v_\ell^T\Sigma_n v_\ell = \lambda_\ell v_\ell^T v_\ell = \lambda_\ell .$$

Donc: $\boxed{v_\ell \in \underset{v\in \operatorname{Span}(v_1,\dots,v_{\ell-1})^\perp}{\arg\max}\ v^T\Sigma_n v .}$

8. The orthonormal eigenvectors associated with the eigenvalues of Σ_n allow to define the principal components as follows. Then, as $V_d = \mathrm{span}\{\underbrace{\vartheta_1, \dots, \vartheta_d}\}$, for all $1 \leqslant i \leqslant n$,

BON.

$$\pi_{V_d}(X_i) = \sum_{k=1}^{d} \langle X_i, \vartheta_k \rangle \vartheta_k = \sum_{k=1}^{d} (X_i^\top \vartheta_k)\vartheta_k = \sum_{k=1}^{d} c_k(i)\vartheta_k \,,$$

where for all $1 \leqslant k \leqslant d$, the k-th principal component is defined as $c_k = \mathbf{X}\vartheta_k$. Prove that $(c_1, \dots, c_d)$ are orthogonal vectors.

$$\pi_{V_d}(X_i) = P^{\perp}_{V_p}(X_i) = \sum_{k=1}^{d} \underbrace{\underbrace{\langle X_i, v_k \rangle}_{X_i^T v_k}}_{c_k(i)} v_k .$$

$$c_k(i) = X_i^T v_k \,. \quad \forall k \in \{1, \dots, d\}$$

$$c_k = \begin{pmatrix} c_k(1) \\ \vdots \\ c_k(n) \end{pmatrix} = \underbrace{\begin{pmatrix} X_1^T \\ \vdots \\ X_n^T \end{pmatrix}}_{X} v_k .$$

$$\langle c_k, c_l \rangle = 0 = v_k^T \underbrace{\overbrace{X^T X}^{n \Sigma_n} v_l}_{n \lambda_l v_l} = n \lambda_l \underbrace{v_k^T v_l}_{=0 \text{ si } k \neq l} .$$

Exercise 3 (Kernel Principal Component Analysis) Let $(X_i)_{1\leq i\leq n}$ be n observations in a general space $\mathcal{X}$, $k : \mathcal{X}\times\mathcal{X}\to\mathbb{R}$ a positive kernel and $\mathbf{K} = (k(X_i, X_j))_{1\leqslant i,j\leqslant n}$. $\mathcal{W}$ denotes the Reproducing Kernel Hilbert Space associated with k and for all $x\in\mathcal{X}$, $\phi(x)$ denotes the function $\phi(x) : y \to k(x, y)$. The aim is now to perform a PCA on $(\phi(X_1), \ldots, \phi(X_n))$. It is assumed that $\sum_{i=1}^{n}\phi(X_i) = 0$.

1. Prove that

$$f_1 = \underset{f\in\mathcal{W}\,;\,\|f\|_{\mathcal{W}}=1}{\operatorname{argmax}} \sum_{i=1}^{n}\langle\phi(X_i), f\rangle_{\mathcal{W}}^2$$

may be written

$$f_1 = \sum_{i=1}^{n}\alpha_1(i)\phi(X_i)\;, \quad \text{where} \quad \alpha_1 = \underset{\alpha\in\mathbb{R}^n\,;\,\alpha^\top\mathbf{K}\alpha=1}{\operatorname{argmax}} \alpha^\top\mathbf{K}^2\alpha\;.$$

2. Prove that $\alpha_1 = \lambda_1^{-1/2} b_1$ where b_1 is the unit eigenvector associated with the largest eigenvalue λ_1 of $\mathbf{K}$. What about $(f_2, \ldots, f_p)$ defined iteratively as in (1)?

3. Write $H_p = \mathrm{span}\{f_1, \ldots, f_p\}$. Prove that, for all $1 \leqslant i \leqslant n$,

$$\pi_{H_p}(\phi(X_i)) = \sum_{j=1}^{p} \lambda_j \alpha_j(i) f_j \ .$$