Warm-up

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

   First note that $AA^\top x = 0$ implies $< A^\top x, A^\top x >= 0$ so that $A^\top x = 0$. The converse is obvious. Therefore, $\mathrm{Ker}(AA^\top) = \mathrm{Ker}(A^\top)$. And using that $\mathrm{Ker}(B^\top) = (\mathrm{Im}(B))^\perp$, we deduce that $\mathrm{Im}(AA^\top)^\perp = \mathrm{Im}(A)^\perp$, which concludes the proof.

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

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

   Let $\pi_H(X)$ be the orthogonal projection of X onto H. Since $\{U_k\}_{1\leq k\leq r}$ is an orthonormal basis of H,

$$\pi_H(X) = \sum_{k=1}^r < X, U_k > U_k = \left(\sum_{k=1}^r U_k U_k^\top\right) X\,.$$

   This implies that for each $X \in H$, $X = \left(\sum_{k=1}^r U_k U_k^\top\right) X$. Since all the column vectors of A are in H, this yields $\left(\sum_{k=1}^r U_k U_k^\top\right) A = A$.

Kernel Principal Component Analysis

Principal Component Analysis

Let $(X_i)_{1\leqslant i\leqslant n}$ be i.i.d. random variables in $\mathbb{R}^d$ and consider the matrix $X \in \mathbb{R}^{n\times d}$ such that the i-th row of X is the observation $X_i^\top$. In this exercise, it is assumed that data are preprocessed so that the columns of X are centered. This means that for all $1 \leqslant k \leqslant d$, $\sum_{i=1}^n X_{i,k} = 0$. Let Σ_n be the empirical covariance matrix:

$$\Sigma_n = n^{-1}\sum_{i=1}^n X_i X_i^\top\,.$$

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

$$U_\star \in \underset{U\in\mathbb{R}^{d\times p}}{\mathrm{argmin}} \sum_{i=1}^n \|X_i - UU^\top X_i\|_2^2\,,$$

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

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

where $\{u_1,\dots,u_r\}\subset\mathbb{R}^n$ and $\{v_1,\dots,v_r\}\subset\mathbb{R}^d$ are two families of orthonormal vectors. The vectors $\{u_1,\dots,u_r\}$ (resp. $\{v_1,\dots,v_r\}$) are the left-singular (resp. right-singular) vectors associated with $\{\sigma_1,\dots,\sigma_r\}$, the singular values of A.

Since the matrix $AA^\top$ is positive semidefinite, its spectral decomposition is given by

$$AA^\top=\sum_{k=1}^r \lambda_k u_k u_k^\top\,,$$

where $\lambda_1\geqslant\ldots\geqslant\lambda_r>0$ are the nonzero eigenvalues of $AA^\top$ and $\{u_1,\dots,u_r\}$ is an orthonormal family of $\mathbb{R}^n$. For all $1\leqslant k\leqslant r$, define $v_k=\lambda_k^{-1/2}A^\top u_k$ so that

$$\begin{aligned}&\|v_k\|^2=\lambda_k^{-1}\langle A^\top u_k;A^\top u_k\rangle=\lambda_k^{-1}u_k^\top AA^\top u_k=1\,,\\&A^\top A v_k=\lambda_k^{-1/2}A^\top AA^\top u_k=\lambda_k v_k\,.\end{aligned}$$

On the other hand, for all $1\leqslant k\neq j\leqslant r$, $\langle v_k;v_j\rangle=\lambda_k^{-1/2}\lambda_j^{-1/2}u_k^\top AA^\top u_j=\lambda_k^{-1/2}\lambda_j^{1/2}u_k' u_j=0$. Therefore, $\{v_1,\dots,v_r\}$ is an orthonormal family of eigenvectors of $A^\top A$ associated with the eigenvalues $\lambda_1\geqslant\ldots\geqslant\lambda_r>0$. Define, for all $1\leqslant k\leqslant r$, $\sigma_k=\lambda_k^{1/2}$ which yields

$$\sum_{k=1}^r\sigma_k u_k v_k^\top=\sum_{k=1}^r u_k u_k^\top A=\left(\sum_{k=1}^r u_k u_k^\top\right)A\,.$$

As $\{u_1,\dots,u_r\}$ is an orthonormal family, $UU^\top=\sum_{k=1}^r u_k u_k^\top$ is the orthogonal projection onto the $\mathrm{range}(AA^\top)=\mathrm{range}(A)$ *which implies*

$$\sum_{k=1}^r\sigma_k u_k v_k^\top=\left(\sum_{k=1}^r u_k u_k^\top\right)A=A\,.$$

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

$$U_\star\in\underset{U\in\mathbb{R}^{d\times p},\,U^\top U=\mathrm{I}_p}{\mathrm{argmax}}\ \{\mathrm{trace}(U^\top\Sigma_n U)\}\,.$$

Let $U\in\mathbb{R}^{d\times p}$ be such that $U^\top U=\mathrm{I}_p$. Then,

$$\begin{aligned}\sum_{i=1}^n\|X_i-UU^\top X_i\|_2^2&=\sum_{i=1}^n\|X_i\|_2^2+\sum_{i=1}^n\|UU^\top X_i\|_2^2-2\sum_{i=1}^n\langle X_i;UU^\top X_i\rangle\,,\\&=\sum_{i=1}^n\|X_i\|_2^2+\sum_{i=1}^n X_i^\top UU^\top X_i-2\sum_{i=1}^n X_i^\top UU^\top X_i\,,\\&=\sum_{i=1}^n\|X_i\|_2^2-\sum_{i=1}^n X_i^\top UU^\top X_i\,,\\&=\sum_{i=1}^n\|X_i\|_2^2-\mathrm{trace}(U^\top XX^\top U)\,.\end{aligned}$$

3. Let $\{\vartheta_1,\dots,\vartheta_d\}$ be orthonormal eigenvectors associated with the eigenvalues $\lambda_1\geqslant\ldots\geqslant\lambda_d$ of Σ_n. Prove that a solution to this problem is given by the matrix $U_\star$ with columns $\{\vartheta_1,\dots,\vartheta_p\}$.

Let $\Sigma_n=VD_nV^\top$ be the spectral decomposition of Σ_n where $D_n=\mathrm{Diag}(\lambda_1,\dots,\lambda_d)$ and $V\in\mathbb{R}^{d\times d}$ is a matrix with orthonormal columns $\{\vartheta_1,\dots,\vartheta_d\}$. For all $U\in\mathbb{R}^{d\times p}$ matrix with orthonormal columns define $B=V^\top U$ so that, as $V\in\mathbb{R}^{d\times d}$ is an orthogonal matrix,

$$VB=VV^\top U=U\quad\text{and}\quad U^\top\Sigma_n U=B^\top V^\top VD_nV^\top VB=B^\top D_nB\,.$$

Therefore,

$$\mathrm{Trace}(U^\top \Sigma_n U) = \mathrm{Trace}(B^\top D_n B) = \sum_{i=1}^{d} \lambda_i \sum_{j=1}^{p} b_{i,j}^2\,. \tag{1}$$

On the other hand,

$$B^\top B = U^\top V V^\top U = U^\top U = \mathrm{I}_p\,,$$

so that the columns of B are orthonormal and

$$\sum_{i=1}^{d}\sum_{j=1}^{p} b_{i,j}^2 = p\,.$$

Hence, introducing for all $1 \leqslant i \leqslant d$, $\alpha_i = \sum_{j=1}^p b_{i,j}^2$, by (1),

$$\mathrm{Trace}(U^\top \Sigma_n U) = \sum_{i=1}^{d} \alpha_i \lambda_i\,,$$

with, for all $1 \leqslant i \leqslant d$, $\alpha_i \in [0,1]$ and $\sum_{i=1}^d \alpha_i = p$. As $\lambda_1 \geqslant \lambda_2 \geqslant \ldots, \lambda_d$,

$$\mathrm{Trace}(U^\top \Sigma_n U) \leqslant \sum_{i=1}^{p} \lambda_i\,.$$

Indeed, the function $f_d : (\alpha_1, \ldots, \alpha_d) \mapsto \sum_{i=1}^d \alpha_i \lambda_i$ is maximized under the constraints $\alpha_i \in [0,1]$ and $\sum_{i=1}^d \alpha_i = p$ by $(\alpha_i^*)_{1 \leqslant i \leqslant d}$ such that $\alpha_1^* = \ldots = \alpha_p^* = 1$. Assume that $(\alpha_1, \ldots, \alpha_d)$ is such that there exists $1 \leqslant j_0 \leqslant p$ such that $\alpha_{j_0} < 1$. Then, $\sum_{j=p+1}^d \alpha_j \geqslant 1 - \alpha_{j_0}$ and we may write, as $\lambda_{j_0} \geqslant \lambda_{p+1} \geqslant \ldots \geqslant \lambda_d$,

$$f_d : (\alpha_1, \ldots, \alpha_d) \leqslant \sum_{i=1, i \neq j_0}^{p} \alpha_i \lambda_i + \lambda_{j_0} + \sum_{i=p+1}^{d} \tilde{\alpha}_i \lambda_i\,,$$

where $(\tilde{\alpha}_i)_{p+1 \leqslant i \leqslant d}$ are in $[0,1]$ and such that $\sum_{i=1, i\neq j_0}^p \alpha_i + 1 + \sum_{i=p+1}^d \tilde{\alpha}_i = p$.
As the columns of $U_\star$ are $\{\vartheta_1, \ldots, \vartheta_p\}$, for all $1 \leqslant i \leqslant d$ and $1 \leqslant j \leqslant p$, $b_{i,j} = \langle \vartheta_i; \vartheta_j \rangle = \delta_{i,j}$. Therefore, for all $1 \leqslant i \leqslant d$, $\sum_{j=1}^p b_{i,j}^2 = 1$ and

$$\mathrm{Trace}(U_\star^\top \Sigma_n U_\star) = \sum_{i=1}^{p} \lambda_i\,,$$

which completes the proof.

4. 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_d defined as

$$V_p \in \underset{V \in \mathcal{F}_d^p}{\mathrm{argmin}} \sum_{i=1}^{n} \|X_i - \pi_V(X_i)\|_2^2\,,$$

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

$$v_1 \in \underset{v \in \mathbb{R}^d\,;\, \|v\|_2 = 1}{\mathrm{argmax}} \sum_{i=1}^{n} \langle X_i, v \rangle^2\,.$$

Write $V_1 = \mathrm{span}\{v_1\}$ for $v_1 \in \mathbb{R}^d$ such that $\|v_1\|_2 = 1$. Then,

$$
\begin{aligned}
\sum_{i=1}^n \|X_i - \pi_{V_1}(X_i)\|_2^2 &= \sum_{i=1}^n \|X_i - \langle X_i; v_1\rangle v_1\|_2^2\,,\\
&= \sum_{i=1}^n \left(\|X_i\|_2^2 - 2\langle X_i; \langle X_i; v_1\rangle v_1\rangle + \|\langle X_i; v_1\rangle v_1\|_2^2\right)\,,\\
&= \sum_{i=1}^n \left(\|X_i\|_2^2 - \langle X_i; v_1\rangle^2\right).
\end{aligned}
$$

Consequently, V_1 is a solution if and only if v_1 is solution to:

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

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

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

Write $V_p = \mathrm{span}\{v_1, \ldots, v_p\}$ where $\{v_1, \ldots, v_p\}$ is an orthonormal family. Then,

$$
\sum_{i=1}^n \|X_i - \pi_{V_p}(X_i)\|^2 = \sum_{i=1}^n \|X_i - \sum_{k=1}^p \langle X_i; v_k\rangle v_k\|_2^2 = \sum_{i=1}^n \left(\|X_i\|_2^2 - \sum_{k=1}^p \langle X_i; v_k\rangle^2\right)\,.
$$

$(v_1, \ldots, v_p)$ is therefore solution to

$$
v = (v_1, \ldots, v_p) \in \operatorname{argmax} \sum_{k=1}^p \sum_{i=1}^n \langle X_i; v_k\rangle^2\,.
$$

The additive form of the function to be maximized allows to build the orthonormal basis of V_p sequentially as claimed.

6. Prove that the vectors $\{v_1, \ldots, v_k\}$ defined by (2) can be chosen as the orthonormal eigenvectors associated with the k largest eigenvalues of the empirical covariance matrix Σ_n.

Note that for all $v \in \mathbb{R}^d$ such that $\|v\|_2 = 1$,

$$
\frac{1}{n}\sum_{i=1}^n \langle X_i, v\rangle^2 = \frac{1}{n}\sum_{i=1}^n (v^\top X_i)(X_i^\top v) = v^\top \Sigma_n v\,.
$$

As $(\vartheta_i)_{1\leqslant i\leqslant d}$ are the orthonormal eigenvectors associated with the eigenvalues $\lambda_1 \geqslant \ldots \geqslant \lambda_d \geqslant 0$ of Σ_n. Then,

$$
\frac{1}{n}\sum_{i=1}^n \langle X_i, v\rangle^2 = v^\top\left(\sum_{i=1}^d \lambda_i \vartheta_i \vartheta_i^\top\right) v = \sum_{i=1}^d \lambda_i \langle v, \vartheta_i\rangle^2 \leqslant \lambda_1 \sum_{i=1}^d \langle v, \vartheta_i\rangle^2
$$

and, as $(\vartheta_i)_{1\leqslant i\leqslant d}$ is an orthonormal basis of $\mathbb{R}^d$, $\sum_{i=1}^d \langle v, \vartheta_i\rangle^2 = \|v\|_2^2 = 1$. Therefore,

$$
\frac{1}{n}\sum_{i=1}^n \langle X_i, v\rangle^2 \leqslant \lambda_1\,.
$$

On the other hand, for all $2 \leqslant i \leqslant d$, $\langle \vartheta_1, \vartheta_i \rangle = 0$ *and* $\langle \vartheta_1, \vartheta_1 \rangle = 1$ *so that* $\sum_{i=1}^d \lambda_i \langle \vartheta_1, \vartheta_i \rangle^2 = \lambda_1$ *which proves that* ϑ_1 *is solution to* (2).

Assume now that $v \in \mathbb{R}^d$ *is such that* $\|v\| = 1$ *and for all* $1 \leqslant j \leqslant k-1$, $\langle v; \vartheta_j \rangle = 0$ *and write*

$$\frac{1}{n}\sum_{i=1}^n \langle X_i, v\rangle^2 = \sum_{i=1}^d \lambda_i \langle v, \vartheta_i\rangle^2 \leq \lambda_k \sum_{i=k}^d \langle v, \vartheta_i\rangle^2 \leq \lambda_k \,,$$

since, as $(\vartheta_i)_{1\leqslant i\leqslant d}$ *is an orthonormal basis of* $\mathbb{R}^d$, $\sum_{i=1}^d \langle v, \vartheta_i\rangle^2 = \sum_{i=k}^d \langle v, \vartheta_i\rangle^2 = \|v\|^2 = 1$. *On the other hand, for all* $1 \leqslant i \leqslant d$, $i \neq k$, $\langle \vartheta_k, \vartheta_i \rangle = 0$ *and* $\langle \vartheta_k, \vartheta_k \rangle = 1$ *so that* $\sum_{i=1}^d \lambda_i \langle \vartheta_k, \vartheta_i \rangle^2 = \lambda_k$ *which proves that* ϑ_k *is solution to* (2).

Therefore, $V_p = \mathrm{span}\{\vartheta_1, \ldots \vartheta_p\}$ *is a solution to* (2) *and, as* $(\vartheta_i)_{1\leqslant i\leqslant p}$ *is an orthonormal family, the projection matrix onto* V_p *is given by* $U_\star U_\star^\top$ *where* $U_\star$ *is a* $\mathbb{R}^{d\times p}$ *matrix with columns* $\{\vartheta_1, \ldots \vartheta_p\}$.

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

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

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

The k*-th principal component is the vector whose components are the coordinates of each* X_i, $1 \leqslant i \leqslant n$, *relative to the basis* $\{\vartheta_1, \ldots, \vartheta_d\}$ *of* V_d. *For all* $1 \leqslant i \neq j \leqslant d$,

$$\langle c_i, c_j\rangle = \vartheta_i^\top X^\top X \vartheta_j = \vartheta_i^\top (n\Sigma_n)\vartheta_j = n\lambda_j \vartheta_i^\top \vartheta_j = 0 \,,$$

as $\{\vartheta_1, \ldots, \vartheta_d\}$ *is an orthonormal family.*

Application to RKHS

Let $(X_i)_{1\leq i\leq n}$ be n observations in a general space $\mathcal{X}$ and $k : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ a positive function. We assume that k is symmetric and that for all $n \geqslant 1$, $(a_i)_{1\leqslant i\leqslant n} \in \mathbb{R}^n$ and $(x_i)_{1\leqslant i\leqslant n} \in \mathcal{X}^n$, $\sum_{1\leqslant i,j\leqslant n} a_i a_j k(x_i, x_j) \geqslant 0$. For all $x \in \mathcal{X}$, $\phi(x)$ denotes the function $\phi(x) : y \to k(x, y)$.

Let $\mathcal{W}$ be a Hilbert space of real-valued functions defined on $\mathcal{X}$, andowed with an inner product denoted by $\langle\cdot,\cdot\rangle_{\mathcal{W}}$, and such that for all $x \in \mathcal{X}$, $\phi(x) \in \mathcal{W}$ and for all $f \in \mathcal{W}$ and all $x \in \mathcal{X}$, $f(x) = \langle f, \phi(x)\rangle_{\mathcal{W}}$. The aim is now to perform a PCA on $(\phi(X_1), \ldots, \phi(X_n))$. It is assumed that $\sum_{i=1}^n \phi(X_i) = 0$. Define

$$K = (k(X_i, X_j))_{1\leqslant i,j\leqslant n} \,.$$

1. Prove that

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

may be written

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

Any solution to the optimization problem lies in the vectorial subspace $V = \mathrm{span}\{\phi(X_i), \ldots, \phi(X_n)\}$. *Let* $f = \sum_{i=1}^n \alpha(i)\phi(X_i)$ *be such that* $\|f\|_{\mathcal{W}} = 1$. *Then,*

$$\|f\|_{\mathcal{W}}^2 = \sum_{i,j=1}^n \alpha_i\alpha_j \langle\phi(X_i), \phi(X_j)\rangle_{\mathcal{W}} = \alpha^\top K\alpha \,.$$

On the other hand, $\langle \phi(X_i), f\rangle_{\mathcal{W}} = f(X_i) = [K\alpha](i)$ so that,

$$\sum_{i=1}^{n} \langle \phi(X_i), f\rangle_{\mathcal{W}}^2 = \sum_{i=1}^{n} f^2(X_i) = \sum_{i=1}^{n} ([K\alpha](i))^2 = (K\alpha_1)^\top K\alpha_1 = \alpha^\top K^2 \alpha \,.$$

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 K.

   Let $\lambda_1 \geqslant \ldots \geqslant \lambda_n \geq 0$ be the eigenvalues of K associated with the orthonormal basis of eigenvectors $(b_1, \ldots, b_n)$. For any $\alpha \in \mathbb{R}^n$ such that $\alpha^\top K \alpha = 1$,

$$\alpha^\top K^2 \alpha = \alpha^\top \left(\sum_{i=1}^{n} \lambda_i b_i b_i^\top \right)^2 \alpha = \sum_{i=1}^{n} \lambda_i^2 \langle \alpha, b_i\rangle^2 \leqslant \lambda_1 \sum_{i=1}^{n} \lambda_i \langle \alpha, b_i\rangle^2 = \lambda_1 \,,$$

   as $\alpha^\top K\alpha = \sum_{i=1}^{n} \lambda_i \langle \alpha, b_i\rangle^2 = 1$. On the other hand,

$$\left(\lambda_1^{-1/2} b_1\right)^T K^2 \left(\lambda_1^{-1/2} b_1\right) = \lambda_1^{-1} \sum_{i=1}^{n} \lambda_i^2 \langle b_1, b_i\rangle^2 = \lambda_1 \,.$$

   Following the same steps, f_j may be written $f_j = \sum_{i=1}^{n} \alpha_j(i)\phi(x_i)$ with $\alpha_j = \lambda_j^{-1/2} b_j$.

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

$$\pi_{H_d}(\phi(X_i)) = \sum_{j=1}^{d} \lambda_j \alpha_j(i) f_j \,.$$

   Note first that the $(f_1, \ldots, f_d)$ is an orthonormal family. Therefore,

$$\pi_{H_d}(\phi(X_i)) = \sum_{j=1}^{d} \langle \phi(X_i), f_j\rangle_{\mathcal{W}} f_j = \sum_{j=1}^{d} \langle \phi(X_i), \sum_{\ell=1}^{n} \alpha_j(\ell)\phi(X_\ell)\rangle_{\mathcal{W}} f_j = \sum_{j=1}^{d} [K\alpha_j](i) f_j \,.$$

   Therefore,

$$\pi_{H_d}(\phi(x_i)) = \sum_{j=1}^{d} \lambda_j^{-1/2} [Kb_j](i) f_j = \sum_{j=1}^{d} \lambda_j^{1/2} b_j(i) f_j = \sum_{j=1}^{d} \lambda_j \alpha_j(i) f_j \,.$$