

Mémoire

Présenté et soutenu par
Ellyott Lankar, Esther Hammer, Eva Zagury

MASTER 1 : MATHÉMATIQUES APPLIQUÉES

ANALYSE EN COMPOSANTE PRINCIPALE POUR DONNÉES FONCTIONNELLES

Tutrice : Angéline Roche
roche@ceremade.dauphine.fr

5 mai 2021

Remerciements

Nos premiers remerciements s'adressent bien évidemment à notre encadrante **Angelina Roche**, pour nous avoir donné l'opportunité de traiter un sujet de mémoire aussi intéressant qu'actuel. Outre la grande disponibilité tout au long de l'année, elle nous a fourni un encadrement de qualité, alliant rigueur et pédagogie dans la transmission de son savoir scientifique.

Nous remercions également notre professeur **Julien Stoehr** pour nous avoir attribué ce sujet qui était parmi nos favoris, ainsi que pour l'organisation de l'ensemble des projets.

Nous tenons enfin à remercier le corps professoral de l'université Paris-Dauphine pour nous avoir apporté un bagage nécessaire, nous permettant l'accomplissement de ce mémoire.

Table des matières

Remerciements	3
Introduction	7
1 Principe de l'ACP	9
1.1 Cas vectoriel	9
1.2 Cas fonctionnel	10
1.2.1 Données fonctionnelles	10
1.2.2 ACP fonctionnelle	10
2 Outils d'analyse	11
2.1 Sur l'intégrale de Bochner	11
2.2 Notions d'espérance et de covariance	12
2.2.1 Espérance	12
2.2.2 Opérateur de covariance	13
2.3 Bases Hilbertiennes	14
2.3.1 Représentation des fonctions par une base fonctionnelle	14
2.3.2 Bases de Fourier	15
2.3.3 Bases Splines	15
2.3.4 Quelques autres exemples de bases	15
3 Approximation par un opérateur de rang fini	17
3.1 Détermination des composantes principales	18
3.1.1 Diagonalisation de l'opérateur de covariance	18
3.1.2 Détermination de l'opérateur de rang fini minimisant la variance	19
3.2 Estimation des fonctions propres	20
3.2.1 Opérateur de covariance empirique	20
3.2.2 Reconstruction des fonctions propres	21
3.3 Représentation des variables	22
4 Application numérique	23
4.1 Méthode d'implémentation	23
4.2 Application sur données simulées	24
4.3 Application sur données réelles	26
Conclusion	33
A Théorèmes généraux	35

B Codes R	37
Bibliographie	39

Introduction

À l'instant où le lecteur lit ces quelques lignes, environ 9000 GB de données sont produites en France, selon le rapport DOMO 2020. Et ce chiffre augmente d'années en années (Forbes 2019, How Much Data Is Collected Every Minute Of The Day).

Depuis quelques années, on assiste à une explosion des quantités de données collectées, liée aux progrès des capacités de stockage de celles-ci. C'est l'ère du Big Data.

Bonne nouvelle ? Pas si simple. Cet important volume a en effet des conséquences qui vont contre l'intuition : au fur et à mesure que la dimension augmente, la quantité de données dont nous avons besoin pour une analyse précise croît exponentiellement. On parle de "fléau de la dimension".

Notamment lorsque la donnée prend une structure particulière, et varie continûment en fonction de variables. Les données collectées appartiennent dans ce cas à un certain espace fonctionnel (par exemple $L^2([0, 1])$) de dimension infinie.

Pour lutter contre ce problème, il est utile de chercher à résumer l'information : l'idée est donc de représenter les données dans un espace de plus petite dimension. Cependant, cette projection entraîne inévitablement une perte d'information, et l'enjeu est alors de trouver (s'il existe) l'espace qui minimisera au mieux cette perte. C'est le principe même de l'Analyse en Composantes Principales (ACP).

Chapitre 1

Principe de l'ACP

L'idée de cette section est de se familiariser avec le concept d'analyse en composantes principales en se plaçant en premier lieu dans un cadre plus simple : le cas vectoriel.

1.1 Cas vectoriel

Soit $M = (X_1, \dots, X_n) \in \mathbb{R}^{n \times d}$ un nuage de points, noté M , i.i.d suivant la même loi qu'une v.a.r X . On suppose ici d grand. L'ACP est une technique ayant pour objectif de rechercher pour un entier $k \leq d$ fixé le ou les sous-espaces affines de dimension k , passant par le centre de gravité du nuage M expliquant au mieux celui-ci. En d'autres termes cela consiste à trouver les vecteurs orthonormaux ψ_1, \dots, ψ_k (appelés **composantes principales**) minimisant le critère :

$$\min_{\psi_1, \dots, \psi_k} \mathbb{E} \left\| X - \sum_{i=1}^k \psi_i \psi_i^t X \right\|^2 \quad (1.1)$$

On appellera dorénavant **inertie du nuage M par rapport au sous-espace engendré par ψ_1, \dots, ψ_k** la fonction à minimiser.

Les résultats qui suivent sont tirés de The Oxford Handbook, chap 8 [1]. Le problème 1.1 revient à trouver successivement les axes ψ_i tel que la projection du nuage sur $Vect(\psi_1, \dots, \psi_k)$ ait une variance maximale.

On appelle K la matrice de covariance de X . Par le théorème spectral, K est diagonalisable dans une base orthonormée, notons ϕ_1, \dots, ϕ_n ses vecteurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. Ainsi,

$$K = \sum_{i=1}^n \lambda_i \phi_i \phi_i^t$$

On remarque que $\lambda_1 = Var(\phi_1^t X)$, de tel sorte que le vecteur unitaire qui maximise la variance est un vecteur propre de K associé à la valeur propre λ_1 .

On continue ensuite la recherche du deuxième axe de projection sur le même principe, en imposant qu'il soit orthogonal à ϕ_1 . On peut ainsi déduire que **les k premières composantes principales sont k vecteurs propres K , associés respectivement aux valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_k$** .

1.2 Cas fonctionnel

1.2.1 Données fonctionnelles

Dans notre mémoire, on se penche sur la question de l'analyse des données collectées et stockées qui sont représentées sous forme de fonction (courbes, images, ...). Il s'agit d'observations de valeurs discrètes qui reflètent une variation homogène. Dans toute cette section, on considère un espace de probabilité (Ω, F, \mathbb{P}) . On note $(X_i)_{1 \leq i \leq n}$, les données, à valeurs dans un espace de fonctions, avec n la taille de l'échantillon.

Définition 1.1. *Soit I un ensemble, un processus stochastique réel est une famille $X = (X_t, t \in I)$ de variables aléatoires à valeurs dans \mathbb{R} .*

On peut ainsi voir le processus X à valeurs dans \mathbb{R} , un espace muni de sa tribu borélienne $B(\mathbb{R})$, comme une famille (pas forcément dénombrable) de v.a. indexée par le temps :

$$\begin{aligned} X_t : (\Omega, F, \mathbb{P}) &\longrightarrow (\mathbb{R}, B(\mathbb{R})) \\ \omega &\longmapsto X(t, \omega) \end{aligned}$$

Une observation d'une telle variable fonctionnelle est appelée donnée fonctionnelle.

On peut par ailleurs voir le processus X comme une fonction réelle paramétrée par un état du monde.

Définition 1.2. *Pour tout $\omega \in \Omega$, on appelle trajectoire de X (associée à l'état ω) la fonction :*

$$\begin{aligned} X : I &\longrightarrow \mathbb{R} \\ t &\longmapsto X(t, \omega) \end{aligned}$$

Dans le cadre de notre mémoire, on observe $X_i(t_j)$, avec $i \in [1, n]$ et $j \in [1, p]$ tels que X_1, \dots, X_n sont i.i.d suivant la même loi qu'un processus X . On supposera que $I = [0, 1]$ et que les trajectoires de X vivent dans l'espace $L^2([0, 1]) \cap C([0, 1])$.

1.2.2 ACP fonctionnelle

La variable X vivant cette fois dans un espace de dimension infinie, la question de réduction de dimension se pose d'autant plus. On va chercher à adapter l'ACP classique aux cas fonctionnel.

L'objectif est toujours le même ; on cherche à minimiser l'inertie par rapport à ψ_1, \dots, ψ_k , famille orthonormale de $L^2([0, 1]) \cap C([0, 1])$ cette fois.

$$\min_{\psi_1, \dots, \psi_k} \mathbb{E} \left\| X - \sum_{i=1}^k \psi_i \psi_i^t X \right\|^2 \quad (1.2)$$

Nous verrons par la suite que malgré un problème de nature bien plus complexe, le résultat sera similaire au cas vectoriel : les k premières composantes principales sont k vecteurs propres de l'opérateur de covariance Γ de X , (éléments de $L^2[0, 1] \cap C[0, 1]$, on les appelle donc **fonctions propres** de Γ).

Chapitre 2

Outils d'analyse

Nos observations $X_i(t_j)$ sont i.i.d suivant la même loi que X , processus à trajectoires continues et de carré intégrable. Cette variable aléatoire fonctionnelle étant donc à valeurs dans un espace de dimension infinie, l'intégrale de Lebesgue n'est pas définie et nous aurons besoin d'un outil plus général.

2.1 Sur l'intégrale de Bochner

L'intégrale de Bochner (du nom de son créateur Salomon Bochner) étend la définition de l'intégrale de Lebesgue aux fonctions à valeurs dans un espace de Banach. Dans cette section, on fixe H un espace de Banach que l'on munit des boréliens ainsi qu'un espace mesuré $(\Omega, \mathcal{F}, \mu)$. Les preuves des résultats énoncés ne seront pas exposées (Voir Dronio, chapitres 1.1 et 1.2 [??]).

On définit une fonction simple comme combinaison linéaire de fonctions caractéristiques d'ensembles partitionnant Ω , i.e de la forme

$$\sum_{i=1}^N a_i \mathbb{1}_{A_i}$$

avec $a_i \in H$, $\mu(A_i) < \infty$ et de telle sorte que $(A_i)_{i=1, \dots, N}$ forme une partition de Ω . Cette fonction est alors mesurable comme combinaison linéaire de fonctions mesurables.

Pour s fonction simple, on définit son intégrale par

$$\int_{\Omega} s \, d\mu := \sum_{i=1}^N a_i \mu(A_i)$$

élément de H avec la convention $\mu(A) \times 0 = 0$.

Cette formule définit bien l'intégrale de s de manière intrinsèque. On remarque de plus que c'est une forme linéaire sur les fonctions simples et qui vérifie l'inégalité triangulaire, c'est-à-dire :

$$\left\| \int_{\Omega} s \, d\mu \right\| \leq \int_{\Omega} \|s\| \, d\mu.$$

Les résultats suivants vont nous servir à définir une notion de mesurabilité stable par passage à la limite, afin de définir l'intégrale d'une fonction en tant qu'approximation

par des fonctions simples. Par ailleurs, sous des hypothèses raisonnables, cette notion est équivalente à la notion classique de mesurabilité.

Définition 2.1. Soit $f : \Omega \rightarrow H$. On dit que f est μ -mesurable, s'il existe une suite de fonctions simples qui converge μ -presque partout vers f .

On note que f ainsi définie est mesurable en tant que limite simple de fonctions mesurables sur Ω à valeurs dans H . Les notions sont même équivalentes dans un cadre particulier.

Théorème 2.2. Si Ω est de mesure σ -finie et H est séparable, alors $f : \Omega \rightarrow H$ est μ -mesurable si et seulement si f est mesurable.

On peut maintenant définir la notion de fonction (Bochner) intégrable.

Définition 2.3. Une fonction $f : \Omega \rightarrow H$ est dite (Bochner-) intégrable s'il existe une suite de fonctions simples $(s_n)_{n \geq 0}$ (dite suite approximante) telle que :

- $s_n \rightarrow f$ μ -presque partout (i.e f est μ -mesurable et $\int_{\Omega} \|f\| \, d\mu < \infty$);
- $\int_{\Omega} \|s_n - f\| \, d\mu \rightarrow 0$

Théorème 2.4. f est intégrable si et seulement si f est μ -mesurable et $\int_{\Omega} \|f\| \, d\mu < \infty$.

Définition 2.5. Si $f : \Omega \rightarrow H$ est intégrable, on définit l'intégrale de f comme suit

$$\int_{\Omega} f \, d\mu := \lim_{n \rightarrow \infty} \int_{\Omega} s_n \, d\mu$$

où $(s_n)_{n \geq 1}$ est une suite approximante pour f .

On peut vérifier à nouveau que cette intégrale est linéaire, et que les théorèmes classiques en théorie de Lebesgue (Convergence dominée, Fubini) sont toujours vrais pour cette "extension".

Revenons maintenant à notre problème. L'espace $L^2([0, 1])$ est un Banach (c'est même un espace de Hilbert) séparable, et donc ici la notion de \mathbb{P} -mesurabilité est équivalente à celle de mesurabilité. On peut alors définir l'intégrale de Bochner de X (qui est mesurable) par rapport à la mesure \mathbb{P} .

2.2 Notions d'espérance et de covariance

Nous avons maintenant les outils nécessaires pour définir des notions essentielles pour la suite de ce mémoire. Les définitions et caractérisations suivantes sont basées essentiellement sur la thèse d'Angelina Roche (chap. 1.1 [3]).

2.2.1 Espérance

Définition 2.6. Si $\mathbb{E}[\|X\|] < \infty$, on peut définir l'espérance de X comme :

$$\mathbb{E}[X] := \int_{\Omega} X(\omega) \, d\mathbb{P}(\omega)$$

où l'intégrale est l'intégrale de Bochner de X comme détaillé précédemment.

On définit ainsi, un élément de $L^2([0, 1]) \cap C([0, 1])$. En effet, X étant supposée continue sur $[0, 1]$, $\mathbb{E}[X]$ l'est aussi par théorème de convergence dominée. Par ailleurs, pour toute fonction simple s , on a :

$$\|\mathbb{E}[s]\|_2 = \left\| \sum_{i=1}^N a_i \mu(A_i) \right\| \leq \sum_{i=1}^N \|a_i\| \mu(A_i) = \mathbb{E}[\|s\|_2].$$

Cette dernière intégrale est une intégrale au sens classique de Lebesgue de la fonction mesurable positive $\|s\| : \Omega \rightarrow \mathbb{R}$. Puis on conclut par passage à la limite pour toute fonction Bochner-intégrable. $\mathbb{E}[X]$ est donc bien de carré intégrable

Une autre possibilité pour définir l'espérance d'une variable fonctionnelle consiste à prendre l'espérance de X point par point.

Définition 2.7. On peut définir l'espérance de X comme suit

$$\begin{aligned} \mathbb{E}[X] : [0, 1] &\longrightarrow \mathbb{R} \\ t &\longmapsto \mathbb{E}[X](t) = \int_{\Omega} X(t, \omega) \, d\mathbb{P}(\omega) \end{aligned}$$

où l'intégrale est l'intégrale classique (de Lebesgue) par rapport à la mesure \mathbb{P} .

Ici, ces deux notions d'espérance coïncident bien sur $L^2([0, 1])$, étant donné que l'on suppose $\mathbb{E}[\|X\|] < \infty$.

On suppose dès à présent X centré, i.e $t \mapsto \mathbb{E}[X(t)]$ est la fonction nulle.

2.2.2 Opérateur de covariance

Dans le cadre vectoriel, si $X = \begin{pmatrix} X_1 \\ \vdots \\ X_d \end{pmatrix}$ variable aléatoire sur \mathbb{R}^d , on appelle matrice de variance-covariance la matrice $\Sigma = (Cov(X_i, X_j))_{1 \leq i, j \leq d}$. L'idée ici est de généraliser la notion de matrice de covariance dans le cadre fonctionnel.

Définition 2.8. On appelle fonction de covariance la fonction :

$$\begin{aligned} K : [0, 1]^2 &\longrightarrow \mathbb{R} \\ (s, t) &\longmapsto \mathbb{E}[X(s)X(t)] \end{aligned}$$

Notons que pour s donné, la variable $X(s)$ est une v.a.r, et ainsi l'espérance est celle définie à partir de l'intégrale de Lebesgue. Notons de plus que K est continue.

Définition 2.9. On appelle opérateur de covariance l'opérateur de noyau K : c'est à dire :

$$\begin{aligned} \Gamma : L^2([0, 1]) &\longrightarrow L^2([0, 1]) \\ f &\longmapsto \Gamma f(s) = \int_0^1 K(s, t) f(t) \, dt \end{aligned}$$

L'opérateur de covariance a une importance particulière dans l'ACP comme nous le verrons par la suite.

Proposition 2.10. Soit $(e_n)_{n \in \mathbb{N}}$ une base hilbertienne de $L^2([0, 1])$ donnée, alors Γ est entièrement caractérisé par la donnée de :

$$\langle \Gamma e_p, e_q \rangle_{p, q \geq 1} = \text{Cov}(\langle X, e_p \rangle, \langle X, e_q \rangle)$$

Démonstration. Soit $f \in L^2([0, 1])$. On écrit $f = \sum_{k \geq 0} \langle f, e_k \rangle e_k$. L'opérateur de covariance pris en f se décompose alors de la façon suivante :

$$\begin{aligned} \Gamma(f)(s) &= \int_0^1 \mathbb{E}[X(t)X(s)]f(t)dt \\ &= \mathbb{E}\left[\int_0^1 X(s)X(t)f(t)dt\right] \\ &= \mathbb{E}[\langle X, f \rangle X(s)] \\ &= \mathbb{E}\left[\sum_{k \geq 0} \langle f, e_k \rangle \langle X, e_k \rangle X(s)\right] \\ &= \sum_{k \geq 0} \langle f, e_k \rangle \mathbb{E}[\langle X, e_k \rangle \sum_{j \geq 0} \langle X, e_j \rangle e_j(s)] \\ &= \sum_{k \geq 0} \sum_{j \geq 0} \langle f, e_k \rangle e_j(s) \mathbb{E}[\langle X, e_k \rangle \langle X, e_j \rangle] \\ &= \sum_{k, j \geq 0} \langle f, e_k \rangle e_j(s) \langle \Gamma e_k, e_j \rangle \end{aligned}$$

Où l'on a utilisé le fait que pour $p, q \geq 1$, on a :

$$\begin{aligned} \langle \Gamma e_p, e_q \rangle &= \int_0^1 \int_0^1 (K(t, u) e_p(u) du) e_q(t) dt \\ &= \int_0^1 \int_0^1 K(t, u) e_p(u) e_q(t) du dt \\ &= \int_0^1 \int_0^1 \mathbb{E}[X(t)X(u)] e_p(u) e_q(t) du dt \\ &= \mathbb{E}\left[\int_0^1 \int_0^1 X(t)X(u) e_p(u) e_q(t) du dt\right] \end{aligned} \tag{2.1}$$

$$= \mathbb{E}[\langle X, e_p \rangle \langle X, e_q \rangle] \tag{2.2}$$

$$= \text{Cov}(\langle X, e_p \rangle, \langle X, e_q \rangle) \tag{2.3}$$

□

Une conséquence immédiate de cette proposition est que $\langle \Gamma f, f \rangle = \text{Var} \langle X, f \rangle \geq 0$, et donc Γ est positif.

2.3 Bases Hilbertiennes

2.3.1 Représentation des fonctions par une base fonctionnelle

La caractérisation de l'opérateur de covariance donnée plus haut laisse à penser que l'on nécessitera d'une base hilbertienne pour représenter l'opérateur de covariance.

On rappellera la méthode des bases fonctionnelles pour effectuer une analyse fonctionnelle et étant dans un espace de Hilbert, on rappellera les énoncés suivants :

Définition 2.11. On dit que $(\varphi_i)_{i \in I}$ est une base de Hilbertienne de H si :

1. $(\varphi_i)_{i \in I}$ est une famille orthonormale de H
2. $\forall x \in H, \exists (\lambda_i)_{i \in I}$ tel que $\sum_{i \in I} \lambda_i \varphi_i = x$

Théorème 2.12. Tout espace préhilbertien séparable possède une base Hilbertienne.

Une base d'une fonction permet d'approcher la fonction par un ensemble de fonctions connues φ_k (indépendantes), en réalisant une combinaison linéaire avec K suffisamment grand. Idéalement, la base fonctionnelle doit avoir les mêmes caractéristiques que la fonction estimée. Ainsi, plusieurs facteurs sont à prendre en compte dont la dimension K de la base, les coefficients a_k et le choix de la base.

Soit x la fonction, on note \hat{x} la fonction estimée, on a :

$$\hat{x}(t) = \sum_{k \in K} a_k \varphi_k(t)$$

Non seulement, il faut choisir la base par rapport à la fonction à estimer mais également selon les variables à étudier. Une base qui a une bonne représentation de la fonction, ne donne pas forcément une bonne représentation des dérivées des variables. On va, par la suite, présenter quelques exemples de bases très souvent utilisées en pratique.

2.3.2 Bases de Fourier

La base de Fourier est définie par

$$\forall t \in [0, 1], \forall n \in \mathbb{N}, \varphi_0(t) = 1, \varphi_{2n-1}(t) = \sqrt{2} \sin n\pi t, \varphi_{2n}(t) = \sqrt{2} \cos n\pi t. \quad (2.4)$$

Étant une base de période $\frac{2\pi}{w}$, elle est utilisée pour les fonctions au comportement périodique. Particulièrement efficace et simple à manipuler pour les fonctions très régulières, elle est inappropriée pour des fonctions qui ont des discontinuités ou qui sont dérivables seulement pour un ordre n petit.

2.3.3 Bases Splines

Une fonction spline est une fonction définie par morceaux par des polynômes tel que, aux noeuds (point de liaison entre les différents polynômes) la fonction est régulière.

Définition 2.13. Soit K un ensemble, une base de spline $(\varphi_k)_{k \in K}$ est définie telle que,

1. $\forall k, \varphi_k$ est une fonction spline
2. $\forall k, \varphi_k$ sont indépendantes

Les fonctions splines sont utilisées pour la reconstruction de fonctions non-périodiques, et la base de spline la plus connue est la B-spline.

2.3.4 Quelques autres exemples de bases

Bases exponentielles

La base exponentielle est constituée de fonction exponentielle de paramètres λ_k ,

Bases d'ondelettes

À l'inverse des bases de Fourier, les bases d'ondelettes sont compatibles avec des fonctions discontinues. Elles combinent les avantages des bases de Fourier et des bases splines. La base d'ondelettes la plus connue est la base de Haar définies de la manière suivante,

$$\varphi_{jk}(t) = 2^{\frac{j}{2}} \varphi(2^j t - k), \forall t \in \mathbb{R}.$$

Chapitre 3

Approximation par un opérateur de rang fini

Revenons au problème de départ. L'enjeu de l'ACP consiste à trouver les fonctions ψ_k (normées et orthogonales) minimisant le critère suivant :

$$\min_{\psi_1, \dots, \psi_k} \mathbb{E} \left\| X - \sum_{i=1}^k \psi_i \int_0^1 \psi_i(t) X(t) dt \right\|_{L^2}^2$$

Réécrivons la fonction à minimiser. Pour $(\psi_i)_{i \in \mathbb{N}}$ famille orthonormée de $L^2([0, 1])$, on a :

$$\begin{aligned} \left\| X - \sum_{i=1}^k \psi_i \langle \psi_i, X \rangle \right\|_{L^2}^2 &= \|X\|_{L^2}^2 - 2 \langle X, \sum_{i=1}^k \psi_i \langle \psi_i, X \rangle \rangle + \left\| \sum_{i=1}^k \psi_i \langle \psi_i, X \rangle \right\|_{L^2}^2 \\ &= \|X\|_{L^2}^2 - 2 \sum_{i=1}^k \langle X, \psi_i \rangle^2 + \langle \sum_{i=1}^k \psi_i \langle \psi_i, X \rangle, \sum_{j=1}^k \psi_j \langle \psi_j, X \rangle \rangle \\ &= \|X\|_{L^2}^2 - 2 \sum_{i=1}^k \langle X, \psi_i \rangle^2 + \sum_{i,j=1}^k \langle \psi_i, X \rangle \langle \psi_j, X \rangle \langle \psi_i, \psi_j \rangle \\ &= \|X\|_{L^2}^2 - 2 \sum_{i=1}^k \langle X, \psi_i \rangle^2 + \sum_{i=1}^k \langle \psi_i, X \rangle^2 \quad \text{par orthonormalité de } (\psi_i)_i \\ &= \|X\|_{L^2}^2 - \sum_{i=1}^k \langle X, \psi_i \rangle^2 \end{aligned}$$

En passant à l'espérance, on obtient finalement :

$$\begin{aligned} \mathbb{E} \left\| X - \sum_{i=1}^k \psi_i \int_0^1 \psi_i(t) X(t) dt \right\|_{L^2}^2 &= \mathbb{E} [\|X\|_{L^2}^2 - \sum_{i=1}^k \langle X, \psi_i \rangle^2] \\ &= \mathbb{E} [\|X\|_{L^2}^2] - \sum_{i=1}^k \int_0^1 \int_0^1 K(t, u) \psi_i(u) \psi_i(t) du dt \end{aligned}$$

$$= \mathbb{E}[\|X\|_{L^2}^2] - \sum_{i=1}^k \langle \Gamma \psi_i, \psi_i \rangle$$

En ayant repris les lignes de calculs (2.1) à (2.3) pour la ligne 3.

Minimiser l'inertie revient donc à maximiser la somme de droite sous la contrainte de d'orthonormalité des ψ_i . L'objectif de la section suivante sera de montrer l'existence d'une solution à ce problème.

3.1 Détermination des composantes principales

3.1.1 Diagonalisation de l'opérateur de covariance

Nous allons voir que la structure particulière de l'opérateur de covariance, et de l'espace dans lequel vit X permet l'existence de fonctions propres pour cet opérateur. On va d'abord énoncer des résultats plus généraux. Notons H un espace de Hilbert séparable (i.e contenant un sous-ensemble dense dénombrable). Par ailleurs, on note $\mathbb{L}(H) := \{\text{ensemble des opérateurs de } H \rightarrow H\}$.

Définition 3.1. $T \in \mathbb{L}(H)$ est auto-adjoint si

$$\forall x, y \in H, \langle Tx, y \rangle = \langle x, Ty \rangle$$

Remarque. Dans le cas où H est de dimension finie, cette définition correspond à celle d'un endomorphisme symétrique.

Définition 3.2. $T \in \mathbb{L}(H)$ est dit de Hilbert-Schmidt s'il existe $(e_n)_{n \in \mathbb{N}}$ une base hilbertienne telle que :

$$\sum_{i \in \mathbb{N}} \|Te_n\|^2 < \infty.$$

Théorème 3.3. Soit $T \in \mathbb{L}(H)$ un opérateur de Hilbert-Schmidt auto-adjoint, alors il existe une base hilbertienne $(\phi_n)_n$ de H formée de vecteurs propres de T .

Ainsi, en notant pour $n \in \mathbb{N}$, λ_n la valeur propre associée à $(\phi_n)_{n \in \mathbb{N}}$, on a,

$$\langle T\phi_p, \phi_q \rangle = \langle \lambda_p \phi_p, \phi_q \rangle = \lambda_p \langle \phi_p, \phi_q \rangle = \lambda_p \delta_{p,q}.$$

De plus $\lambda_j = \langle \Gamma \phi_j, \phi_j \rangle = \text{Var}(\langle X, \phi_j \rangle) \geq 0$.

Remarque. 1. Si $\lambda_i > 0$, alors les ϕ_i ainsi définis sont éléments de $\mathfrak{S}(\Gamma) \subset L^2([0, 1]) \cap C([0, 1])$.

2. Si $\lambda_i > 0$, alors ϕ_i est point fixe de l'opérateur $f \mapsto \frac{1}{\lambda_i} \Gamma(f)$

Pour une démonstration du théorème, le lecteur pourra se référer aux notes de cours M. L. Gallardo, partie 3 [4]

Appliquons le théorème à l'opérateur de covariance Γ .
 Γ est un opérateur auto-adjoint. En effet, pour $f, g \in L^2[0, 1]$, on a

$$\langle \Gamma f, g \rangle = \left\langle \int_0^1 K(\cdot, t) f(t) dt, G \right\rangle$$

$$\begin{aligned}
&= \int_0^1 \int_0^1 K(s, t) f(t) dt g(s) ds \\
&= \int_0^1 \int_0^1 K(t, s) f(t) g(s) ds dt \\
&= \langle f, \Gamma g \rangle
\end{aligned}$$

Montrons maintenant que Γ est de Hilbert-Schmidt.

Proposition 3.4. *Si $K \in L^2([0, 1])$ et est symétrique alors $\Gamma : f \mapsto \int_0^1 K(\cdot, t) f(t) dt$ est de Hilbert-Schmidt.*

Démonstration. Soit $(\varphi_n)_{n \in \mathbb{N}}$ une base hilbertienne de $L^2([0, 1])$.

$$\begin{aligned}
\sum_{n \in \mathbb{N}} \|T\varphi_n\|^2 &= \sum_{n \in \mathbb{N}} \int_0^1 \left(\int_0^1 (K(s, t) \varphi_n(t) dt) \right)^2 ds \\
&= \int_0^1 \sum_{n \in \mathbb{N}} \left(\int_0^1 (K(s, t) \varphi_n(t) dt) \right)^2 ds \\
&= \int_0^1 \|K(\cdot, t)\|^2 ds \text{ d'après le théorème A.1} \\
&= \int_0^1 \left(\int_0^1 (K(s, t)^2 dt) \right) ds \\
&= \|K\|_{L^2([0, 1]^2)}^2 < \infty
\end{aligned}$$

□

Dans notre cas, K est symétrique, et donc Γ est bien de Hilbert-Schmidt. On peut ainsi appliquer de théorème 3.3 à cet opérateur.

3.1.2 Détermination de l'opérateur de rang fini minimisant la variance

On note $(\lambda_j)_{j \geq 1}$ les valeurs propres de Γ associées à $(\phi_j)_{j \geq 1}$. Puisque Γ est positif, il vient que $\lambda_j \geq 0 \forall j \geq 1$. On va supposer $\lambda_j > 0$. On remarque d'une part que $\sum_{i=1}^k \langle \Gamma \phi_i, \phi_i \rangle = \sum_{i=1}^k \lambda_i$. Montrons d'autres parts que :

$$\max_{\psi_1, \dots, \psi_k} \sum_{i=1}^k \langle \Gamma \psi_i, \psi_i \rangle = \sum_{i=1}^k \lambda_i$$

On aura alors trouvé une famille de fonctions de $L^2([0, 1]) \cap C([0, 1])$ solution du problème. Au vu de la forme du problème, on peut utiliser le principe de programmation dynamique pour sa résolution (voir Cardaliaguet p.45 [5]). On a alors l'équation suivante :

$$\begin{cases} V(j) = V(j-1) + \max_{\psi_j \perp (\psi_1, \dots, \psi_{j-1})} \langle \Gamma \psi_j, \psi_j \rangle & j \in \{1, \dots, k\} \\ V(0) = 0 \end{cases} \quad (3.1)$$

$$\text{où } V(j) := \max_{\psi_1, \dots, \psi_j} \sum_{i=1}^j < \Gamma \psi_i, \psi_i >.$$

Résolution de 3.1 pour $j = 1$:

En indiquant les valeurs propres et les fonctions propres telles que $\lambda_1 \geq \lambda_2 \geq \dots > 0$, on obtient par le théorème de Mercer A.2 une écriture du noyau :

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_i \phi_i(s) \phi_i(t)$$

En l'injectant dans l'expression à maximiser, on trouve pour $\psi \in L^2([0, 1])$ normé :

$$\begin{aligned} < \Gamma \psi, \psi > &= < < K(\cdot, \cdot), \psi >, \psi > \\ &= < < \sum_{i=1}^{\infty} \lambda_i \phi_i(\cdot) \phi_i(\cdot), \psi >, \psi > \\ &= \sum_{i=1}^{\infty} \lambda_i < \phi_i(\cdot) < \phi_i(\cdot), \psi >, \psi > \\ &= \sum_{i=1}^{\infty} \lambda_i < \phi_i, \psi >^2 \\ &\leq \lambda_1 \sum_{i=1}^{\infty} < \phi_i, \psi >^2 \\ &\leq \lambda_1 \|\psi\|^2 \quad \text{d'après l'inégalité de Bessel} \\ &= \lambda_1 \quad \text{car } \psi \text{ est normé} \end{aligned}$$

En remarquant que $< \Gamma \phi_1, \phi_1 > = \lambda_1$, on obtient qu'une fonction qui atteint le maximum est la fonction propre normalisée ϕ_1 associée à λ_1 .

On applique alors le principe de programmation dynamique en résolvant 3.1 pour tout j de la même manière que pour le cas $j = 1$, et on trouve directement que la base de l'ACP est la famille $(\phi_j)_{j \in [1, k]}$.

3.2 Estimation des fonctions propres

En pratique, le calcul de l'opérateur de covariance est difficile voire impossible car il nécessite de connaître la loi jointe de $(X(s), X(t))$ pour tout $s, t \in [0, 1]$. On s'en tient alors à l'estimation de l'opérateur, de ses fonctions propres et ses valeurs propres. Notons $(X_i)_{i \in [1, n]}$ nos observations, avec n la taille de cet échantillon.

3.2.1 Opérateur de covariance empirique

Afin de définir l'estimation de Γf , qu'on notera $\widehat{\Gamma} f$, on utilisera la moyenne empirique et la matrice de variance empirique issue des données $(X_i)_{i \in [1, n]}$. On a,

$$\begin{aligned}
\widehat{\Gamma}f(s) &= \frac{1}{n} \sum_{i=1}^n \langle f, X_i \rangle X_i(s) \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^1 f(t) X_i(t) dt X_i(s) \\
&= \int_0^1 f(s) \frac{1}{n} \sum_{i=1}^n X_i(s) X_i(t) ds \\
&:= \int_0^1 f(s) \widehat{K}(s, t) ds
\end{aligned}$$

où l'on a noté \widehat{K} un estimateur de la fonction de covariance. Il s'agit donc de l'opérateur de covariance pour lequel on a remplacé le noyau K par une estimation de celui-ci. À noter que $\widehat{\Gamma}$ est aussi un opérateur de H-S auto-adjoint, et le théorème 3.3 s'applique toujours. Par ailleurs, cet estimateur est non biaisé. En effet,

$$\mathbb{E}[\widehat{\Gamma}f(s)] = \mathbb{E}\left[\int_0^1 f(t) X_i(t) X_i(s) dt\right] = \int_0^1 f(t) \mathbb{E}[X_i(t) X_i(s)] dt = \Gamma f(s)$$

Concernant la consistance de cet estimateur, on pourra se référer au livre de D.Bosq, p. 102 [6]

3.2.2 Reconstruction des fonctions propres

On note $(\varphi_k)_{k \in K}$ et $(\lambda_k)_{k \in K}$ les fonctions et valeurs propres de $\widehat{\Gamma}f$ et on ordonnera également ses valeurs propres tel que : $\lambda_1 \geq \lambda_2 \geq \dots$

En pratique, il est difficile de trouver les valeurs propres estimées, on procède donc de la façon suivante.

En s'appuyant sur la proposition 2.10, on remarque qu'il suffit d'approcher $\widehat{\Gamma}$ par une matrice $\hat{G} = (\langle \widehat{\Gamma}e_j, e_k \rangle)_{1 \leq j, k \leq K}$, avec $(e_j)_{j \geq 1}$ une base hilbertienne bien choisie, et K suffisamment grand. On a,

$$\widehat{\Gamma}f = \sum_{j \in \mathbb{N}} \langle \widehat{\Gamma}f, e_j \rangle e_j = \sum_{j \in \mathbb{N}} \langle \widehat{\Gamma}(\sum_{k \in \mathbb{N}} \langle f, e_k \rangle e_k), e_j \rangle e_j = \lim_{n \rightarrow +\infty} \sum_{j, k=1}^n \langle f, e_k \rangle \hat{G}_{k,j} e_j$$

Ainsi, pour notre K suffisamment grand, $\widehat{\Gamma}\hat{\varphi}$ peut être approché par $\sum_{j,k=1}^K \langle \hat{\varphi}, e_k \rangle \hat{G}_{k,j} e_j$. En utilisant cette approximation, on trouve :

$$\hat{\lambda} \langle \hat{\varphi}, e_k \rangle = \langle \widehat{\Gamma}\hat{\varphi}, e_k \rangle \approx \sum_{i,j=1}^K \langle \hat{\varphi}, e_i \rangle \hat{G}_{i,j} \langle e_j, e_k \rangle = \sum_{j=1}^K \hat{G}_{k,j} \langle \hat{\varphi}, e_j \rangle$$

Ainsi, on obtient que $\sum_{j=1}^K \hat{G}_{k,j} \approx \hat{\lambda}$.

\hat{G} est symétrique définie positive, on peut donc appliquer le théorème spectral. Soient $\hat{v}_1, \dots, \hat{v}_K$ les vecteurs propres de \hat{G} . On va dans un premier temps calculer $\hat{G} = (\langle \widehat{\Gamma}e_j, e_k \rangle)_{1 \leq j, k \leq K}$. Puis nous calculerons les fonctions propres $\hat{v}_1, \dots, \hat{v}_K$ de \hat{G} et ses valeurs propres

$\hat{\delta}_i$ respectives qui vérifient $\hat{\delta}_1 \geq \dots \geq \hat{\delta}_K$. Pour finir, on reconstruira $\hat{\varphi}_l = \sum_{j=1}^K v_{l,j} e_j$, pour obtenir les fonctions propres de $\hat{\Gamma}$ où $v_{l,j}$ représente la l -ième coordonnée du j -ième vecteur propre de \hat{G} .

3.3 Représentation des variables

Après avoir estimé les K premières composantes principales, on peut maintenant représenter notre nuage $(X_i)_{i \in [1,n]}$ dans cette base.

Définition 3.5. *Le développement de Kharunen-Loève d'une fonction aléatoire X est sa représentation dans la base de fonctions propres φ_k de son opérateur de covariance Γ .*

Supposons $\mathbb{E}[X] = 0$ et $\int_0^1 \mathbb{E}[X^2] < \infty$, on a :

$$X(t) = \sum_{i=1}^{\infty} \langle X, \varphi_i \rangle \varphi_i(t) = \sum_{i=1}^{\infty} \sqrt{\lambda_i} \xi_i \varphi_i(t) \quad \text{où} \quad \xi_i = \frac{\langle X, \varphi_i \rangle}{\sqrt{\lambda_i}}$$

En pratique, on ne dispose pas du développement de Kharunen-Loève ou du moins très rarement. On va donc utiliser les estimations précédemment calculées. Ainsi pour tout $1 \leq i \leq n$, X_i qui est élément de $L^2([0, 1]) \cap C([0, 1])$ peut être réécrit de la sorte :

$$X_i(t) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} \xi_{ij} \varphi_j(t) \quad \forall t \in [0, 1]$$

Où $\xi_{ij} = \frac{\langle X_i, \varphi_j \rangle}{\sqrt{\lambda_j}}$. On le projette finalement sur l'espace engendré par les K premières composantes principales :

$$P_{Vect(\varphi_1, \dots, \varphi_k)}(X_i(t)) = \sum_{j=1}^k \sqrt{\lambda_j} \xi_{ij} \varphi_j(t) \quad \forall t \in [0, 1]$$

Chapitre 4

Application numérique

Dans ce chapitre, nous allons effectuer une ACPf sur des données simulées et des données réelles en utilisant les méthodes d'estimation vues dans les précédents chapitres. La mise en oeuvre est réalisée à l'aide du logiciel R. Dans un premier temps nous allons travailler sur des données simulées, puis dans un second temps nous allons utiliser des données réelles grâce au package fda de R. Les données réelles en question porteront sur : la météo Canadienne ainsi que l'évolution de la force appliquée entre le pouce et l'index d'un individu.

Le code correspondant est disponible ici.

4.1 Méthode d'implémentation

Nous supposons dans tout ce chapitre que K et n sont deux entiers naturels qui représentent respectivement le nombre de points d'observations et le nombre de courbes observées.

On observe $X_i(t_j)$ avec $i \in \{1, \dots, n\}, j \in \{1, \dots, K\}$ et $t_j \in [0, 1]$, que l'on suppose centrés (quitte à considérer $X_i(t_j) - \frac{1}{n} \sum_{i=1}^n X_i(t_j)$). On rappelle la fonction de covariance empirique :

$$\forall s, t \in [0, 1], \hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^n X_i(s)X_i(t)$$

Déroulé de la méthode

On pose $(e_j)_{j \geq 1}$ une base hilbertienne de $L^2[0, 1]$.

Première étape : calculer \hat{G}

Définition 4.1. Soit f une fonction continue, la somme de Riemann associée à une division des intervalles équitablement répartie est

$$\int_0^1 f(t)dt = \lim_{K \rightarrow +\infty} \frac{1}{K} \sum_{j=1}^K f(t_j)$$

avec par défaut $t_1 = 0$ et $\forall j \in \{2, \dots, K-1\}$ $t_j = \frac{j-1}{K-1}$ et $t_K = 1$.

En réutilisant la définition précédente, on a :

$$\forall j, k \in \{1, \dots, K\}, \hat{G} = \langle \hat{\Gamma} e_j, e_k \rangle = \int_0^1 \int_0^1 \hat{K}(s, t) e_j(s) e_k(t) ds dt$$

On va donc approcher \hat{G} par sa somme de Riemann :

$$\frac{1}{K^2} \sum_{l, h=1}^K \hat{K}(t_l, t_h) e_j(t_l) e_k(t_h).$$

Deuxième étape : calcul des vecteurs propres et des valeurs propres de \hat{G}

En reprenant les notations du paragraphe 3.2.2, nous avons $(\hat{v}_k)_{1 \leq k \leq K}$, les fonctions propres de \hat{G} et $\hat{\delta}_1, \dots, \hat{\delta}_K$ ses valeurs propres rangées par ordre décroissant. En utilisant la fonction `eigen` de R on obtient les valeurs propres et fonctions propres de \hat{G} . Certes, la moyenne est un aspect très important des données, mais nous disposons déjà d'une technique simple pour l'identifier. Par conséquent, nous soustrayons généralement la moyenne de chaque variable des valeurs de variable correspondantes avant de procéder à l'ACPF.

```

1 # X représente la matrice contenant les données
2 X.cent <- scale(X, scale=FALSE) # On centre les données
3 Khat <- crossprod(X.cent)/(n*K)
4 Ghat <- crossprod(t(L), Khat)
5 Ghat <- tcrossprod(Ghat, L) / K
6 eigen(Ghat)$values
7 eigen(Ghat)$vectors

```

Troisième étape : reconstruction des fonctions propres de $\hat{\Gamma}f$

Il nous reste à reconstruire les fonctions propres de $\hat{\Gamma}f$ que l'on notera $(\hat{\varphi}_j)_{1 \leq j \leq K}$. Pour cela, il faut effectuer un produit matriciel entre les vecteurs propres de \hat{G} précédemment calculés et les éléments de la base hilbertienne e_1, \dots, e_K . On a $\forall j \in \{1, \dots, K\}$

$$\hat{\varphi}_j = \sum_{i=1}^K \hat{v}_i^\top e_i \quad (4.1)$$

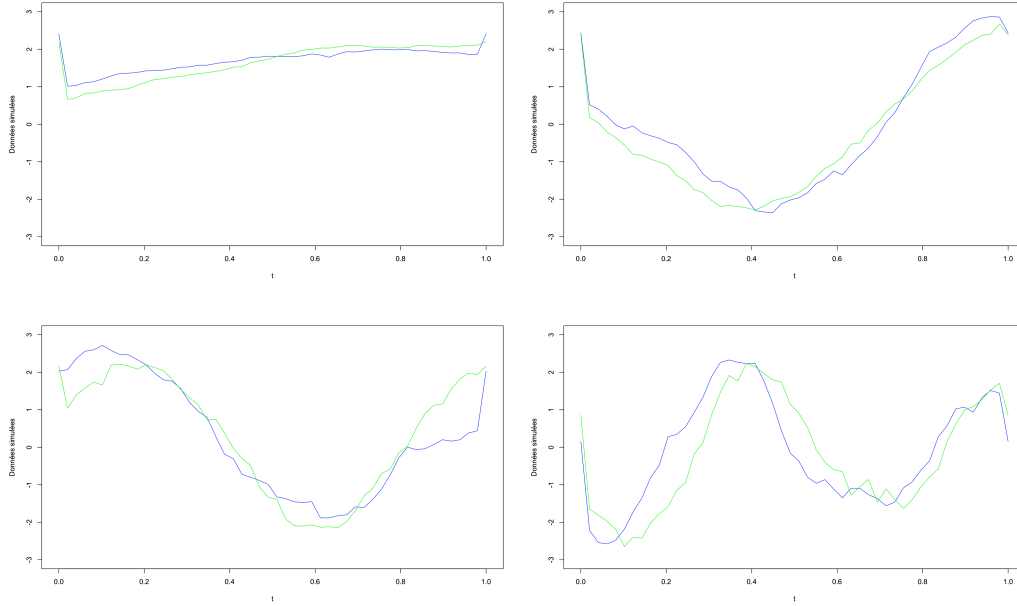
4.2 Application sur données simulées

La méthode est tout d'abord testée sur des données simulées. Dans cette partie nous supposons $n = 1000$ et $K = 500$ et $J = 50$. On pose $t = (t_1, \dots, t_K)$ une séquence avec un pas de $\frac{1}{K-1} = \frac{1}{499}$. On génère la base de Fourier comme base de $L^2[0, 1]$ dont la définition est donnée au chapitre [2.4]. On prendra ici $w = 2\pi$.

On simule X_1, \dots, X_K à partir de la fonction `simuX` définie en annexe. On y applique les trois étapes décrites précédemment. On simule également un mouvement brownien standard. Les résultats peuvent différer d'une simulation à une autre à coefficient -1 près. En effet, cela dépend de la base.

On pose J_m une matrice réelle de taille $J \times K$ contenant dans la première ligne que des 0.5 et qui pour tout $i \in \{1, \dots, J\}$, la i -ème ligne contient que des $0.5 + i$.

On pose également la matrice P de taille $J \times K$ qui contient pour tout $i \in \{1, \dots, J\}$ et $j \in \{1, \dots, d\}$ $P_{i,j} = \frac{1}{K}$.

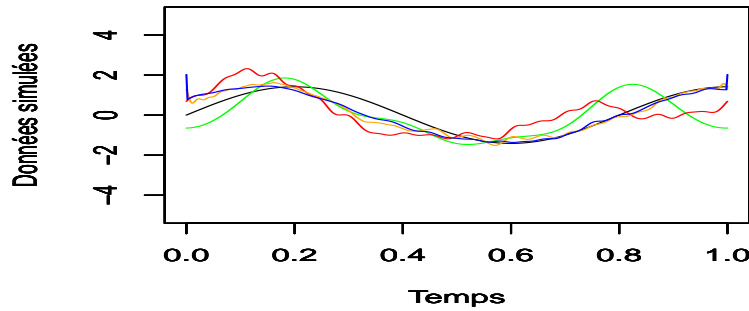


Représentation des quatre premières fonctions propres de X en bleu et du mouvement brownien standard en vert

On peut remarquer dans ces 4 graphiques que les fonctions propres estimées coïncident avec les fonctions propres d'un mouvement brownien standard. C'est bien le résultat attendu car nous avons simulé les données à partir de la loi normale. On définit l'effet de bord qui consiste à étudier le phénomène d'instabilité numérique au bord de l'intervalle. En essayant de calculer la solution approchée du problème, il arrive qu'on parvienne à avoir une bonne approximation sauf sur le bord. C'est ce qu'on peut voir dans le premier graphique sur le bord droit, la courbe bleue remonte tandis que la courbe verte continue sans manifester d'écart important. Ces effets apparaissent en fonction du choix de la base hilbertienne choisie.

Les réelles fonctions propres sont égales $\forall j \in \{1, \dots, K\}$, $\varphi_j(t) = \sqrt{2} \sin(\pi(j - 0.5)x)$.

On pose $\eta = \sqrt{2} \sin(\pi W)$ avec W une matrice de taille $J \times K$ tel que $\forall i \in \{1, \dots, J\}$ et $\forall j \in \{1, \dots, K\}$, $(W)_{i,j} = (J_m)_{i,j} * P_{i,j}$, qui représente la matrice contenant toutes les réelles fonctions propres. On remarque que si l'on prend les k premières colonnes de X avec $1 \leq k \leq K$ alors plus k est grand plus on se rapproche de η . On pose $n_1 = 10$, $n_2 = 50$, $n_3 = 100$, $n_4 = 500$.



La courbe verte représente les n_1 premières colonnes de X , la courbe rouge les n_2 premières, la courbe orange les n_3 premières et la courbe bleu les n_4 premières, enfin la courbe noire représente η . On remarque sur le graphique que plus l'ACPF est réalisée sur de nombreuses observations plus elle converge vers la valeur réelle. On peut également apercevoir un effet de bord sur la courbe bleue.

4.3 Application sur données réelles

Dans toute cette partie, nous supposons que le nombre de variables est égale à d et le nombre d'observation est égal à K , avec d et K deux entiers naturels. On utilisera également la base de Fourier comme base de $L^2[0, 1]$ donnée au point 2.4.

La section 3.1.2 nous donne le résultat suivant : les valeurs propres associées à chaque fonction propre correspondent aux variances des composantes principales. La variance totale du jeu de données est donc donnée par la somme des valeurs propres de l'opérateur de covariance. Nous allons estimer ses K premières valeurs propres par $\hat{\lambda}_1, \dots, \hat{\lambda}_K$ pour avoir une idée de la variance totale.

Pour déterminer le nombre K de composantes principales nécessaires à la compréhension de l'essentiel de l'information, on va donc chercher à conserver les composantes principales ayant les plus grandes valeurs propres.

Définition 4.2. *La proportion désigne la part de la variabilité des données expliquée par chaque composante principale. Pour un certain k tel que $1 \leq k \leq K$ elle est égale à :*

$$\frac{\hat{\lambda}_k}{\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_d}$$

Si cette valeur est élevée, la composante principale explique une grande part de la variabilité et est donc suffisamment utile pour être conservée.

Définition 4.3. *On appelle valeur cumulée, la proportion cumulée de la variabilité des données de composantes principales consécutives.*

Pour un certain k tel que $1 \leq k \leq K$, la proportion de variance cumulée ou valeur cumulée expliquée par les k premières composantes principales est,

$$\frac{\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_k}{\hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_d}$$

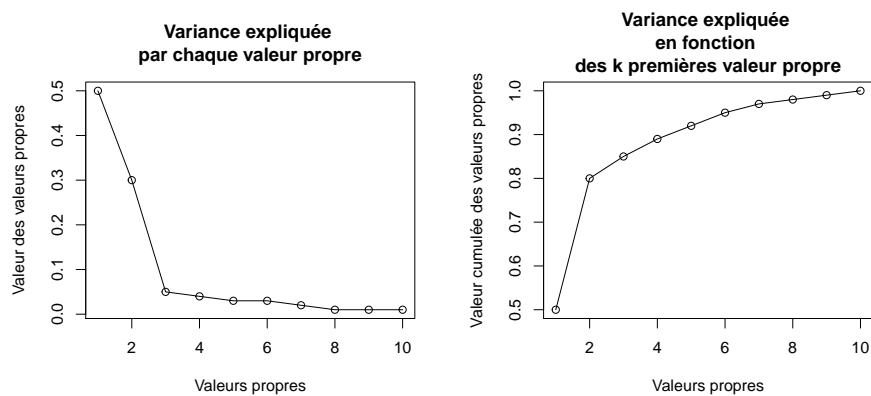
Ce ratio peut être utilisé pour choisir le nombre de composantes principales à conserver pour expliquer un niveau acceptable de variance. Ce niveau est déterminé par le statisticien.

Définition 4.4. *On appelle score les combinaisons linéaires de l'échantillon de données déterminées par les coefficients associés à chaque composante principale.*

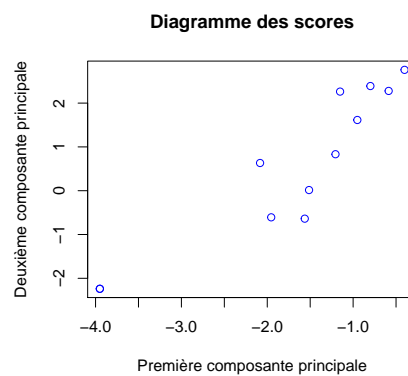
Notons \mathbf{A} la matrice de taille $n \times K$ contenant les K composantes principales. On pose :

$$\begin{aligned} A : \mathbb{R}^n &\longrightarrow \mathbb{R}^K \\ f &\longmapsto f^\top \mathbf{A} \end{aligned}$$

Pour obtenir le score d'une observation, nous devons indiquer chacune de ses valeurs dans l'équation linéaire de la composante principale.



Le graphe de gauche montre le pourcentage de variance expliqué par chaque composante principale. A partir de 4 composantes principales, ajouter une autre n'est pas très informatif. Le graphe de droite montre le pourcentage cumulé de variance expliqué par chacune des composantes principales. Si on se fixe une proportion de variance expliquée de 95%, on peut se contenter de 6 composantes principales.



Données CanadianWeather sur les précipitations.

Le diagramme des scores affiche les scores de la deuxième composante principale par rapport à ceux de la première. Il permet de connaître la structure des données.

Utilisation du package `fda` (functional data analysis) de R, et des jeux de données associés.

Jeu de données `CanadianWeather`

Cet ensemble contient la température et les précipitations quotidiennes moyennes de 1960 à 1994 dans 31 stations situées au Canada. Nous allons nous concentrer sur les données de température pour cette étude.

St. Johns	Halifax	Sydney	Yarmouth	Charlottvl
Fredericton	Scheffervll	Arvida	Bagottville	Quebec
Sherbrooke	Montreal	Ottawa	Toronto	London
Thunderbay	Winnipeg	The Pas	Churchill	Regina
Pr. Albert	Uranium Cty	Edmonton	Calgary	Kamloops
Vancouver	Victoria	Pr. George	Pr. Rupert	Whitehorse
Dawson	Yellowknife	Iqaluit	Inuvik	Resolute

Table 4.1 – Noms des stations météorologiques canadiennes

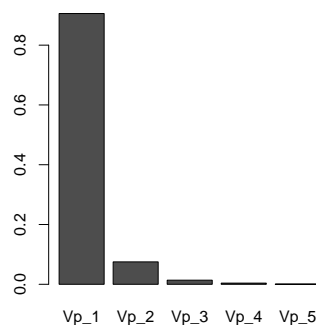
Après avoir effectué toutes les étapes décrites dans la section 4.1.1 sur les données, nous commençons par étudier les valeurs propres.

```

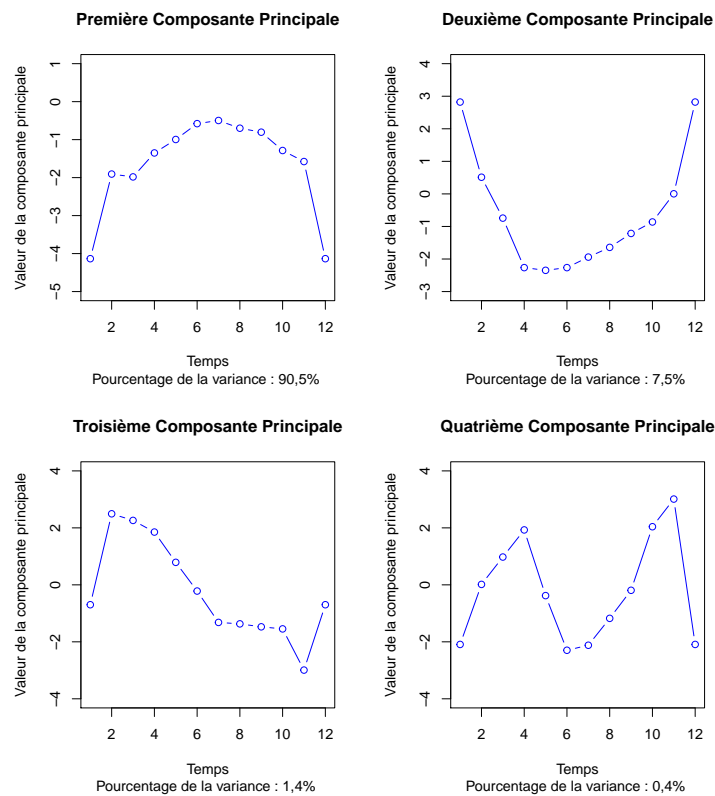
1 > V[1 :5]
2 [1] 0.9057876283 0.0749526316 0.0137975456 0.0039379790 0.0007172274
3 > V[1]+V[2]
4 [1] 0.9807403
```

On obtient la valeur des cinq premières valeurs propres. En sommant les deux premières, on obtient que 98% des données sont concentrées dans les deux premiers axes principaux.

Poids des 5 premières valeur propre



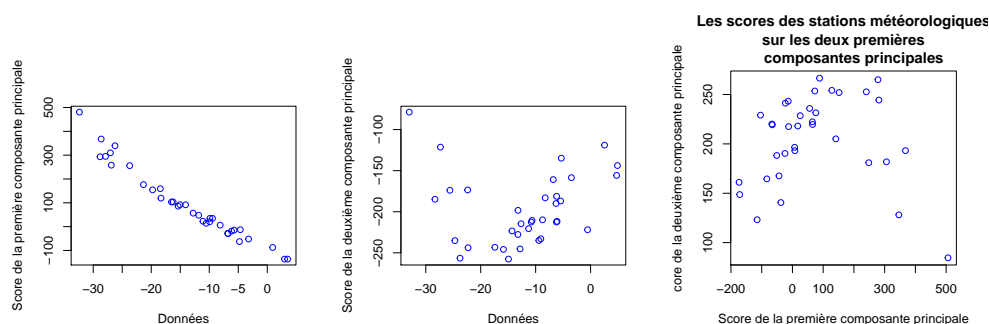
Ce graphique correspond au barplot des cinq premières valeurs propres. Il permet de remarquer du premier coup d'œil les valeurs propres importantes.



Plot des quatre premières composantes principales

On étudie la première composante principale des données. On remarque que tout au long de l'année sa valeur est négative. Cela signifie que le poids placé sur les températures hivernales est environ 4 fois supérieur au poids placé sur les températures estivales. Cela signifie que la plus grande variabilité entre les stations météorologiques sera trouvée en pondérant fortement les températures hivernales, avec seulement une légère contribution des mois d'été. Le temps au Canada est donc plus variable en hiver. La première composante principale a une proportion de 90,5 %. Cette proportion est la plus élevée, elle domine donc fortement toutes les autres variations des autres composantes principales. Les stations météorologiques pour lesquelles le score associé à la première composante principale est faible auront des hivers bien plus chauds que la moyenne combinés à des étés chauds, et les deux scores les plus faibles correspondent aux villes de Vancouver et Victoria. Le score le plus élevé revient à Resolute qui est une ville située dans le Haut-Arctique ce qui n'est pas étonnant.

Étudions maintenant la seconde composante principale. Elle ne représente que 7,5% de la variation totale, et consiste en une contribution négative pour les mois d'hiver et une contribution positive pour les mois d'été, correspondant donc à une mesure d'uniformité de la température tout au long de l'année. L'un des scores associés à la deuxième composante le plus faible est attribué à Prince Rupert, pour lequel l'écart est relativement faible entre l'hiver et l'été. La station Winnipeg, par contre, ont des étés chauds et des hivers très froids et reçoivent d'importants scores positifs de deuxième composante. La troisième (resp. quatrième) composante principale explique de faibles proportions de la variation : 1,4 % (resp. 0,4%).

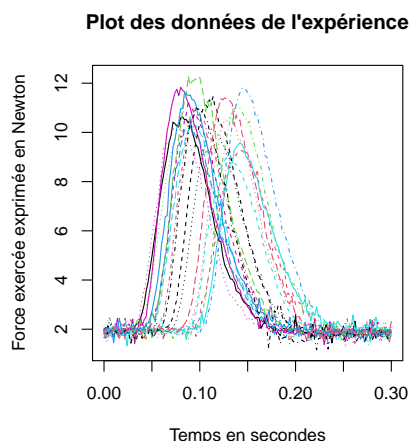


Plot des scores des deux premières composantes principales ainsi que le plot des scores de la deuxième composante principale en fonction de la première

L'examen des scores de chaque composante est important. Dans le troisième plot ci-dessus, les stations sont dans le coin supérieur gauche car elles ont des hivers plus chauds que la plupart des stations donc faible sur la première composante principale et moins de variation de température été-hiver (élevé sur PC 2). Resolute est situé dans le coin inférieur droit car a un hiver extrêmement froid donc élevé sur la première composante principale et faible sur la deuxième car grande variation de température été-hiver.

Jeu de données pinch

Nous étudions maintenant quelques données biomécaniques. Ces données représentent vingt enregistrements au cours d'une manipulation de la force exercée lors d'un bref pincement par le pouce et l'index. Le sujet devait maintenir une certaine force de fond sur un dynamomètre, puis presser le multimètre en visant une valeur maximale spécifiée, pour revenir ensuite à un niveau de repos fixé à 2 Newtons. La force était échantillonnée 500 fois par seconde. Le but de l'expérience était d'étudier la neurophysiologie du groupe musculaire pouce-index.



Vingt enregistrements de la force exercée par le pouce et l'index.

```

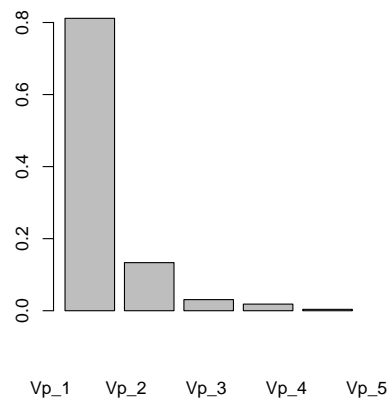
1 > V[1:5]
2 [1] 0.820049101 0.134921476 0.022871035 0.016662250 0.003675789
3 > V[1]+V[2]
4 [1] 0.9549706

```

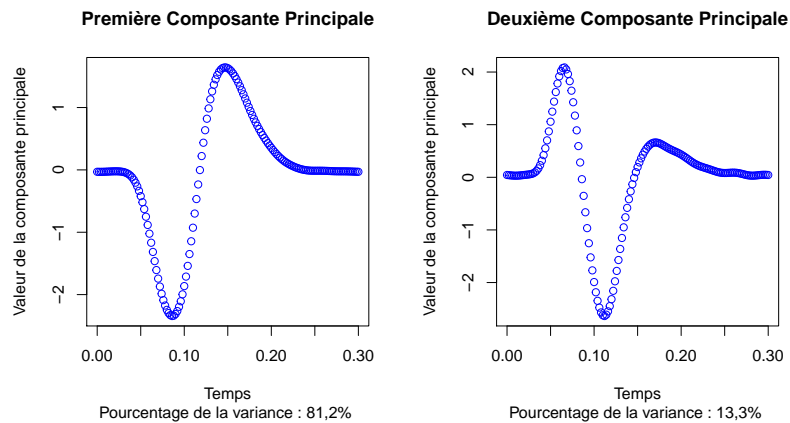
Code R permettant d'obtenir les cinq premières valeurs propres ainsi que la somme des deux premières

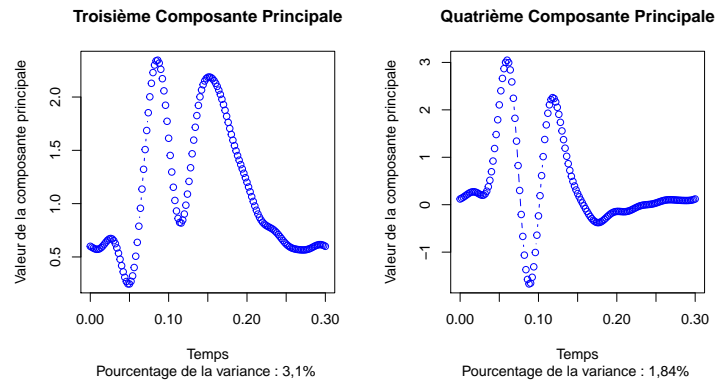
On obtient les valeurs des cinq premières valeurs propres. En sommant les deux premières, on obtient que plus de 95% des données sont expliquées par les deux premières composantes principales.

Poids des cinq premières valeurs propres



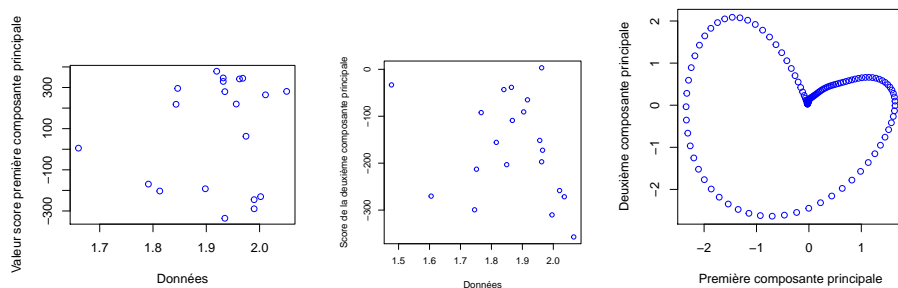
Ce résultat se confirme avec le barplot des cinq premières valeurs propres. La première est largement plus significative que les autres. Les valeurs propres après la cinquième sont très proches de 0 et n'expliquent quasi plus les données.





On étudie la première composante principale des données. On remarque qu'à partir de 0.05 secondes sa valeur commence à devenir négative. C'est exactement au même moment que l'on ordonne à l'individu de presser son pouce et son index. Cette décroissance de la valeur de la première composante principale s'opère jusqu'à l'individu ait commencé à réaliser les mouvements de l'étude pour les 20 enregistrements. Puis on remarque que les valeurs passent du négatif au positif durant tout le temps que l'individu presse son pouce et son index pour les 20 enregistrements. Cela signifie que la plus grande variabilité entre les enregistrements sera trouvée lors de l'effort, avec seulement une légère contribution lors du début de l'expérience. La première composante principale a une proportion de 81,2 %. Cette proportion est la plus élevée, elle domine donc fortement toutes les autres variations des autres composantes principales. Les enregistrements pour lesquelles le score correspondant à la première composante principale est faible auront des forces exercées plus faible que la moyenne, et le score le plus faible correspond à l'enregistrement 3. Le score le plus élevé revient à l'enregistrement 15.

Étudions maintenant la seconde composante principale. Elle représente 13,3% de la variation totale, et consiste en une contribution positive pour les efforts faibles et une contribution négative pour les efforts importants. Sur cette composante, l'un des scores les plus faible est attribué à l'enregistrement 15, pour lequel l'écart est relativement faible entre le repos et l'effort. L'enregistrement 3, par contre, a des forces exercées élevées pendant l'effort et faible pendant le repos et reçoivent d'importants scores positif de deuxième composante. La troisième (resp. quatrième) composante principale explique de faibles proportions de la variation : 3,1 % (resp 1.84%).



Les deux premiers graphiques représentent respectivement les scores de la première et de la deuxième composante principale. Le troisième représente le plot de la deuxième composante principale en fonction de la première.

Conclusion

Le concept d'Analyse en Composantes Principales pour données fonctionnelles ne diffère finalement pas du cas vectoriel. En effet, bien que le problème soit plus délicat à résoudre (on cherche ici à minimiser l'inertie sur un espace de fonctions), il suffit toujours de projeter les données sur l'espace engendré par les k premiers vecteurs propres de l'opérateur de covariance pour minimiser l'inertie.

La principale subtilité ici est la nature des données, qui est dans son essence plus complexe, et on arrive ainsi à un résultat plus puissant : les données vivant initialement dans un espace de dimension infinie sont résumées dans un espace de dimension fini. Du moins en théorie. En pratique, on se heurte à la difficulté d'identification des fonctions propres de l'opérateur de covariance, les composantes principales. Une solution avancée dans ce mémoire est de les estimer à partir des données. Une autre approche possible, mais non explorée dans ce mémoire, peut être déduite de la remarque [2] chap. 3. On pourrait estimer ces fonctions à partir d'un algorithme du point fixe.

Annexe A

Théorèmes généraux

Dans cette annexe, on rappelle les théorèmes utilisés dans le mémoire.

Théorème A.1. Théorème de Bessel. *Soit $(\varphi_n)_{n \in \mathbb{N}}$ une base orthonormale, alors les propositions sont équivalentes :*

1. (φ_n) est complète
2. $\forall f \in L^2([0, 1]), f = \sum_{i=1}^{+\infty} \langle f, \varphi_i \rangle \varphi_i$
3. $\|f\|^2 = \sum_{i=1}^{+\infty} \langle f, \varphi_i \rangle^2$

.

Nous énonçons ensuite le théorème de Mercer dans le cadre de ce mémoire, soit pour K défini sur $[0, 1] \times [0, 1]$, et $\Gamma_K \in \mathbb{L}(L^2([0, 1]))$ l'opérateur intégral associé à K .

Théorème A.2. Théorème de Mercer. *Si K est un noyau continu de type positif et symétrique, alors :*

1. *Il existe une base hilbertienne $(e_i)_i$ de $L_2([0, 1])$ constituée de fonctions propres de Γ_K*
2. *Les fonctions propres correspondant à des valeurs propres strictement positives sont continues sur $[0, 1]$*
3. *K admet la représentation*

$$K(s, t) = \sum_{i=1}^{\infty} \lambda_i e_i(s) e_i(t)$$

où la convergence est absolue pour chaque $s, t \in [0, 1]$, et uniforme sur $[0, 1] \times [0, 1]$.

Ce théorème est un corollaire du théorème de diagonalisation des opérateurs de H-S auto-adjoints.

Annexe B

Codes R

Dans cette annexe sont proposés les codes utilisés dans le chapitre 4.

Code R permettant de générer la base de Fourier

```
1  fourier_pair <- function(t, k) # Calcul des éléments pairs de la base
2
3    return(sqrt(2) * cos(k * pi * t))
4
5  fourier_impair <- function(t, k) # Calcul des éléments impairs de la base
6
7    return(sqrt(2) * sin((k-1) * pi * t))
8
9  base_fourier <- function(n, t){ # Fonction renvoyant la base Fourier complète
10
11    l <- matrix(1,n,length(t)) # On crée une matrice qui va contenir la base
12                                # de Fourier
13
14    for(i in 2 : n){
15
16      if(i %% 2 == 0){
17
18        l[i,] = fourier_pair(t, i)
19
20      }
21
22      else{
23
24        l[i,] = fourier_impair(t, i)
25
26      }
27    }
28    return(l) # On retourne la matrice l contenant la base de Fourier complète
29  }
```

Code R permettant de réaliser les 3 étapes décrites dans la partie 4.1

```

1 M_ACPF <- function(x, l, d, p){
2
3   x.cent <- scale(x, scale = FALSE) # On centre les données contenues dans x
4
5   k_hat <- crossprod(x.cent) / (d * p) # Calcul de Khat
6
7   g_hat <- tcrossprod(l, t(k_hat)) # Calcul de Ghat
8
9   g_hat <- tcrossprod(g_hat, l) / p # Calcul de Ghat
10
11   return(g_hat)
12
13 }
```

Code R permettant de simuler X dans la partie 4.2

```

1 simuX = function(n, p, j = 50, lambda = (1:J)^(-2), loi = 'norm', ...) {
2
3   if (loi == 'norm')
4
5     ksi = rnorm(n * j)
6
7   else
8
9     ksi = runif(n * j, -sqrt(3), sqrt(3))
10
11   ksi = matrix(ksi, n, j) * matrix(rep(sqrt(lambda), n), n, j, byrow = TRUE)
12
13   x = matrix(rep(seq(0, 1, length.out = p), j), j, p, byrow = TRUE)
14
15   jm = matrix(rep(seq(0.5, j - 0.5, by=1), p), j, p)
16
17   base = sqrt(2) * sin(pi * jm * x)
18
19   pertinit <- matrix(rep(rnorm(n), p), n, p, byrow = FALSE)
20
21   tcrossprod(ksi, t(base)) + pertinit
22
23 }
```

Code R permettant de calculer η de la partie 4.2

```

1 k = 500 j = 50
2 jm = matrix(rep(seq(0.5, J-0.5, by=1), p), j, k)
3
4 x = matrix(rep(seq(0,1,length.out = p), j), j, k, byrow = TRUE)
5
6 eta = sqrt(2) * sin(pi * jm * x)
```

Bibliographie

- [1] P. Hall *Principal component analysis for functional data : methodology, theory, and discussion. The Oxford handbook of functional data analysis* 210234, Oxford Univ. Press, Oxford, 2011.
- [2] J. Droniou. *Intégration et Espaces de Sobolev à Valeurs Vectorielles*. 2001. hal-01382368.
- [3] A. Roche. *Modélisation statistique pour données fonctionnelles : approches non-asymptotiques et méthodes adaptatives*. Statistiques [math.ST]. Université Montpellier II, 2014. Français. tel- 01023919.
- [4] M. L. Gallardo. [http ://www.lmpt.univ-tours.fr/~gallardo/EspHilbert6bis.pdf](http://www.lmpt.univ-tours.fr/~gallardo/EspHilbert6bis.pdf).
- [5] P. Cardaliaguet. *Optimisation et programmation dynamique*. Master mention Mathématiques et Applications, 1ère année.
- [6] D. Bosq. *Linear Processes in Function Spaces Theory and Applications*. Springer Edition.