

---

UNIVERSITE Dr. MOULAY TAHAR DE SAÏDA  
FACULTE DES SCIENCES & TECHNOLOGIE  
DEPARTEMENT DE MATHEMATIQUES & INFORMATIQUE

---

Mémoire de master

# Modèle linéaire fonctionnel pour la variable explicative est bruitée

---

*CHERIFI Fouad*

# Table des matières

<b>1</b>	<b>Introduction générale</b>	<b>5</b>
<b>2</b>	<b>La méthode Moindres carrés orthogonaux</b>	<b>16</b>
2.1	Cas multivarié . . . . .	16
2.2	Cas fonctionnel . . . . .	21
2.2.1	Construction de l'estimateur splines de régression . . . . .	22
2.2.2	Résultat de convergence . . . . .	29
2.2.3	Commentaires . . . . .	31
<b>3</b>	<b>La méthode de régression fonctionnelle sur composantes principales</b>	<b>32</b>
3.1	L'ACP fonctionnelle pour la régression . . . . .	32
3.1.1	La régression sur composantes principales . . . . .	34
3.1.2	Régression sur composantes principales pour les variables sont bruitées . . . . .	36
3.2	Résultats asymptotiques . . . . .	41
<b>4</b>	<b>simulations</b>	<b>49</b>
	<b>Conclusion et Perspectives</b>	<b>53</b>
	<b>Bibliographie</b>	<b>56</b>

## Dédicace

Je dédie ce mémoire à Mes parents :

Ma mère, qui a oeuvré pour ma réussite, de par son amour, son soutien, tous les sacrifices consentis et ses précieux conseils, pour toute son assistance et sa présence dans ma vie, reçois à travers ce travail aussi modeste soit-il, l'expression de mes sentiments et de mon éternelle gratitude.

Mon père, qui peut être fier et trouver ici le résultat de longues années de sacrifices et de privations pour m'aider à avancer dans la vie. Puisse Dieu faire en sorte que ce travail porte son fruit ; Merci pour les valeurs nobles, l'éducation et le soutien permanent venu de toi. Mes frères et soeurs qui n'ont cessé d'être pour moi des exemples de persévérance, de courage et de générosité.

Je souhaite que Dieu leur préserve une longue vie.

## Remerciements

Grace à " Dieu " le tout puissant, j'ai pu achever ce mémoire.

Je tiens à exprimer toute ma reconnaissance à mon directrice de mémoire Madame Benziadi Fatima. Je la remercie de m'avoir encadré, orienté, aidé et conseillé.

J'adresse mes sincères remerciements à tous les professeurs, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté à me rencontrer et répondre à mes questions.

Je remercie mes très chers parents qui ont toujours été là pour moi, Vous avez tout sacrifié pour vos enfants n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je suis redevable d'une éducation dont je suis fier.

Je remercie mon frère, ma soeur pour leur encouragement

Je tiens à remercier sincèrement les membres du jury qui me font le grand honneur d'évaluer ce travail.

Enfin, je remercie tous mes Ami(e)s Pour leur sincère amitié et confiance, et à qui je dois ma reconnaissance et mon attachement.

À tous ces intervenants, je présente mes remerciements, mon respect et ma gratitude.

# Chapitre 1

## Introduction générale

La statistique fonctionnelle a connu un très important développement ces dernières années. Cette branche de la statistique vise à étudier des données qui, de part leur structure et le fait qu'elles soient collectées sur des grilles très fines, peuvent être assimilées à des courbes, par exemple fonctions du temps. Le besoin de considérer ce type de données, maintenant couramment rencontré sous le nom de données fonctionnelles dans la littérature, est avant tout un besoin pratique. Compte tenu des capacités actuelles des appareils de mesure et de stockage informatique, les situations pouvant fournir de telles données sont multiples et issues de domaines variés : on peut imaginer par exemple des courbes de croissance, de température, des images observées par satellite, . . .

Cependant, au-delà de cet aspect pratique, il est nécessaire de donner un cadre théorique pour l'étude de ces données. Bien que la statistique fonctionnelle ait les mêmes objectifs que les autres branches de la statistique (analyse de données, inférence, . . .), les données ont cette particularité de prendre leurs valeurs dans des espaces de fonctions, et les méthodes usuelles de la statistique multivariée sont ici mises en défaut. Par exemple, considérons que l'on dispose des observations de  $n$  courbes en  $p$  points de discrétisation, ces courbes étant utilisées comme prédicteur d'une autre variable. Si on regroupe

ces données notées  $x_{ij}$  (pour  $i$  allant de 1 à  $n$  et  $j$  allant de 1 à  $p$ ) sous forme d'une matrice de taille  $n \times p$ ,

$$X = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}$$

la méthode des moindres carrés ordinaires, très courante en statistique multivariée, peut donner de très mauvais résultats dans cette situation, puisque cette méthode amène à l'inversion de la matrice  $X^T X$  qui peut se révéler difficile voire impossible pour deux raisons. La première est que  $p$  est généralement grand (on peut même avoir  $p > n$ , et ainsi une matrice  $X^T X$  non inversible). La seconde raison est qu'il y a de fortes chances d'avoir une colinéarité importante entre les  $p$  prédicteurs du fait qu'ils sont les points de mesure d'une même fonction. Pour contourner ce problème, des solutions ont été envisagées, les plus courantes étant

- la "ridge regression" introduite initialement par Hoerl et Kennard, 1980[37], qui consiste à ajouter un terme de pénalisation dans le critère des moindres carrés. Cela amène à inverser la matrice  $(X^T X + \lambda I_p)$  avec  $\lambda$  réel strictement positif et  $I_p$  matrice identité de taille  $p$  au lieu de  $X^T X$ ,
- la régression sur composantes principales, qui consiste à réduire la dimension  $p$  en utilisant les  $k$  premières composantes principales issues de l'analyse en composantes principales du tableau  $X$  avec  $k$  entier non nul "convenablement" choisi,
- la régression "partial least squares" (voir Helland, 1990[38]), qui est une méthode algorithmique basée à chaque étape sur la régression par moindres carrés ordinaires sur les résidus de l'étape précédente.

Ces différentes méthodes sont étudiées et comparées dans un article de Frank et Friedman (1993)[27], en vue de leur application dans le domaine de la chimiométrie. Comme souligné par Hastie et Mallows (1993)[39] dans leur discussion de cet article, l'approche qui consiste à voir une courbe uniquement à travers un vecteur de points de mesure est réductrice, ne serait-ce que par le fait que les points de mesure doivent être les mêmes pour chaque courbe observée, ce qui n'est pas forcément le cas en pratique. Cette approche conduit également au problème que l'on perd la structure de courbe si on utilise uniquement les mesures de la courbe en certains points. C'est pourquoi il paraît préférable de traiter les données en tenant compte de leur nature fonctionnelle.

Les tout premiers travaux dans lesquels on retrouve cette idée de données fonctionnelles sont finalement relativement "anciens". Rao (1958) et Tucker (1958)[52] envisagent ainsi l'analyse en composantes principales et l'analyse factorielle pour des données fonctionnelles et considèrent même explicitement les données fonctionnelles comme un type particulier de données. Par la suite, Ramsay (1982)[53] dégage la notion de données fonctionnelles et soulève la question de l'adaptation des méthodes de la statistique multivariée à ce cadre fonctionnel. À partir de là, les travaux pour explorer la statistique fonctionnelle commencent à se multiplier, pour finalement aboutir aujourd'hui à des ouvrages faisant référence en la matière, comme par exemple les monographies de Ramsay et Silverman (2002 et 2005)[54].

Les travaux réalisés constituent ainsi à l'heure actuelle une littérature très dense, que ce soit sur un plan théorique ou appliqué. D'un point de vue théorique, la notion de variable aléatoire fonctionnelle (c'est-à-dire une variable aléatoire à valeurs dans un espace de fonctions) est apparue, ce qui a nécessité d'explicitier des notions simples pour une telle variable, comme

par exemple l'espérance ou l'opérateur de covariance. Des techniques déjà existantes en statistique multivariée (comme par exemple l'analyse en composantes principales) peuvent alors se développer dans ce cadre fonctionnel, utilisant notamment les connaissances en théorie des opérateurs dans les espaces de Hilbert (voir par exemple Dunford et Schwarz, 1963[22], Gohberg et Krein, 1971[31]). Parmi les points de départ de cette généralisation au cadre fonctionnel des méthodes de la statistique multivariée, Deville (1974)[23] introduit une analyse en composantes principales de courbes, tandis que la thèse de Dauxois et Pousse (1976)[24] va au-delà de l'analyse en composantes principales dans un espace de Hilbert, s'attachant à traiter un certain nombre de méthodes regroupées sous le nom d'analyses factorielles, dont l'analyse en composantes principales fait partie. L'article de Dauxois, Pousse et Romain (1982)[25] aborde davantage les aspects asymptotiques du problème, fournissant notamment des résultats de convergence pour l'opérateur de covariance empirique.

Dans ce travail, on propose d'apporter une contribution à l'étude des données fonctionnelles dans le contexte où la variable fonctionnelle sert à expliquer un phénomène représenté par une autre variable. Le problème qui va nous intéresser est celui de la régression dans le cas où la variable explicative est fonctionnelle. C'est un sujet sur lequel la littérature est très conséquente. D'un point de vue très général, ce modèle de régression fonctionnelle peut s'écrire

$$Y_i = r(X_i) + \epsilon_i, \quad i = 1, \dots, n, \quad (1.1)$$

où, pour tout  $i = 1, \dots, n$ ,

- $X_i$  appartient à un espace de Hilbert  $H$  dont le produit scalaire et la seminorme associée seront notés respectivement  $\langle \cdot, \cdot \rangle$  et  $\|\cdot\|_H$ . Notons que  $X_i$  pourra être aléatoire ou non, suivant que l'on considère un modèle à plan aléatoire ou à plan fixe,



- $Y_i$  est une variable aléatoire réelle (variable d'intérêt),
- $\epsilon_i$  est une variable aléatoire d'erreur,
- les variables aléatoires considérées sont toutes définies sur un même espace probabilisé  $(\Omega, \mathcal{A}, \mathbb{P})$ .

Le but est alors d'estimer l'opérateur  $r : H \rightarrow \mathbb{R}$  inconnu, sur la base des données  $(X_i, Y_i)_{i=1, \dots, n}$ .

Dans ce travail, on va s'intéresser à un modèle un peu plus particulier que le modèle précédent (1.1). Il s'agit du modèle linéaire fonctionnel, qui s'écrit sous la forme

$$Y_i = \langle \alpha, X_i \rangle + \epsilon_i, \quad i = 1, \dots, n, \quad (1.2)$$

où le but est d'estimer  $\alpha \in H$  inconnu, sur la base des données  $(X_i, Y_i)_{i=1, \dots, n}$ . Ce modèle est toujours l'objet de travaux récents, comme en témoignent par exemple les articles de Cardot, Ferraty et Sarda (1999, 2003)[13] étudiant le cas d'une variable réponse réelle et donnant une méthode d'estimation de  $\alpha$  à l'aide de ce qu'ils introduisent comme la régression sur composantes principales fonctionnelle, ou encore à l'aide de fonctions splines. Des vitesses de convergence sont également obtenues pour les estimateurs qu'ils construisent. Dans ce travail, c'est à ce modèle (1.2) que l'on va s'intéresser. C'est un modèle très populaire en analyse de données fonctionnelles, il est important de noter toutefois que d'autres modèles qui lui sont liés (par exemple des extensions de ce modèle linéaire fonctionnel (1.2) sont également sujets d'études récentes. Là encore, il semble impossible de lister les modèles existants relatifs à (1.1) et (1.2). On va cependant donner un tour d'horizon des modèles les plus fréquemment rencontrés.

- Il est possible de considérer que la variable d'intérêt est elle aussi de nature fonctionnelle, comme la variable explicative. Cuevas, Febrero et

Fraiman (2002)[14] ou encore Chiou, Müller et Wang (2004)[15] se sont ainsi intéressés à ce modèle linéaire fonctionnel avec une variable réponse fonctionnelle, c'est-à-dire lorsque  $Y_i$  (pour  $i = 1, \dots, n$ ) appartient aussi à un espace de Hilbert. Cuevas, Febrero et Fraiman (2002)[14] travaillent dans le cadre d'un modèle à plan fixe (c'est-à-dire que  $X_1, \dots, X_n$  sont non aléatoires), alors que Chiou, Müller et Wang (2004)[15] considèrent  $X_1, \dots, X_n$  aléatoires, basant leur méthode d'estimation sur les décompositions de Karhunen-Loève des  $X_i$  et  $Y_i$ ,  $i = 1, \dots, n$ . L'article plus bibliographique de Müller (2005)[48] passe en revue diverses méthodes d'estimation pour des modèles linéaires fonctionnels avec une variable d'intérêt réelle ou fonctionnelle et une variable explicative multidimensionnelle ou fonctionnelle. Il étend aussi ces méthodes, considérant notamment des variables explicatives dont les points de mesure peuvent être irrégulièrement espacés et en faible nombre, situation étudiée dans les articles de Yao, Müller et Wang (2005a,2005b)[62].

- Une autre extension possible du modèle linéaire fonctionnel est le modèle linéaire fonctionnel généralisé, qui est la version fonctionnelle du modèle linéaire généralisé introduit par Wedderburn (1974)[61] puis repris dans un ouvrage par McCullagh et Nelder (1989)[49]. Ce modèle linéaire fonctionnel généralisé a été notamment étudié par Cardot et Sarda (2005)[16] ainsi que Müller et Stadtmüller (2005)[50]. Dans ce modèle, on suppose que la loi conditionnelle de  $Y_i$  sachant  $X_i = x$  appartient à la famille exponentielle. Par exemple, cela permet de traiter le cas particulier important de la régression fonctionnelle binomiale, où on a  $Y_i \in \{0, 1\}$  pour  $i = 1, \dots, n$  (voir Müller et Stadtmüller, 2005[50]).
- Un autre modèle qui connaît une grande popularité récente pour des raisons pratiques (et qui est en fait un cas particulier du modèle linéaire fonctionnel généralisé cité ci-dessus) est la version fonctionnelle de la classification, développé, entre autres, dans les travaux de Berlinet, Biau et Rouvière (2005)[3]. Dans ce modèle de classification, la va-

riable réponse  $Y_i$  est cette fois un label associé à la courbe  $X_i$ . Le but est de construire une règle de classification de façon à pouvoir attribuer un label à une nouvelle observation. Berlinet, Biau et Rouvière (2005)[3] proposent ainsi une construction de règle de classification basée sur une décomposition de  $X_1, \dots, X_n$  dans une base d'ondelettes. Ils montrent aussi une certaine forme d'optimalité pour cette règle, prouvant qu'asymptotiquement elle prédit aussi bien que la meilleure règle possible, la règle de Bayes (voir à ce sujet Devroye, Györfi et Lugosi, 1996[26]). Müller et Stadtmüller (2005)[50] ont également abordé ce problème en interprétant le problème de classification (à deux labels) avec variable explicative fonctionnelle comme un cas particulier du modèle linéaire fonctionnel généralisé, avec une variable réponse binaire.

- Enfin, une dernière approche importante consiste à revenir au modèle (1.1) et d'estimer directement l'opérateur  $r$  de façon non paramétrique. Cette approche a été développée par Ferraty et Vieu (2002, 2003)[28] qui donnent un estimateur à noyau de l'opérateur  $r$  et obtiennent là aussi des vitesses de convergence pour cet estimateur. On reviendra un peu plus loin sur cette approche. Ces techniques non paramétriques avec variables fonctionnelles sont récentes et les principaux travaux peuvent être trouvés dans la monographie de Ferraty et Vieu (2006)[29].

Comme cela a déjà été souligné, l'étude de ces divers modèles est motivée au départ par des problèmes pratiques. La variété des domaines dans lesquels les outils de la statistique fonctionnelle interviennent est considérable. La monographie de Ramsay et Silverman (2002)[54] est à elle seule une mine de situations concrètes de données fonctionnelles et de méthodes différentes, allant par exemple de l'étude de la forme d'os déterrés par des archéologues à l'étude de l'enregistrement de l'activité du cerveau lorsqu'on fait prononcer à un individu une syllabe (par électromyographie, c'est-à-dire

en enregistrant avec des électrodes placées à la surface de la peau l'activité électrique provoquée par le mouvement des lèvres et en mesurant l'accélération des lèvres). Il est important de noter que, parmi tous ces cas concrets, se pose souvent le problème du traitement préalable des données. Ramsay et Silverman (2002)[54] mettent notamment en évidence, sur des données relatives à des courbes de croissance, le fait que ces courbes montrent deux types de variabilité, l'amplitude (qui se rapporte à des variations en taille pour des caractéristiques particulières comme le pic de croissance de la puberté) et la phase (qui se rapporte à des variations dans le temps de caractéristiques particulières). Un prétraitement des données doit viser à éliminer la phase de façon à pouvoir concentrer l'étude sur la variation en amplitude. C'est l'objet de travaux tels que les articles de Ramsay et Li (1998)[55], ainsi que Kneip, Li, Mac Gibbon et Ramsay (2000)[45]. Sans vouloir en faire une liste exhaustive, on souhaite citer quelques domaines dans lesquels apparaissent les données fonctionnelles, pour donner une idée du type de problèmes que la statistique fonctionnelle permet de résoudre.

- En biologie, on trouve en premier lieu le travail précurseur de Rao (1958)[52] concernant une étude de courbes de croissance. Plus récemment, un autre exemple est l'étude des variations de l'angle du genou durant la marche (voir Ramsay et Silverman, 2002[54]). Concernant la biologie animale, des études de la ponte de mouches méditerranéennes ont été faites par plusieurs auteurs (Chiou, Müller, Wang et Carey, 2003[17], Chiou, Müller et Wang, 2003[15], Cardot, 2006[18]). Les données consistent en des courbes donnant pour chaque mouche la quantité d'oeufs pondus en fonction du temps.
- La chimiométrie fait aussi partie des champs d'étude propices à l'utilisation de méthodes de la statistique fonctionnelle. Parmi les travaux existants sur le sujet, on peut citer Frank et Friedman (1993)[27] dont on a déjà parlé en début d'introduction, ainsi que Hastie et Mallows (1993)[39]

qui ont commenté l'article de Frank et Friedman (1993)[27] en apportant un exemple de courbes mesurant la log-intensité d'un rayon laser réfracté en fonction de l'angle de réfraction. Plus récemment, Ferraty et Vieu (2002)[28] se sont intéressés à l'étude de la contenance de graisse de morceaux de viande (variable d'intérêt) étant données les courbes d'absorptions de longueurs d'ondes infra-rouge de ces morceaux de viande (variable explicative).

- Des applications liées à l'environnement ont été notamment étudiées par Aneiros-Perez, Cardot, Estevez-Perez et Vieu (2004)[2] qui ont travaillé sur un problème de prévision de pollution. Ces données consistent en des mesures de pics de pollution par l'ozone chaque jour (variable d'intérêt) étant données des courbes de polluants ainsi que de courbes météorologiques de la veille (variables explicatives). Ces données seront également utilisées dans la partie appliquée de ce travail (chapitre 4), et seront alors explicitées en détail à ce moment-là.
- La climatologie est un domaine où les données fonctionnelles apparaissent naturellement. Une étude du phénomène El Niño (courant chaud de l'océan Pacifique) a ainsi été réalisée par Besse, Cardot et Stephenson (2000)[5]. Dans ce travail, les données consistent en des mesures de la température de ce courant en fonction du temps, et la prédiction est faite en utilisant un modèle autorégressif fonctionnel (voir à ce sujet Bosq, 2000[6]).
- En linguistique, des travaux ont également été réalisés, notamment concernant la reconnaissance vocale. On peut citer par exemple Hastie, Buja et Tibshirani (1995)[40], Berlinet, Biau et Rouvière (2005)[3] ou encore Ferraty et Vieu (2003)[28]. Ces travaux sont fortement liés aux méthodes de classification lorsque la variable explicative est une courbe. Brièvement, les données sont des courbes correspondant à des enregistrements de phonèmes prononcés par différents individus. On associe

un label à chaque phonème (variable d'intérêt) et le but est d'établir une classification de ces courbes en utilisant comme variable explicative la courbe enregistrée.

- Dans le domaine de la graphologie, l'apport des techniques de la statistique fonctionnelle a là aussi trouvé une application. Les travaux sur ce problème sont par exemple ceux de Hastie, Buja et Tibshirani (1995)[40] et Ramsay (2000)[54]. Ce dernier modélise par exemple la position du stylo (abscisses et ordonnées en fonction du temps) à l'aide d'équations différentielles.
- Les applications à l'économie sont aussi relativement nombreuses. Des travaux ont notamment été effectués par Kneip et Utikal (2001)[46], et récemment par Benko, Härdle et Kneip (2005)[7], basés notamment sur une analyse en composantes principales fonctionnelle. Cette méthode d'estimation sera analysée lorsqu'on l'utilisera (voir chapitre 3), même si on peut déjà souligner que l'idée de base est, lors de l'estimation de l'opérateur de covariance d'estimer des produits scalaires entre les courbes observées au lieu d'estimer des courbes elles-mêmes.

Dans ce mémoire, nous avons essayé de présenter quelques méthodes d'estimation linéaire pour la fonction de régression pour des données fonctionnelles bruitées, notre travail se divise en quatre chapitre, dans le première chapitre, une introduction générale, on cite en particulier quelque définition et des historiques sur les données fonctionnelles et l'estimation non paramétrique de la régression.

Le but essentiel de ce travail est présenté dans le deuxième et le troisième chapitre, dont on donne deux méthodes d'estimation linéaire le première est la méthode des moindres carrés orthogonaux (cas multivarié et cas fonctionnel), et le deuxième est la méthode de régression fonctionnelle sur composantes principales, on s'intéresse à l'étude des propriétés asymptotiques de

ces deux estimateurs. En suite, une étude de simulation est présentée dans le quatrième chapitre, et on termine ce travail par une conclusion et quelques perspectives.

# Chapitre 2

## La méthode Moindres carrés orthogonaux

### 2.1 Cas multivarié

L'objet de cette section est de donner une description de la méthode des moindres carrés orthogonaux ("Total Least Squares" en anglais, abrégé en TLS) dans le cas où la variable explicative est multivariée, c'est-à-dire élément de  $\mathbb{R}^p$ , et de donner une méthode algorithmique de résolution de ce problème (voir par exemple Golub et Van Loan, 1980[32]). Ces travaux, bien que connus dans ce cadre multivarié, nous seront très utiles dans notre cadre fonctionnel, c'est pourquoi on les rappelle ici. On considère donc que le modèle s'écrit, pour  $i$  allant de 1 à  $n$ ,

$$\begin{cases} Y_i &= \langle \alpha, X_i \rangle + \epsilon_i, \\ W_i &= X_i + \delta_i, \end{cases} \quad (2.1)$$

Avec  $X_i = (X_{i1}, \dots, X_{ip})^\tau$ ,  $W_i = (W_{i1}, \dots, W_{ip})^\tau$  et  $\delta_i = (\delta_{i1}, \dots, \delta_{ip})^\tau$  vecteurs de  $\mathbb{R}^p$ . On doit alors estimer  $\alpha = (\alpha_1, \dots, \alpha_p)^\tau \in \mathbb{R}^p$  sur la base des observations disponibles  $(W_1, Y_1), \dots, (W_n, Y_n)$ . Dans toute la suite, on adopte les notations matricielles suivantes : on note  $X, W$  et  $\delta$  les matrices de taille  $n \times p$  et de termes généraux respectifs  $X_{ij}, W_{ij}$  et  $\delta_{ij}$  pour  $i$  allant de 1 à  $n$  et  $j$  allant de 1 à  $p$ , et on note  $Y = (Y_1, \dots, Y_n)^\tau$  et  $\epsilon = (\epsilon_1, \dots, \epsilon_n)^\tau$ . L'idée des



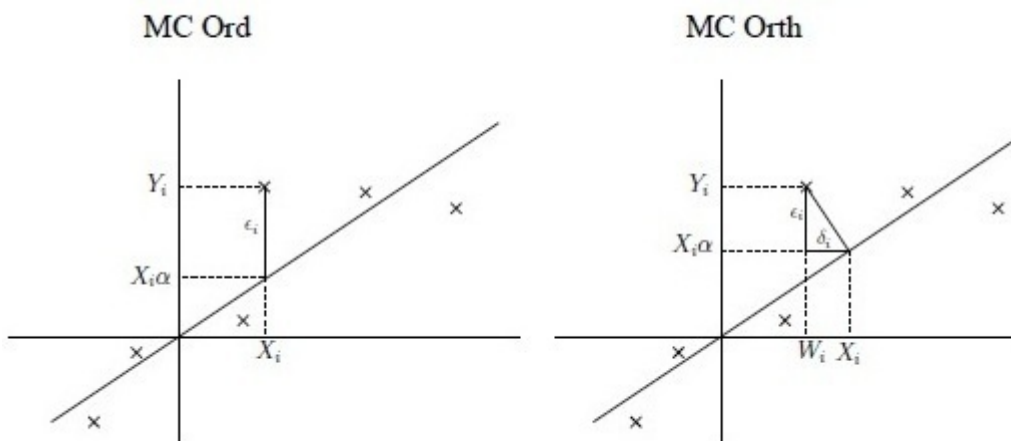
moindres carrés orthogonaux est de rajouter un terme d'erreur sur la variable explicative dans le problème de minimisation, à savoir

$$\frac{1}{n} \sum_{i=1}^n \|W_i - X_i\|^2,$$

pour finalement déterminer simultanément une estimation de  $\alpha$  et  $X_1, \dots, X_n$  en résolvant le problème de minimisation

$$\min_{\alpha \in \mathbb{R}^p, X_i \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^n [(Y_i - X_i^\tau \alpha)^2 + (X_i - W_i)^\tau (X_i - W_i)] \right\}. \quad (2.2)$$

La représentation graphique dans le cas univarié ( $p = 1$ ) permet de justifier l'appellation de moindres carrés orthogonaux. En effet, lorsqu'on résout un problème de moindres carrés ordinaires, on cherche la droite qui minimise la somme des carrés des distances "verticales" des points du nuage jusqu'à la droite. Lorsqu'on résout le problème de minimisation (2.2), on cherche en fait la droite qui minimise la somme des carrés des distances "orthogonales" des points du nuage jusqu'à la droite (voir les figures ci-dessous).



La résolution du problème de minimisation (2.2) est basée sur de l'algèbre linéaire, notamment la décomposition en valeurs singulières d'une matrice rectangulaire (un ouvrage de référence en algèbre matricielle est par exemple celui de Golub et Van Loan, 1996)[33]. Initialement proposé par Golub et Van Loan (1980)[32], cette méthode des moindres carrés orthogonaux a ensuite été reprise dans un ouvrage de Van Huffel et Vandewalle (1991)[59]. Le résultat est le suivant.

**Théorème 2.1.1.** *La solution en  $a \in \mathbb{R}^p$  au problème de minimisation (2.2), notée  $\hat{\alpha}_{TLS}$ , est donnée par*

$$\hat{\alpha}_{TLS} = (W^T W - \sigma_{min}^2 I_p)^{-1} W^T Y, \quad (2.3)$$

où  $I_p$  désigne la matrice identité de taille  $p$  et  $\sigma_{min}^2$  est la plus petite valeur propre non nulle de la matrice  $(W, Y)^T (W, Y)$ , où  $(W, Y)$  désigne la matrice obtenue en concaténant les matrices  $W$  et  $Y$ .

**Remarque 2.1.1.** *On remarque que cette écriture fait apparaître la solution du problème de minimisation comme une version corrigée de l'estimateur par moindres carrés ordinaires. En effet, il apparaît un terme  $-\sigma_{min}^2 I_p$  qui peut être vu comme un terme de "dérégularisation". Celui-ci peut s'interpréter comme un terme de correction qui vise à diminuer le biais induit par la présence de la matrice  $W^T W$  dans l'expression de l'estimateur, au lieu de la matrice  $X^T X$  (non disponible) car on observe  $W$  et non  $X$ . On donne la preuve de ce résultat, l'idée étant reprise ensuite dans notre cadre de variable explicative fonctionnelle.*

**Preuve 2.1.1.** *On introduit la norme de Frobenius d'une matrice  $A$  de terme général  $a_{ij}$  ( $i = 1, \dots, n$  et  $j = 1, \dots, p$ ), notée  $\|\cdot\|_F$  et définie par*

$$\|A\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p a_{ij}^2 = Tr(A^T A).$$

Le problème de minimisation (2.2) s'écrit alors

$$\min_{((W,Y-(\delta,\epsilon)) \begin{pmatrix} \alpha \\ -1 \end{pmatrix})=0} \|(\delta, \epsilon)\|_F^2,$$

ce qui revient à résoudre, en notant  $A = (W, Y)$ ,  $E = (\delta, \epsilon)$  et  $x = (\alpha^\tau, -1)^\tau$

$$\min_{Ax=Ex} \|E\|_F^2.$$

En notant  $\|\cdot\|$  la norme matricielle euclidienne usuelle, on a ainsi

$$\|E\|_F^2 \geq \|E\| = \sup_{x \neq 0} \frac{x^\tau E^\tau E x}{x^\tau x} = \sup_{x \neq 0} \frac{x^\tau A^\tau A x}{x^\tau x}.$$

Or, on remarque que, si on prend

$$E = \frac{Axx^\tau}{x^\tau x},$$

on a bien  $Ax = Ex$  et

$$\|E\|_F^2 = \text{Tr}(E^\tau E) = \frac{xx^\tau A^\tau A xx^\tau}{(x^\tau x)^2} = \frac{x^\tau A^\tau A x}{x^\tau x}.$$

Il ne reste alors plus qu'à minimiser  $\frac{x^\tau A^\tau A x}{x^\tau x}$  en  $x$ . On considère donc la diagonalisation de  $A^\tau A$  et on note  $\sigma_{\min}^2$  la plus petite valeur propre non nulle, associée au vecteur propre noté  $v_{\min}$ . Ainsi, la solution au problème de minimisation est obtenue pour  $x = kv_{\min}$ . La dernière composante (la  $(p+1)^{\text{eme}}$ ) donne la valeur de  $-1/v_{\min_{p+1}}$ . La solution au problème de minimisation (2.2) est donc donnée par

$$\hat{\alpha}_{TLS} = \left( -\frac{1}{v_{\min_{p+1}}} \right) \begin{pmatrix} v_{\min_1} \\ \vdots \\ v_{\min_p} \end{pmatrix} \text{ et } (\hat{\delta}, \hat{\epsilon}) = Av_{\min} v_{\min}^\tau.$$

L'écriture de  $\widehat{\alpha}_{TLS}$  s'obtient alors immédiatement de la façon suivante. Comme

$\begin{pmatrix} \widehat{\alpha}_{TLS} \\ -1 \end{pmatrix}$  est vecteur propre de la matrice  $A^T A$  associé à la valeur propre  $\sigma_{min}^2$ , on a donc

$$\begin{pmatrix} W^T \\ Y^T \end{pmatrix} (W, Y) \begin{pmatrix} \widehat{\alpha}_{TLS} \\ -1 \end{pmatrix} = \sigma_{min}^2 \begin{pmatrix} \widehat{\alpha}_{TLS} \\ -1 \end{pmatrix},$$

ce qui donne (en considérant les  $p$  premières composantes de ce vecteur)

$$W^T W \widehat{\alpha}_{TLS} - W^T Y = \sigma_{min}^2 \widehat{\alpha}_{TLS},$$

et cela achève la preuve du la théorème 2.1.1.

Il est possible, pour faire face à des éventuels problèmes de conditionnement (dû au fait que les valeurs propres de la matrice  $W^T W$  peuvent décroître rapidement vers zéro), de considérer une version régularisée du problème de minimisation (2.2). Ce cas de figure a notamment été envisagé dans un article de Golub, Hansen et O'Leary (1999)[34] qui considèrent le problème de minimisation

$$\min_{a \in \mathbb{R}^p, X_i \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^n [(Y_i - X_i^T a)^2 + (X_i - W_i)(X_i - W_i)^T] + \rho a^T L^T L a \right\}, \quad (2.4)$$

où  $L$  est une matrice de taille  $p \times p$  fixée et  $\rho$  est un paramètre de régularisation qui permet de contrôler le mauvais conditionnement de la matrice  $W$ . Golub, Hansen et O'Leary (1999)[34] montrent alors le résultat suivant, dont la preuve se calque sur celle du la Théorème (2.1.1), en incorporant simplement en plus la régularisation (cette preuve ne sera donc pas donnée ici).

**Théorème 2.1.2.** *La solution en  $a \in \mathbb{R}^p$  au problème de minimisation (2.4), notée  $\widehat{\alpha}_{TLS,pen}$  est donnée par*

$$\widehat{\alpha}_{TLS,pen} = (W^\tau W + \rho L^\tau L - \sigma_{min,pen}^2 I_p)^{-1} W^\tau Y, \quad (2.5)$$

où  $\sigma_{min,pen}^2$  est la plus petite valeur propre non nulle de la matrice

$$(W, Y)^\tau (W, Y) + \begin{pmatrix} L^\tau L & 0 \\ 0 & 0 \end{pmatrix}.$$

Ce cas faisant intervenir une pénalisation est important dans notre contexte fonctionnel, où, comme cela a été souligné, l'apport d'une pénalisation s'avère fondamental. Ce dernier résultat permettra ainsi la généralisation de la méthode des moindres carrés orthogonaux au cas d'une variable explicative fonctionnelle.

## 2.2 Cas fonctionnel

Dans cette section, le but est de généraliser la méthode des moindres carrés orthogonaux à notre cadre fonctionnel. On souhaite proposer un estimateur spline, qui étant obtenu en modifiant l'estimateur par splines de régression introduit par Cardot, Ferraty et Sarda (1999, 2003)[13]. On va d'abord s'attacher à donner la méthode de construction de l'estimateur par splines de régression, qui a été étudié le premier chronologiquement (voir Crambes, 2005[19]). On cherche un estimateur  $\widehat{\alpha} = B_{k,q}^\tau \widehat{\theta}$  avec  $\theta \in \mathbb{B}^{k+q}$  solution du problème de minimisation donnée par

$$\min_{\theta \in \mathbb{B}^{k+q}} \left\{ \frac{1}{n} \sum_{i=1}^n (Y_i - \langle B_{k,q}^\tau \theta, X_i \rangle)^2 + \rho \|(B_{k,q}^\tau \theta)^{(m)}\|_{L^2}^2 \right\}.$$

Ce problème de minimisation admet une solution explicite. Celle-ci est donnée par

$$\widehat{\theta}_{FLS,X} = \frac{1}{n} \left( \frac{1}{n} D_X^\tau D_X + \rho G_k \right)^{-1} D_X^\tau Y, \quad (2.6)$$

avec

$$D_X = \begin{pmatrix} \langle B_1, X_1 \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, X_1 \rangle \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \langle B_1, X_n \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, X_n \rangle \end{pmatrix},$$

et

$$G_k = \begin{pmatrix} \langle B_1^{(m)}, B_1^{(m)} \rangle & \cdot & \cdot & \cdot & \langle B_1^{(m)}, B_{k+q}^{(m)} \rangle \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \langle B_{k+q}^{(m)}, B_1^{(m)} \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}^{(m)}, B_{k+q}^{(m)} \rangle \end{pmatrix}.$$

### 2.2.1 Construction de l'estimateur splines de régression

Ici, les courbes  $X_1, \dots, X_n$  ne sont pas connues, les courbes réellement observées  $W_1, \dots, W_n$  sont définies par

$$W_i(t_j) = X_i(t_j) + \delta_{ij}.$$

Pour étendre la méthode des moindres carrés orthogonaux à ce contexte, on va donc considérer le problème de minimisation

$$\min_{\theta \in \mathbb{R}^{k+q}, X_i \in L^2(I)} \left\{ \frac{1}{n} \sum_{i=1}^n \left[ (Y_i - \langle B_{k,q}^T \theta, X_i \rangle)^2 + \|\widetilde{X}_i - \widetilde{W}_i\|^2 \right] + \rho \|(B_{k,q}^T \theta)^{(m)}\|_{L^2}^2 \right\}, \quad (2.7)$$

où  $\widetilde{X}_i$  et  $\widetilde{W}_i$  sont les versions splines de  $X_i$  et  $W_i$ . Plus précisément, en notant  $X$  la matrice  $n \times p$  de terme général  $(X_i(t_j))$  pour  $i = 1, \dots, n$  et  $j = 1, \dots, p$  et si  $\beta$  est la matrice  $p \times (k+q)$  de terme général  $B_r(t_j)$  pour  $j = 1, \dots, p$  et  $r = 1, \dots, k+q$ , alors  $\widetilde{X}_i$  est la matrice  $n \times (k+q)$  définie par

$$\widetilde{X} = X\beta,$$

et  $\widetilde{X}_i$  est la  $i$ ème ligne de  $\widetilde{X}$ . Maintenant, en utilisant une technique analogue à celle présentée dans le cas multivarié, on montre le résultat suivant.

**Théorème 2.2.1.** *La solution en  $\theta \in \mathbb{R}^{k+q}$  du problème de minimisation (2.7), notée  $\widehat{\theta}_{FTLS}$  (FTLS pour "Functional Total Least Squares") est donnée par*

$$\widehat{\theta}_{FTLS} = \frac{1}{n} \left( \frac{1}{n} D_W^\tau D_W + \rho G_k - \sigma_{\min}^2 B_k \right)^{-1} D_W^\tau Y, \quad (2.8)$$

où les matrices  $D_W$  et  $B_k$  sont définies par

$$D_W = \begin{pmatrix} \langle B_1, W_1 \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, W_1 \rangle \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \langle B_1, W_n \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, W_n \rangle \end{pmatrix},$$

$$B_k = \begin{pmatrix} \langle B_1, B_1 \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, B_1 \rangle \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \langle B_1, B_{k+q} \rangle & \cdot & \cdot & \cdot & \langle B_{k+q}, B_{k+q} \rangle \end{pmatrix},$$

et  $\sigma_{\min}^2$  est la plus petite valeur propre de la matrice

$$\frac{1}{n} \left( \frac{D_W}{\sqrt{p}}, Y \right)^\tau \left( \frac{D_W}{\sqrt{p}}, Y \right) + \gamma (\gamma^\tau \gamma)^{-1} (\rho K_k) (\gamma^\tau \gamma)^{-1} \gamma^\tau,$$

avec  $\gamma$  matrice  $(p+1) \times (k+q+1)$  donnée par

$$\gamma = \begin{pmatrix} \frac{\beta}{\sqrt{p}} & 0 \\ 0 & 0 \end{pmatrix},$$

et  $K_k$  matrice  $(k+q+1) \times (k+q+1)$  donnée par

$$K_k = \begin{pmatrix} G_k & 0 \\ 0 & 0 \end{pmatrix}.$$

**Preuve 2.2.1.** La solution du problème de minimisation (2.7) et donnée par

$$\hat{\theta}_{FTLS} = \frac{1}{n} \left( \frac{1}{n} D_W^\tau D_W + \rho G_k - \sigma_{\min}^2 B_k \right)^{-1} D_W^\tau Y.$$

En utilisant la matrice  $\beta$ , on écrit  $\alpha = \beta\theta$  avec  $\theta \in \mathbb{R}^{k+q}$ . On tire alors de l'écriture du modèle

$$\left( \left( \frac{W\beta}{p}, Y \right) - \left( \frac{\delta\beta}{p}, \epsilon \right) \right) \begin{pmatrix} \theta \\ -1 \end{pmatrix} = 0, \quad (2.9)$$

ce qui nous permet d'écrire le problème de minimisation (2.7) sous la forme

$$\min_{(2.9)} \left\{ \frac{1}{n} \left\| \begin{pmatrix} \delta \\ \sqrt{p} \end{pmatrix}, \epsilon \right\|_F^2 + \rho \theta^\tau G_k \theta \right\},$$

où la notation  $\|\cdot\|_F$  désigne toujours la norme de Frobenius matricielle. On considère alors le problème de minimisation

$$\min_{Ax=Ex} \left\{ \frac{1}{n} \left\| \begin{pmatrix} \delta \\ \sqrt{p} \end{pmatrix}, \epsilon \right\|_F^2 + \rho x^\tau K_k x \right\}, \quad (2.10)$$

avec  $A = \left( \frac{W\beta}{p}, Y \right)$ ,  $E = \left( \frac{\delta\beta}{p}, \epsilon \right)$  et  $x = \begin{pmatrix} \theta \\ -1 \end{pmatrix}$ . En notant que

$$\begin{aligned} \frac{1}{n} x^\tau \gamma^\tau \begin{pmatrix} \delta \\ \sqrt{p} \end{pmatrix}, \epsilon \begin{pmatrix} \delta \\ \sqrt{p} \end{pmatrix}, \epsilon \gamma x &= \frac{1}{n} x^\tau E^\tau E x \\ &= \frac{1}{n} x^\tau A^\tau A x \\ &= \frac{1}{n} x^\tau \gamma^\tau \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \gamma x, \end{aligned}$$

on voit que la quantité

$$\begin{aligned} & \frac{1}{n} x^\tau \gamma^\tau \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \gamma x + \rho x^\tau K_k x \\ &= \frac{1}{n} x^\tau \gamma^\tau \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \begin{pmatrix} W \\ \sqrt{p} \end{pmatrix}, Y \gamma x + x^\tau \gamma^\tau \gamma (\gamma^\tau \gamma)^{-1} (\rho K_k) (\gamma^\tau \gamma)^{-1} \gamma^\tau \gamma x \end{aligned}$$



est minimisée pour  $\gamma x$  vecteur propre de la matrice

$$\frac{1}{n} \left( \frac{W}{\sqrt{p}}, Y \right)^\tau \left( \frac{W}{\sqrt{p}}, Y \right) + \gamma (\gamma^\tau \gamma)^{-1} (\rho K_k) (\gamma^\tau \gamma)^{-1} \gamma^\tau,$$

correspondant à la plus petite valeur propre non nulle, notée  $\sigma_{\min}^2$ . En utilisant la définition de cette valeur propre, on déduit que

$$\left( \frac{1}{n} \left( \frac{W}{\sqrt{p}}, Y \right)^\tau \left( \frac{W}{\sqrt{p}}, Y \right) + \gamma (\gamma^\tau \gamma)^{-1} (\rho K_k) (\gamma^\tau \gamma)^{-1} \gamma^\tau \right) \gamma \hat{x} = \sigma_{\min}^2 \gamma \hat{x},$$

ce qui donne, en prémultipliant par  $\gamma^\tau$ ,

$$\frac{1}{n} \left( \frac{W\beta}{p}, Y \right)^\tau \left( \frac{W\beta}{p}, Y \right) \begin{pmatrix} \hat{\theta} \\ -1 \end{pmatrix} = \sigma_{\min}^2 \begin{pmatrix} \frac{\beta^\tau \beta \hat{\theta}}{p} \\ -1 \end{pmatrix}.$$

Finalement, en gardant les  $k + q$  première lignes, on obtient

$$\hat{\theta} = \frac{1}{n} \left( \frac{1}{n} D_W^\tau D_W + \rho G_k - \sigma_{\min}^2 B_k \right)^{-1} D_W^\tau Y,$$

ce qui achève la preuve du la théorème (2.2.1).

**Théorème 2.2.2.** *Sous l'hypothèse suivante.*

(H.0) les variables  $X_i$  vérifient (p.s.)

$$\sup_{i=1, \dots, n} \sup_{t \in [0,1]} |X_i(t)| \leq c_0,$$

où  $c_0$  ne dépend pas de  $n$ . On a alors

$$\frac{1}{n} D_W^\tau D_W = \frac{1}{n} D_X^\tau D_X + \frac{\sigma_\delta^2}{p} B_k + R_1, \quad (2.11)$$

où  $R_1$  est une matrice vérifiant

$$\|R_1\| = o_{\mathbb{P}} \left( \frac{1}{n^{1/2} p^{1/2} k^{1/2}} \right).$$

Enfin, comme  $\sigma_\delta^2$  n'est pas toujours connu, on peut l'estimer. On choisit ici de l'estimer nonparamétriquement, en utilisant les travaux de Gasser, Sroka et Jennen-Steinmetz (1986)[35]. Comme on est en présence de points de mesure équidistants, cet estimateur est donné par

$$\widehat{\sigma}_\delta^2 = \frac{1}{n} \sum_{i=1}^n \frac{1}{6(p-2)} \sum_{j=1}^{p-1} [W_i(t_{j-1}) - W_i(t_j) + W_i(t_{j+1}) - W_i(t_j)]^2. \quad (2.12)$$

Finalement, l'estimateur de  $\theta$  est donné par

$$\widehat{\theta}_{FTLS} = \frac{1}{n} \left( \frac{1}{n} D_W^\tau D_W + \rho G_k - \frac{\widehat{\sigma}_\delta^2}{p} B_k \right)^{-1} D_W^\tau Y, \quad (2.13)$$

et l'estimateur de  $\alpha$  est donné par

$$\widehat{\alpha}_{FTLS,k} = B_{k,q}^\tau \widehat{\theta}_{FTLS}.$$

**Preuve 2.2.2.** On a

$$\frac{1}{n} D_W^\tau D_W = \frac{1}{n} D_X^\tau D_X + \frac{\sigma_\delta^2}{p} B_k + R_1,$$

avec

$$\|R_1\| = O_{\mathbb{P}} \left( \frac{1}{n^{1/2} p^{1/2} k^{1/2}} \right).$$

En utilisant le fait que  $W_i(t_j) = X_i(t_j) + \delta_{ij}$  pour  $i = 1, \dots, n$  et  $j = 1, \dots, p$ , on peut écrire

$$\frac{1}{n} D_W^\tau D_W = \frac{1}{n} D_X^\tau D_X + \left( \frac{1}{n} \sum_{i=1}^n M_{irs} \right)_{r,s=1,\dots,k+q},$$

avec  $M_{irs} = \langle B_r, X_i \rangle \langle B_s, \delta_i \rangle + \langle B_r, \delta_i \rangle \langle B_s, X_i \rangle + \langle B_r, \delta_i \rangle \langle B_s, \delta_i \rangle$ . Étudions maintenant cette variable aléatoire  $M_{irs}$ . Tout d'abord, en utilisant l'indépendance entre  $X_i$  et  $\delta_i$ , on peut écrire

$$\begin{aligned}
\mathbb{E}(M_{irs}) &= \mathbb{E}(\langle B_r, \delta_i \rangle \langle B_s, \delta_i \rangle) \\
&= \frac{1}{p^2} \sum_{j_1=1}^p \sum_{j_2=1}^p B_r(t_{j_1}) B_s(t_{j_2}) \mathbb{E}(\delta_i(t_{j_1}) \delta_i(t_{j_2})) \\
&= \frac{\sigma_{\delta}^2}{p^2} \sum_{j=1}^p B_r(t_j) B_s(t_j) \\
&= \frac{\sigma_{\delta}^2}{p} \langle B_r, B_s \rangle.
\end{aligned} \tag{2.14}$$

D'autre part, on a

$$\begin{aligned}
\mathbb{E}(M_{irs}^2) &= \mathbb{E}(\langle B_r, X_i \rangle^2) \mathbb{E}(\langle B_s, \delta_i \rangle^2) \\
&\quad + \mathbb{E}(\langle B_r, \delta_i \rangle^2) \mathbb{E}(\langle B_s, X_i \rangle^2) \\
&\quad + \mathbb{E}(\langle B_r, \delta_i \rangle^2 \langle B_s, \delta_i \rangle^2) \\
&\quad + 2\mathbb{E}(\langle B_r, X_i \rangle \langle B_s, X_i \rangle) \mathbb{E}(\langle B_r, \delta_i \rangle \langle B_s, \delta_i \rangle).
\end{aligned} \tag{2.15}$$

En utilisant des résultats techniques sur les B-splines (voir Cardot, 2000[20]), on note que

$$|\mathbb{E}(\langle B_r, \delta_i \rangle \langle B_s, \delta_i \rangle)| = \left| \frac{\sigma_{\delta_i}^2}{p^2} \sum_{j=1}^p B_r(t_j) B_s(t_j) \right| = O\left(\frac{1}{p^k}\right), \tag{2.16}$$

et, avec l'hypothèse (H.0),

$$|\mathbb{E}(\langle B_r, X_i \rangle \langle B_s, X_i \rangle)| = O\left(\frac{1}{k^2}\right). \tag{2.17}$$

Avec le même type de calculs, on a aussi

$$\begin{aligned}
& \mathbb{E}(\langle B_r, \delta_i \rangle^2 \langle B_s, \delta_i \rangle^2) \\
&= \frac{1}{p^4} \sum_{j_1=1}^p \sum_{j_2=1}^p \sum_{j_3=1}^p \sum_{j_4=1}^p B_r(t_{j_1}) B_r(t_{j_2}) B_s(t_{j_3}) B_s(t_{j_4}) \mathbb{E}(\delta_{ij_1} \delta_{ij_2} \delta_{ij_3} \delta_{ij_4}) \\
&= \frac{1}{p^4} \sum_{j=1}^p B_r(t_j)^2 B_s(t_j)^2 \mathbb{E}(\delta_{ij}^4) \\
&\quad + \frac{1}{p^4} \sum_{j_1=1}^p \sum_{j_2=1}^p B_r(t_{j_1}) B_r(t_{j_2}) B_s(t_{j_1}) B_s(t_{j_2}) \mathbb{E}(\delta_{ij_1}^2) \mathbb{E}(\delta_{ij_2}^2),
\end{aligned}$$

d'où

$$\mathbb{E}(\langle B_r, \delta_i \rangle^2 \langle B_s, \delta_i \rangle^2) = O\left(\frac{1}{p^4} \left[ \sum_{j=1}^p B_r(t_j) B_s(t_j) \right]^2\right),$$

ce qui donne finalement

$$\mathbb{E}(\langle B_r, \delta_i \rangle^2 \langle B_s, \delta_i \rangle^2) = O\left(\frac{1}{p^2 k^2}\right). \quad (2.18)$$

Maintenant, avec (2.16), (2.17) et (2.18), la relation (2.15) devient

$$\mathbb{E}(M_{irs}^2) = O\left(\frac{1}{pk^3}\right) + O\left(\frac{1}{p^2 k^2}\right),$$

soit, en prenant  $p > k$

$$\mathbb{E}(M_{irs}^2) = O\left(\frac{1}{pk^3}\right). \quad (2.19)$$

On peut maintenant conclure la preuve du la théorème 2.2.2. En utilisant (2.14) et (2.19), on a donc

$$\frac{1}{n} \sum_{i=1}^n M_{irs} = \frac{\sigma_\delta^2}{p} \langle B_r, B_s \rangle + o_{\mathbb{P}}\left(\frac{1}{n^{1/2} p^{1/2} k^{3/2}}\right).$$

C'est donc là qu'apparaissent les matrices  $B_k$  et  $R_1$  : il existe une matrice  $R_1$  telle que

$$\left( \frac{1}{n} \sum_{i=1}^n M_{irs} \right)_{r,s=1,\dots,k+q} = \frac{\sigma_\delta^2}{p} B_k + R_1,$$

avec, pour  $r, s = 1, \dots, k+q$ ,  $R_{1rs} = o_{\mathbb{P}} \left( \frac{1}{n^{1/2} p^{1/2} k^{3/2}} \right)$ , soit, avec le théorème 1.19 de Chatelin (1983)[21],

$$\|R_1\| = O_{\mathbb{P}} \left( \frac{1}{n^{1/2} p^{1/2} k^{1/2}} \right),$$

ce qui termine la preuve du la théorème 2.2.2.

## 2.2.2 Résultat de convergence

Le résultat de convergence donne une borne supérieure pour la vitesse de convergence de  $\widehat{\alpha}_{FTLS}$  vers  $\alpha$  au sens de la semi-norme induite par l'opérateur de covariance  $\Gamma_X$ . On supposera que  $k = k_n \rightarrow +\infty$  et  $\rho = \rho_n \rightarrow 0$  quand  $n \rightarrow +\infty$ . On fait également tendre  $p$  vers l'infini. La preuve de ce résultat sera basée sur une décomposition de la forme  $\widehat{\alpha}_{FTLS} - \alpha = \widehat{\alpha}_{FTLS} - \widehat{\alpha} + \widehat{\alpha} - \alpha$ , où  $\widehat{\alpha}$  est l'estimateur de  $\alpha$  par splines de régression introduit par Cardot, Ferraty et Sarda (1999, 2003)[13], supposant que les courbes  $X_1, \dots, X_n$  sont directement accessibles et non bruitées. Comme ces derniers ont déjà établi un résultat de convergence concernant  $\widehat{\alpha}$ , il faudra évaluer l'écart entre cet estimateur et celui par moindres carrés orthogonaux. Ainsi, pour établir notre résultat de convergence, on aura besoin des hypothèses faites par Cardot, Ferraty et Sarda (2003)[13] assurant la convergence de leur estimateur.

### Hypothèses

(H.1) La variable  $X$  vérifie

$$\|X\| \leq c_1 < \infty \text{ p.s.}$$

(H.2) La fonction  $\alpha$  admet une dérivée d'ordre  $p'$  et  $\alpha^{(p')}$  vérifie

$$\left| \alpha^{(p')}(t) - \alpha^{(p')}(s) \right| \leq c_2 |t - s|^v,$$

pour  $s, t \in [0, 1]$ , où  $c_2 > 0$  et  $v \in [0, 1]$ . Dans ce qui suit, on pose  $d = p' + v$  et on suppose que  $q \geq d \geq m$ .

(H.3) Les valeurs propres de  $\Gamma_X$  sont strictement positives.

On fait aussi les hypothèses suivantes qui nous permettront de contrôler la vitesse de l'écart entre  $\widehat{\alpha}_{FTLS}$  et  $\widehat{\alpha}$ .

(H.4) Les variables  $\delta_{ij}$  vérifient

$$\sup_{i=1, \dots, n} \sup_{j=1, \dots, p} \mathbb{E}(\delta_{ij}^4) \leq c_3.$$

où  $c_3$  ne dépend pas de  $n$  et de  $p$ .

(H.5) Les variables  $Y_i$  et  $\delta_{ij}$  sont indépendantes pour tout  $i = 1, \dots, n$  et  $j = 1, \dots, p$  et il existe une constante  $c_4 > 0$  indépendante de  $n$  telle que  $\sup_{i=1, \dots, n} \mathbb{E}(Y_i^2) \leq c_4$ .

On a alors le résultat suivant.

**Théorème 2.2.3.** *Sous les hypothèses qui précèdent, en supposant de plus que  $1/p = o(\rho_n/k_n)$ , qu'il existe une constante  $c_5 > 0$  indépendante de  $n$  telle que  $k_n^{1/2} \|(\langle \Gamma_{X,n} \alpha, B_j \rangle)_{j=1, \dots, k_n+q}\| \geq c_5$  et qu'il existe  $\beta, \gamma \in ]0, 1[$  tels que  $k_n \sim n^\beta$ ,  $\rho_n \sim n^{-(1-\gamma)/2}$ , on a*

$$\|\widehat{\alpha}_{FTLS, k_n} - \alpha\|_{\Gamma_X}^2 = o_{\mathbb{P}} \left( \frac{1}{k_n^{2d}} + \frac{k_n}{n\rho_n} + \rho_n + \frac{k_n}{n p \rho_n} \right).$$

**Remarque 2.2.1.** *La preuve ne sera pas donnée ici. Elle est simplement basée sur la comparaison entre  $\widehat{\alpha}_{FTLS, k_n}$  et  $\widehat{\alpha}$  et entre  $\widehat{\alpha}$  et  $\alpha$ . Un résultat de convergence a été déjà obtenu par Cardot, Ferraty et Sarda (2003)[13]*

concernant  $\|\widehat{\alpha} - \alpha\|_{\Gamma_X}$ . Plus précisément, sous les hypothèses (H.1) - (H.3), dès que  $1/p = o(\rho_n/k_n)$ , on a

$$\|\widehat{\alpha} - \alpha\|_{\Gamma_X}^2 = o_{\mathbb{P}} \left( \frac{1}{k_n^{2d}} + \frac{k_n}{n\rho_n} + \rho_n \right).$$

Pour prouver la théorème 2.2.2, on montre finalement que

$$\|\widehat{\alpha}_{FTLS, k_n} - \widehat{\alpha}\|_{\Gamma_X}^2 = o_{\mathbb{P}} \left( \frac{k_n}{n p \rho_n} \right).$$

### 2.2.3 Commentaires

**Vitesse de convergence.** On remarque que  $k_n/(n p \rho_n)$  sera négligeable par rapport à  $k_n/(n \rho_n)$  si  $p$  est assez grand. Cela signifie qu'à partir du moment où le nombre de points de mesure est suffisamment grand, l'effet du bruit est négligeable. Sous cette hypothèse, une vitesse optimale peut être trouvée en choisissant un  $\rho_n$  et un  $k_n$  particuliers (voir Cardot, Ferraty et Sarda, 2003[13]). En prenant  $\rho_n \sim n^{-2d/(4d+1)}$  et  $k_n \sim n^{1/(4d+1)}$ , on a alors

$$\|\widehat{\alpha}_{FTLS, k_n} - \alpha\|_{\Gamma_X}^2 = o_{\mathbb{P}} \left( n^{-2d/(4d+1)} \right).$$

**Effet de la dérégularisation.** Regardons ce qui se passe si on estime  $\theta$  sans faire intervenir de dérégularisation, mais en utilisant les courbes disponibles  $W_1, \dots, W_n$ . Plus précisément, on a  $\widehat{\alpha}_W = B_{k,q}^{\tau} \widehat{\theta}_W$  avec

$$\widehat{\theta}_W = \frac{1}{n} \left( \frac{1}{n} D_W^{\tau} D_W + \rho G_k \right)^{-1} D_W^{\tau} Y.$$

Alors, avec des arguments comparables à ceux utilisés pour la preuve du résultat précédent, on obtient, si  $p$  est assez grand, la même vitesse qu'avec l'estimateur par moindres carrés orthogonaux. Dans ce cas, la correction induite par la méthode des moindres carrés orthogonaux ne semble pas avoir un impact fondamental sur la vitesse de convergence.

## Chapitre 3

# La méthode de régression fonctionnelle sur composantes principales

Dans ce cadre d'une variable explicative bruitée, ce qui précédait visait à généraliser la méthode des moindres carrés orthogonaux au cas d'une variable explicative fonctionnelle. Dans cet chapitre, on va présenter une autre approche. Cette approche est finalement au départ plus directe que les moindres carrés orthogonaux (qui traite globalement les courbes bruitées). L'idée de départ est d'effectuer un lissage (par exemple un lissage à noyau) de chaque courbe bruitée, puis de produire une estimation du paramètre fonctionnel à l'aide par exemple d'une régression sur composantes principales. Rappelons que dans cette partie, le modèle considéré est donné par (2.1) et que le but est de donner une méthode d'estimation de  $\alpha$  à l'aide des observations  $(W_1, Y_1), \dots, (W_n, Y_n)$ .

### 3.1 L'ACP fonctionnelle pour la régression

L'ACP fonctionnelle est un domaine fondamental de la statistique fonctionnelle, elle s'intéresse à l'étude de la covariance d'un ensemble des données fonctionnelles, elle la représente dans un espace fini sous forme linéaire optimal. En pratique, on dispose de  $n$  observations de  $k$  variables fonctionnelles



sur  $n$  individus et on plaçons ces données  $x_{ij}$  dans un tableau  $X$  rectangulaire de dimension  $n \times p$

$$X = \begin{pmatrix} x_{11} & \cdot & \cdot & \cdot & x_{1p} \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ \cdot & & & & \cdot \\ x_{n1} & \cdot & \cdot & \cdot & x_{np} \end{pmatrix}.$$

Les auteurs qui sont intéressés par ce sujet se sont : Dauxois et Pousse (1976)[24] . Koenker et Basseett (1978)[47]. Breiman, Friedman, Olshen et Stone (1984)[8]. Dauxois, Pousse et Romain (1982)[25]. Besse et Ramsay (1986)[9]. Ramsay et Dalzell (1991)[56] . les articles de Ramsay et Silverman (2002,2005)[54] Ainsi que Ramsay et Silverman (1997)[54] étudient un exemple du climat dans un pays donné par l'intermédiaire des variations de la température mensuelle moyenne au cours de l'année en différents points de pays. Chaque station météo est alors un individu qui est décrit par une fonction qui au mois associe la température moyenne observée. Besse et Cardot (2000)[10] étudient l'évolution des diverses mesures géophysique sur une zone donnée pendant un an. Chaque année d'observations correspond à un individu et ils ont étudié la série temporelle des individus. Cardot (2000,2006)[20] étudie l'analyse en composantes principales fonctionnelle conditionnelle. Le principe de l'ACP fonctionnelle est la projection orthogonale d'un nuage de  $n$  points fonctionnels  $x_1, x_2, \dots, x_n$  dans  $L^2([0, 1])$  sur un espace vectorielle fini engendré par  $q$ , ( $q < p$ ) fonctions de  $L^2([0, 1])$   $\psi_1, \psi_2 \dots, \psi_q$ , sans perte aucune information, ces vecteurs vérifient que :

$$\xi(\psi_1, \psi_2 \dots, \psi_q) = \sum_{i=1}^q \|x_i - \sum_{i=1}^q \alpha_i(x_i) \psi_i\|_2^2,$$

soit minimale,

- $\|\cdot\|_2$  désigne la norme de  $L^2$ ,

- $\alpha_i(x_i) = \langle x_i, \psi_i \rangle$  est le produit scalaire dans  $L^2$  qui est la coordonnée de la projection de  $x_i$  sur l'espace engendré par  $\psi_i$ .

La recherche de ces vecteurs revient à faire une analyse spectrale de l'opérateur de la covariance suivant :

$$\Sigma(s, t) = \frac{1}{n} \sum_{i=1}^n (x_i(s) - \mu(s))(x_i(t) - \mu(t)),$$

où  $\mu$  est la fonction moyenne de  $x_{i=1, \dots, n}$ . L'ACP des  $X$  cherche les valeurs propres de l'opérateur défini par :

$$\Gamma(\psi) = \langle \Sigma, \psi \rangle.$$

La fonction propre  $\psi_1$  associée à la plus grande valeur propre  $\lambda_1$  est la solution maximale du problème suivant :

$$\begin{cases} \max \langle \Gamma(\psi), \psi \rangle \\ \|\psi\|_2 = 1 \end{cases}$$

**Théorème 3.1.1.** (*La décomposition Karhunen-Loeve*)

$$X_t = \mathbb{E}(X_t) + \sum_{k=1}^{\infty} \xi_k \psi_k(t)$$

$\xi_k$  v.a indépendant tel que  $\mathbb{E}(\xi_k) = 0$  et  $\mathbb{V}(\psi_k) = \lambda_k$ , donc

$$X_t \simeq X_k = \mathbb{E}(X_t) + \sum_{k=1}^K \xi_k \psi_k(t),$$

avec  $\frac{\mathbb{V}(X_k)}{\mathbb{V}(X_k)} \simeq 1$

### 3.1.1 La régression sur composantes principales

Les méthodes classiques de l'estimation (moindre carrée) de la fonction de la régression sont basées sur l'inversion de la matrice de variance-covariance

$\Gamma_X = XX^t$ . Ces méthodes ne sont pas applicables dans le cas fonctionnel. Parmi les solutions qui ont été proposées pour résoudre ce problème, la régression sur composantes principales fonctionnelles. Cette approche a pour but de réduire la dimension de  $X$ , en travaillant par les  $k$  premières composantes principales d'une ACP de  $X$ . les premiers travaux dans ce domaine sont les travaux de Ramsay (1982)[53], les travaux de Breiman, Freidman, Olshen et Stone (1984)[8]. Ferraty et Sarda (1999,2003)[13], les travaux de Kneip et Utikal (2001)[46]. Benko, Müller et Stadtmüller (2005)[11]. Härd et Kneip (2005)[7] ainsi que les travaux de Yao, Müller et Wang (2005)[62].

L'idée de base de cette méthode est la diagonalisation de la matrice de la variance-covariance empirique  $\Gamma_{(X,n)}$  associée à  $X$ , pour ce fait, on note par  $(\psi_s)_{s \geq 1}$  les fonctions propres de  $\Gamma_{(X,n)}$  associées aux valeurs propres  $(\lambda_s)_{s \geq 1}$  (ordonnées au sens décroissant)

$$\Gamma_{X,n}\psi_s = \lambda_s\psi_s.$$

On introduisant les coefficients (les composants principaux) :

$$\eta_{is} = \langle X_i, \psi_s \rangle \quad i = 1, \dots, n, \quad s \geq 1.$$

Pour déterminer les fonctions propres  $(\psi_s)_{s \geq 1}$  de  $\Gamma_{(X,n)}$ , il faut pas passer par  $\Gamma_{(X,n)}$  d'après Kneip et Utikal (2001)[46], ils ont utilisé une autre matrice  $M$  carrée de dimension  $n \times n$  définie par

$$M = \left( \frac{1}{n} \langle X_{i_1}, X_{i_2} \rangle \right)_{(i=1, \dots, n)}.$$

Ils ont montré que les valeurs propres de  $\Gamma_{(X,n)}$  et de  $M$  sont les mêmes. De plus, si on note par  $P_s = (p_{1s}, \dots, p_{ns})^\tau$  le vecteur propre de  $M$  associé à la valeur propre  $\lambda_s$  on a

$$\eta_{is} = \sqrt{\lambda_s} p_{is}.$$

### 3.1.2 Régression sur composantes principales pour les variables sont bruitées 36

Pour tout  $i = 1, \dots, n$  et  $s \geq 1$  tel que  $\lambda_s > 0$ , on obtient la fonction propre :

$$\psi_s = \frac{1}{\sqrt{\lambda_s}} \sum_{i=1}^n p_{is} X_i = \frac{\sum_{i=1}^n \eta_{is} X_i}{\eta_{is}^2} \quad s \geq 1.$$

Finalement, on construit l'estimateur de  $r$  donné par l'approximation d'ordre  $k \geq 1$ , en utilisant les  $k$  premières composantes principales, par :

$$\hat{r}_k = \frac{1}{n} \sum_{s=1}^k \sum_{i=1}^n \frac{Y_i}{\lambda_s} \langle X_i, \psi_s \rangle \psi_s.$$

### 3.1.2 Régression sur composantes principales pour les variables sont bruitées

L'idée de cette procédure de lissage est donc la suivante. Lorsqu'on doit prendre en compte des courbes bruitées  $W_1, \dots, W_n$ , on les lisse dans le but de construire une estimation des "vraies" courbes  $X_1, \dots, X_n$ . On peut alors utiliser ces nouvelles courbes lissées  $\widetilde{W}_1, \dots, \widetilde{W}_n$  pour produire un estimateur de  $\alpha$  au moyen d'une régression sur composantes principales fonctionnelle (voir les travaux de Cardot, Ferraty et Sarda, 1999, 2003[13]). Cette régression sur composantes principales fonctionnelle sera néanmoins adaptée, reprenant une idée utilisée par Kneip et Utikal (2001)[46] ainsi que Benko, Härdle et Kneip (2005)[7]. La procédure d'estimation comprend ainsi deux étapes qui vont être détaillées dans ce qui suit.

**Étape 1 : lissage des courbes bruitées.** Cette première étape consiste à lisser les observations bruitées  $W_1, \dots, W_n$  à l'aide d'un estimateur à noyau de type Nadaraya-Watson, introduit à l'origine simultanément par Nadaraya (1964) et Watson (1964)[51]. Pour plus de détails sur cet estimateur à noyau tant d'un point de vue théorique qu'appliqué, on renvoie à Härdle (1991)[41]

ou Sarda et Vieu (2000)[58]. Plus précisément, on définit, pour  $i = 1, \dots, n$  et pour  $t \in [0, 1]$ ,

$$\widetilde{W}_i(t) = \frac{\sum_{j=1}^p W_i(t_j) K\left(\frac{t-t_j}{h_i}\right)}{\sum_{j=1}^p K\left(\frac{t-t_j}{h_i}\right)}. \quad (3.1)$$

Dans cette expression (3.1), la fonction  $K$ , paire et d'intégrale égale à 1, est appelée noyau et le nombre réel  $h_i > 0$  est appelé largeur de fenêtre. C'est le paramètre qui permet de contrôler le lissage de la courbe estimée  $\widetilde{W}_i$ , alors que le choix du noyau est moins fondamental, si ce n'est que l'estimateur construit hérite des propriétés de régularité du noyau choisi (continuité, dérivabilité, . . . ). Le choix de la largeur de fenêtre étant très important, de nombreux travaux ont été réalisés pour permettre de déterminer ce paramètre en pratique, comme par exemple la validation croisée (voir Härdle ,1991[41]).

**Étape 2 :régression fonctionnelle sur composantes principales**

Cette étape consiste à construire une estimation de  $\alpha$  au moyen d'une régression sur composantes principales fonctionnelle (voir Cardot, Ferraty et Sarda, 1999, 2003[13]) en utilisant comme variable explicative la version lissée  $\widetilde{W}_i$  de la courbe bruitée  $W_i$ . Cette méthode est basée sur la diagonalisation de l'opérateur de covariance empirique  $\Gamma_{W,n}$  associé à  $W$ . On note  $(\lambda_r)_{r \geq 1}$  la suite des valeurs propres de  $\Gamma_{W,n}$  (rangées par ordre décroissant) et  $(\psi_r)_{r \geq 1}$  la suite de fonctions propres associées, telles que pour tout  $r \geq 1$ ,

$$\Gamma_{W,n} \psi_r = \lambda_r \psi_r.$$

En introduisant les coefficients

$$\eta_{ir} = \langle W_i, \psi_r \rangle, \quad (3.2)$$

pour tout  $i = 1, \dots, n$  et pour tout  $r \geq 1$ , on a alors

$$\sum_{i=1}^n \eta_{ir} = 0,$$

pour tout  $r \geq 1$  et

$$\sum_{i=1}^n \eta_{ir} \eta_{is} = \lambda_r \Pi_{[r=s]},$$

pour  $r, s \geq 1$ , avec  $\Pi_{[r=s]} \begin{cases} 1 & \text{si } r = s \\ 0 & \text{sinon} \end{cases}$

On utilise alors l'idée suivante provenant de Kneip et Utikal (2001)[46], l'idée également reprise dans Benko, Härdle et Kneip (2005)[7]. Pour déterminer les fonctions propres  $\psi_r$ ,  $r \geq 1$ , il n'est pas nécessaire de passer par l'opérateur  $\Gamma_{W,n}$ . En effet, on peut considérer à la place la matrice  $M$  de taille  $n \times n$  définie par

$$M_{i_1 i_2} = \frac{1}{n} \langle W_{i_1}, W_{i_2} \rangle, \quad (3.3)$$

pour  $i_1, i_2 = 1, \dots, n$ . L'avantage d'utiliser cette matrice est que l'on estime les produits scalaires entre les courbes, donc des nombres réels. On obtiendra ainsi des vitesses de convergences avec un biais en  $h_i^2$  et une variance en  $1/n$ . L'étude de ces vitesses sera détaillée dans la section suivante.

Ainsi, en utilisant ce qui a été fait dans la première étape, on construit une estimation de la matrice  $M$  définie par (3.3), en utilisant les estimations  $\widetilde{W}_i$  de  $W_i$ , pour  $i = 1, \dots, n$ . L'estimateur le plus naturel  $\widehat{M}$  de  $M$  semble être la matrice de taille  $n \times n$  et de terme général  $\widehat{M}_{i_1 i_2} = \frac{1}{n} \langle \widetilde{W}_{i_1}, \widetilde{W}_{i_2} \rangle$ , pour  $i_1, i_2 = 1, \dots, n$ . Cependant, comme cela a été souligné, on cherche à estimer ici des produits scalaires entre des courbes et non les courbes elles-mêmes. Si

on regarde l'estimateur  $\langle \widetilde{W}_{i_1}, \widetilde{W}_{i_2} \rangle$  lorsque  $i_1 = i_2 = i$ , on remarque que ce terme s'écrit

$$\begin{aligned} \langle \widetilde{W}_i, \widetilde{W}_i \rangle &= \sum_{j_1=1}^p \sum_{j_2=1}^p W_i(t_{j_1}) W_i(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right) K\left(\frac{t-t_{j_2}}{h_i}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} dt \\ &= \sum_{j_1=1}^p \sum_{j_2=1, j_2 \neq j_1}^p W_i(t_{j_1}) W_i(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right) K\left(\frac{t-t_{j_2}}{h_i}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} dt \\ &\quad + \sum_{j_1=1}^p W_i(t_{j_1})^2 \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right)^2}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} dt. \end{aligned}$$

Ainsi, le terme

$$\sum_{j_1=1}^p W_i(t_{j_1})^2 \int_I \frac{K\left(\frac{t-t_{j_1}}{h_i}\right)^2}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_i}\right)\right]^2} dt$$

produit un biais dans l'estimation de  $M_{ii}$ . L'idée est donc de le supprimer dans l'estimateur. Cette idée a été utilisée par Kneip et Utikal (2001)[46] dans le cadre de l'estimation de densités puis par Benko, Härdle et Kneip (2005)[7] concernant l'analyse en composantes principales fonctionnelle. On se rend compte cependant que cette idée avait déjà été proposée auparavant par Hall et Marron (1987)[42] et Jones et Sheater (1991)[44] dans le cadre de l'intégration de densités. Concernant l'estimation de l'intégrale du carré de fonctions de régression (qui nous intéressera plus particulièrement ici), on peut citer les travaux de Benhenni et Cambanis (1992)[12], Ruppert, Sheater et Wand (1993)[57], ainsi que Huang et Fan (1999)[43]. Ceci permet de gagner

### 3.1.2 Régression sur composantes principales pour les variables sont bruitées 40

au niveau du biais de l'estimation de la matrice  $M$ . Dans la suite, on considère donc l'estimateur, pour  $i_1, i_2 = 1, \dots, n$ ,

$$\widehat{M}_{i_1 i_2} = \begin{cases} \frac{1}{n} \sum_{j_1=1}^p \sum_{j_2=1}^p W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1}}\right) K\left(\frac{t-t_{j_2}}{h_{i_2}}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_{i_1}}\right)\right] \left[\sum_{j_4=1}^p K\left(\frac{t-t_{j_4}}{h_{i_2}}\right)\right]} dt, \\ \text{si } i_1 \neq i_2, \\ \\ \frac{1}{n} \sum_{j_1=1}^p \sum_{j_2=1, j_2 \neq j_1}^p W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1}}\right) K\left(\frac{t-t_{j_2}}{h_{i_2}}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_{i_1}}\right)\right]^2} dt, \text{ si } i_1 = i_2. \end{cases}$$

De plus, en ayant à l'esprit que l'on estime des produits scalaires plutôt que des courbes, il semble plus approprié de choisir une même largeur de fenêtre  $h_{i_1 i_2}$  pour l'estimation du produit scalaire entre le couple de courbes  $\{W_{i_1}, W_{i_2}\}$ , pour  $i_1, i_2 = 1, \dots, n$ . Une légère modification de l'estimateur ci-dessus nous donne alors

$$\widehat{M}_{i_1 i_2} = \begin{cases} \frac{1}{n} \sum_{j_1=1}^p \sum_{j_2=1}^p W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1 i_2}}\right) K\left(\frac{t-t_{j_2}}{h_{i_1 i_2}}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_{i_1 i_2}}\right)\right]^2} dt, \text{ si } i_1 \neq i_2, \\ \\ \frac{1}{n} \sum_{j_1=1}^p \sum_{j_2=1, j_2 \neq j_1}^p W_{i_1}(t_{j_1}) W_{i_2}(t_{j_2}) \int_I \frac{K\left(\frac{t-t_{j_1}}{h_{i_1 i_2}}\right) K\left(\frac{t-t_{j_2}}{h_{i_1 i_2}}\right)}{\left[\sum_{j_3=1}^p K\left(\frac{t-t_{j_3}}{h_{i_1 i_2}}\right)\right]^2} dt, \text{ si } i_1 = i_2. \end{cases} \quad (3.4)$$

Avec cette estimation  $\widehat{M}$  de  $M$ , on calcule les valeurs propres  $\widehat{\lambda}_r$  et les vecteurs propres  $\widehat{p}_r$  correspondants, pour  $r = 1, \dots, n$ . On en déduit les esti-



mations  $\widehat{\eta}_{ir}$  et  $\widehat{\psi}_r$  de  $\eta_{ir}$  et  $\psi_r$  en utilisant les relations

$$\eta_{ir} = \sqrt{\lambda_r} p_{ir},$$

et

$$\psi_r = \frac{1}{\sqrt{\lambda_r}} \sum_{i=1}^n p_{ir} W_i = \frac{\sum_{i=1}^n \eta_{ir} W_i}{\sum_{i=1}^n \eta_{ir}^2}.$$

Finalement, on construit l'estimateur de  $\alpha$  donné par l'approximation d'ordre  $K \geq 1$ , c'est-à-dire en utilisant les  $K$  premières composantes principales (voir Cardot, Ferraty et Sarda, 1999, 2003[13]). Notre estimateur est ainsi donné par

$$\widehat{\alpha}_K = \frac{1}{n} \sum_{r=1}^K \sum_{i=1}^n \frac{Y_i}{\widehat{\lambda}_r} \langle \widehat{W}_i, \widehat{\psi}_r \rangle \widehat{\psi}_r. \quad (3.5)$$

## 3.2 Résultats asymptotiques

Les résultats que l'on va établir ici sont directement inspirés des travaux de Kneip et Utikal (2001)[46] et Benko, Härdle et Kneip (2005)[7].

**Théorème 3.2.1.** *Pour tous  $i_1, i_2 = 1, \dots, n$ , si on prend  $h_{i_1 i_2}$  de la forme  $p^{-\varsigma}$  avec  $\varsigma \in [1/4, 1/2[$ , on a*

$$\widehat{M}_{i_1 i_2} - M_{i_1 i_2} = o_{\mathbb{P}} \left( \frac{1}{np^{1/2}} \right).$$

**Remarque 3.2.1.** *La preuve du la théorème 3.2.1 est basée sur les propositions suivantes :*

**Proposition 3.2.1.** *On a*

$$\mathbb{E}(\widehat{\theta}) - \theta = \frac{\mu_2(k)}{2} \left( \int_0^1 [r(x)s''(x) + r''(x)s(x)] dx \right) h^2 + o(h^2).$$

**Proposition 3.2.2.**

$$\mathbb{V}(\widehat{\theta}) = 2 \left[ \tau^2 \left( \int_0^1 r(x)^2 dx \right) + \sigma^2 \left( \int_0^1 s(x)^2 dx \right) \right] \times \left( \int_0^2 \Psi(z) dz \right) \frac{1}{n} + o\left(\frac{1}{n}\right).$$

**Proposition 3.2.3.** *On a*

$$\mathbb{E}(\widehat{\theta}) - \theta = \mu_2(K) \left( \int_0^1 r(x)r''(x) dx \right) h^2 - R(K) \left( \int_0^1 r(x)^2 dx \right) \frac{1}{nh} + o\left(h^2 + \frac{1}{nh}\right).$$

**Proposition 3.2.4.** *On a*

$$\mathbb{V}(\widehat{\theta}) = 8\sigma^2 \left( \int_0^1 r(x)^2 dx \right) \left( \int_0^2 \Psi(z) dz \right) \frac{1}{n} + o\left(\frac{1}{n}\right).$$

**Preuve 3.2.1.**

Dans le cas où  $i_1 \neq i_2$ , on a, d'après les propositions 3.2.1 et 3.2.2,

$$\begin{aligned} & \mathbb{E} \left( \widehat{M}_{i_1 i_2} - M_{i_1 i_2} \right) \\ &= \frac{1}{n} \left\{ \frac{u_2(k)}{2} \left( \int_0^1 [X_{i_1}(t)X_{i_2}''(t) + X_{i_1}''(t)X_{i_2}(t)] dt \right) h_{i_1 i_2}^2 + o(h_{i_1 i_2}^2) \right\}, \end{aligned}$$

et

$$\begin{aligned} & \mathbb{V} \left( \widehat{M}_{i_1 i_2} - M_{i_1 i_2} \right) \\ &= \frac{1}{n^2} \left\{ 2 \left[ \sigma_\delta^2 \int_0^1 [X_{i_1}(t)^2 + X_{i_2}(t)^2] dt \right] \left( \int_0^2 \Psi(z) dz \right) \frac{1}{p} + o\left(\frac{1}{p}\right) \right\}. \end{aligned}$$

De même, dans le cas où  $i_1 = i_2 = i$ , on a, d'après les propositions 3.2.3 et 3.2.4,

$$\mathbb{E} \left( \widehat{M}_{ii} - M_{ii} \right) = \frac{1}{n} \left\{ u_2(k) \left( \int_0^1 X_i(t)X_i''(t) dt \right) h_{ii}^2 + o(h_{ii}^2) \right\},$$

et

$$\mathbb{V}(\widehat{M}_{ii} - M_{ii}) = \frac{1}{n^2} \left\{ 4\sigma_\delta^2 \left( \int_0^1 X_i(t)^2 dt \right) \left( \int_0^2 \Psi(z) dz \right) \frac{1}{p} + o\left(\frac{1}{p}\right) \right\}.$$

On en déduit, pour tous  $i_1, i_2 = 1, \dots, n$ ,

$$\mathbb{V}(\widehat{M}_{i_1 i_2} - M_{i_1 i_2}) = o\left(\frac{h_{i_1 i_2}^2}{n}\right),$$

et

$$\mathbb{V}(M_{i_1 i_2}) = o\left(\frac{1}{n^2 p}\right).$$

Ainsi, en prenant  $h_{i_1 i_2}$  de la forme  $p^{-\varsigma}$  avec  $\varsigma \in [1/4, 1/2[$ , on s'assure d'un biais négligeable, et le résultat du la théorème 3.2.1 est immédiat.

**Théorème 3.2.2.** *Si on note  $\|\cdot\|$  la norme matricielle euclidienne usuelle, on a*

$$\|\widehat{M} - M\|^2 = O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

*Les deux théorèmes suivantes donnent le comportement des valeurs propres et des vecteurs propres de  $\widehat{M}$  par rapport à ceux de  $M$ .*

**Preuve 3.2.2.**

On a

$$\|\widehat{M} - M\|^2 \leq \text{Tr} \left[ (\widehat{M} - M)^\tau (\widehat{M} - M) \right] = \sum_{i_1=1}^n \sum_{i_2=1}^n (\widehat{M}_{i_1 i_2} - M_{i_1 i_2})^2,$$

et le résultat est alors immédiat en utilisant théorème 3.2.1.

**Théorème 3.2.3.** *On suppose que les valeurs propres de  $M$  sont telles que, pour tout  $r = 1, \dots, L$ , il existe des constantes  $0 < C_{1r} < +\infty$  et  $0 < C_{2r} \leq C_{3r} < +\infty$  vérifiant*

$$\min_{s=1, \dots, n, s \neq r} |\lambda_r - \lambda_s| \geq C_{1r},$$

et

$$C_{2r} \leq \lambda_r \leq C_{3r}.$$

Alors, on a, pour tout  $r = 1, \dots, K$ ,

$$\widehat{\lambda}_r - \lambda_r = o_{\mathbb{P}} \left( \frac{1}{n^{1/2}p^{1/2}} + \frac{1}{p} \right).$$

**Preuve 3.2.3.**

Sous les hypothèses du la théorème 3.2.3, on a, pour tout  $r = 1, \dots, K$ ,

$$\widehat{\lambda}_r - \lambda_r = o_{\mathbb{P}} \left( \frac{1}{n^{1/2}p^{1/2}} + \frac{1}{p} \right).$$

On commence par donner un résultat provenant de Kneip et Utikal (2001)[46].

On a

$$\widehat{\lambda}_r - \lambda_r = Tr \left[ P_{\varepsilon_r} \left( \widehat{M} - M \right) \right] + R_1,$$

où  $P_{\varepsilon_r}$  désigne la matrice de projection sur le sous-espace propre associé à la  $r^{eme}$  valeur propre  $\lambda_r$  et  $R_1$  vérifie

$$|R_1| \leq \frac{6 \left\| \widehat{M} - M \right\|^2}{\min_{s=1, \dots, n, s \neq r} |\lambda_r - \lambda_s|}.$$

En utilisant théorème 3.2.2 et l'hypothèse de cette théorème concernant les valeurs propres de  $M$ , on en déduit que

$$R_1 = O_{\mathbb{P}}\left(\frac{1}{p}\right),$$

ce qui donne

$$\widehat{\lambda}_r - \lambda_r = p_r^\tau \left( \widehat{M} - M \right) p_r + O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

D'autre part, on a

$$\begin{aligned} & \mathbb{E} \left[ \left( p_r^\tau \left( \widehat{M} - M \right) p_r \right)^2 \right] \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n \sum_{i_4=1}^n p_{i_1 r} p_{i_2 r} p_{i_3 r} p_{i_4 r} \mathbb{E} \left[ \left( \widehat{M}_{i_1 i_2} - M_{i_1 i_2} \right) \left( \widehat{M}_{i_3 i_4} - M_{i_3 i_4} \right) \right] \\ &= \sum_{i_1=1}^n \sum_{i_2=1}^n \sum_{i_3=1}^n p_{i_1 r}^2 p_{i_2 r} p_{i_3 r} \mathbb{E} \left[ \left( \widehat{M}_{i_1 i_2} - M_{i_1 i_2} \right) \left( \widehat{M}_{i_1 i_3} - M_{i_1 i_3} \right) \right] \\ &\leq \sum_{i_1=1}^n p_{i_1 r}^2 \sum_{i_2=1}^n |p_{i_2 r}| \sum_{i_3=1}^n |p_{i_3 r}| \times o\left(\frac{1}{n^2 p}\right), \end{aligned}$$

ce qui donne finalement, vu que  $\sum_{i_1=1}^n p_{i_1 r}^2 = 1$  et que  $\sum_{i_2=1}^n |p_{i_2 r}| = o(n^{1/2})$ ,

$$p_r^\tau \left( \widehat{M} - M \right) p_r = o_{\mathbb{P}}\left(\frac{1}{n^{1/2} p^{1/2}}\right),$$

et achève la preuve du la théorème 3.2.3.

**Théorème 3.2.4.** *Sous les mêmes hypothèses qu'à la théorème 3.2.3, on a*

$$\|\widehat{p}_r - p_r\| = O_{\mathbb{P}}\left(\frac{1}{p^{1/2}}\right).$$

**Preuve 3.2.4.**

Sous les hypothèses de la théorème 3.2.3, on a

$$\|\widehat{p}_r - p_r\| = O_{\mathbb{P}}\left(\frac{1}{p^{1/2}}\right).$$

On commence là aussi par énoncer un résultat provenant de Kneip et Utikal (2001)[46]. On a

$$\widehat{p}_r - p_r = -S_r \left(\widehat{M} - M\right) p_r + R_2,$$

où  $S_r$  est la matrice définie par

$$S_r = \sum_{s \neq r} \frac{1}{\lambda_s - \lambda_r} P_{\varepsilon_s},$$

et  $R_2$  vérifie

$$\|R_2\| \leq \frac{6 \|\widehat{M} - M\|^2}{\min_{\lambda \neq \lambda_r} |\lambda - \lambda_r|^2}.$$

En utilisant théorème 3.2.2 et l'hypothèse de ce théorème concernant les valeurs propres de  $M$ , on en déduit que

$$\|R_2\| = O\left(\frac{1}{p}\right).$$

En posant  $q_r = \left(\widehat{M} - M\right) p_r$ , on obtient donc

$$\|\widehat{p}_r - p_r + S_r q_r\| = O_{\mathbb{P}}\left(\frac{1}{p}\right).$$

Calculons maintenant  $\|S_r\|$ . On a

$$\begin{aligned} \|S_r\| &= \sup_{\|v\|=1} \left[ v^\tau \left( \sum_{s_1 \neq r} \frac{1}{\lambda_{s_1} - \lambda_r} p_{s_1} p_{s_1}^\tau \right) \left( \sum_{s_2 \neq r} \frac{1}{\lambda_{s_2} - \lambda_r} p_{s_2} p_{s_2}^\tau \right) v \right]^{1/2} \\ &\leq \frac{1}{l_r} \sup_{\|v\|=1} \left[ v^\tau \left( \sum_{s_1 \neq r} p_{s_1} p_{s_1}^\tau \right) \left( \sum_{s_2 \neq r} p_{s_2} p_{s_2}^\tau \right) v \right]^{1/2}, \end{aligned}$$

où  $l_r = \min\{|\lambda_{r-1} - \lambda_r|, |\lambda_{r+1} - \lambda_r|\}$ . Comme le suprémum ci-dessus n'est autre que  $\left\| \sum_{s \neq r} p_s p_s^\tau \right\|$ , on en déduit qu'il est inférieur à  $\left\| \sum p_s p_s^\tau \right\| = 1$ , d'où

$$\|S_r\| \leq \frac{1}{l_r}.$$

D'après les hypothèses faites sur les  $\lambda_r$ , on en déduit alors que

$$\|S_r\| = O(1).$$

Pour tout  $v \in \mathbb{R}^n$ , on a alors

$$\begin{aligned} \mathbb{E}[(v^\tau S_r q_r)^2] &= v^\tau S_r \mathbb{E}(q_r q_r^\tau) S_r v \\ &\leq \|v\|^2 \|S_r\|^2 \mathbb{E} \left[ p_r^\tau \left( \widehat{M} - M \right)^\tau \left( \widehat{M} - M \right) p_r \right]. \end{aligned}$$

D'après ce qui précède, on obtient ainsi

$$\mathbb{E}[(v^\tau S_r q_r)^2] = o\left(\frac{\|v\|^2}{np}\right).$$

Finalement, en prenant comme vecteur  $v$  le vecteur dont toutes les composantes sont nulles sauf la  $i$ ème qui vaut 1, il vient

$$\mathbb{E}[(S_r q_r)_i^2] = o\left(\frac{1}{np}\right),$$

d'où

$$|\widehat{p}_{ir} - p_{ir}| = o_{\mathbb{P}} \left( \frac{1}{n^{1/2} p^{1/2}} \right),$$

ce qui prouve le résultat du la théorème 3.2.4.



# Chapitre 4

## simulations

Nous avons simulé des échantillons  $(X_i, Y_i), i = 1, \dots, n$ , à partir du modèle

$$Y = \int_0^1 \alpha(t)X(t)dt + \epsilon,$$

- dans lequel  $X(t)$  est un mouvement brownien définie sur  $[0, 1]$ ,
- $\epsilon$  est normale avec une moyenne 0 et la variance  $0.2 \text{ var}(\alpha(X))$ .
- L'espace de Hilbert  $H$  est  $L^2[0, 1]$  et les éléments propres de l'opérateur de covariance  $\Gamma_X$  sont connus pour être (voir Ash et Gardner, 1975[1]) :

$$\lambda_j = \frac{1}{(j - 0.5)^2 \pi^2}, \quad V_j(t) = \sqrt{2} \sin\{(j - 0.5)\pi t\}, \quad t \in [0, 1], \quad j = 1, 2, \dots$$

Nous avons fait des simulations pour deux fonctions différentes  $\alpha$  :

- $\alpha_1(t) = \sin(\pi t/2) + 0.5 \sin(3\pi t/2) + 0.25 \sin(5\pi t/2)$ .
- $\alpha_2(t) = \sin(4\pi t)$ .

Dans le premier cas, la fonction  $\alpha_1$  est une combinaison linéaire des fonctions propres associé aux trois plus grandes valeurs propres de  $\Gamma$ , de sorte que la meilleure  $k_n$  dimension devrait être 3. Nous avons essayé plusieurs tailles d'échantillon dans chaque cas :

$n = 50, 200, 1000$ . Pour faire face pratiquement avec les fonctions aléatoires browniens  $X_i(t)$ , leurs chemins d'échantillon ont été discrétisés par 100 points équidistants dans  $[0, 1]$ .

Le but de notre étude est d'examiner la meilleure  $k_n$  dimension pour la procédure d'estimation et nous avons donc considéré comme le critère d'erreur suivant :

$$R(\hat{\alpha}_{k_n}) = \frac{\int_0^1 (\alpha(t) - \hat{\alpha}_{k_n}(t))^2 dt}{\int_0^1 \alpha^2(t) dt}.$$

**Tableau 1**

Erreur quadratique pour l'estimation de  $\alpha_1$

$k_n$	n=50	n=200	n=1000
2	3.33	6.58	2.99
3	5.46	3.93	<b>1.76</b>
4	11.27	3.93	1.79
5	10.5	3.92	1.96
6	74.3	3.97	3.72

**Tableau 2**

Erreur quadratique pour l'estimation de  $\alpha_2$

$k_n$	n=50	n=200	n=1000
4	81.8	40.8	15.76
5	9.9	18.14	1.9
6	13.6	6.02	1.02
7	15	5.58	0.92
8	19.2	4.38	0.49
9	51.6	7.64	0.54
10	53	11.62	1.71

Tableau 1 (resp. Le tableau 2) donne l'erreur quadratique des estimations de  $\alpha_1$  (resp.  $\alpha_2$ ) pour chaque taille de l'échantillon et les différentes dimensions  $k_n$ . Dans chaque cas, on peut remarquer que  $R(\hat{\alpha}_{k_n})$  ressemble à une fonction convexe de dimension  $k_n$  et un trop grand  $k_n$  donne de mauvaises estimations de  $\alpha$  en augmentant la variance de l'estimation. Aussi, il

apparaît que, pour le premier exemple, la meilleure cote sélectionnée pour la procédure d'estimation est raisonnablement proche de la dimension théorique "optimale". Ce dernier point illustre le bon comportement de notre estimateur. Dans la vraie étude de la vie, cette erreur de critère quadratique ne peut être calculé et d'autre part, il ressort de la simulation ci-dessus que la qualité de l'estimateur dépend fortement du choix de la valeur de la dimension,  $k_n$ . Une méthode de sélection fondée sur les données telles que la validation croisée pénalisée peut être utilisé dans ce cas (Voir Vieux, 1995[60], qui utilise un tel critère pour le choix de l'ordre dans les modèles autorégressifs non linéaires).

Nous avons tiré les estimations  $\hat{\alpha}_{k_n}$  de la fonction  $\alpha_1$ .

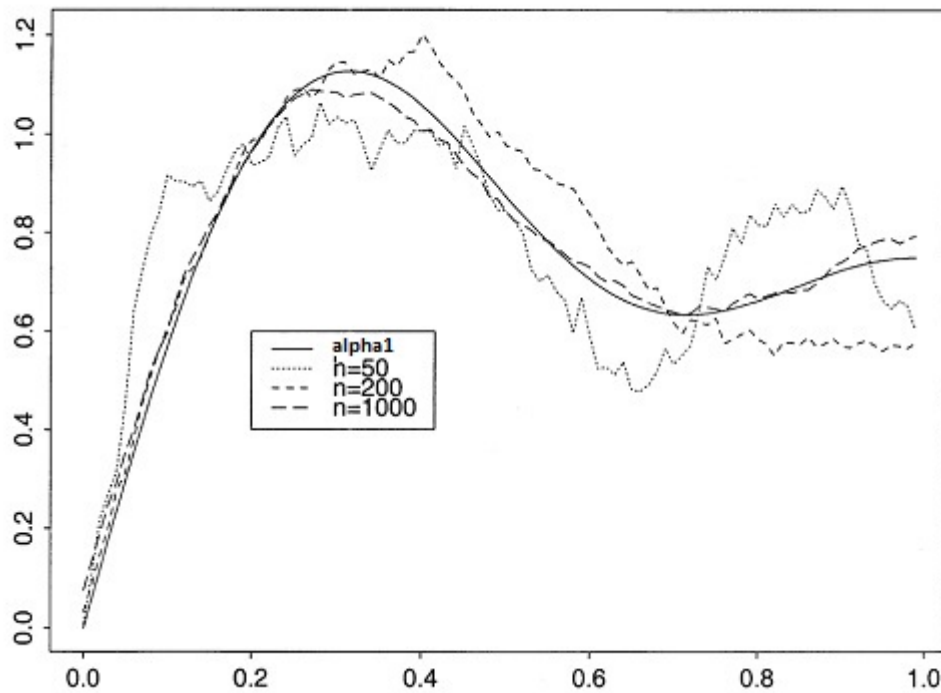


FIGURE 4.1 –

Figure 4.1 montre la bonne performance de la procédure d'estimation pour assez grande taille de l'échantillon. Pour de plus petite taille de l'échantillon de l'estimateur présente des caractéristiques grossières, même si la forme générale de la fonction est restituée. Nous pensons que cet aspect de l'estimateur pourrait être corrigé par l'introduction d'une procédure de lissage préliminaire comme dans Besse et Cardot (1996)[10]. Nous allons enquêter sur ce sujet dans une autre étude.

# Conclusion et Perspectives

## Conclusion

Jusqu'à présent, au vu de l'écriture des modèles (1.1) et (1.2), on a toujours implicitement fait l'hypothèse que les courbes  $X_1, \dots, X_n$  sont observées sans erreur. Cette hypothèse peut se révéler assez peu réaliste en pratique puisque de nombreuses erreurs (comme entre autres des erreurs de mesure) peuvent empêcher de connaître les courbes  $X_1, \dots, X_n$  exactement. Il semble alors plus réaliste de considérer que la variable explicative réellement disponible est une variable  $W_i$  (pour  $i = 1, \dots, n$ ) telle qu'en chaque point de mesure  $t_j$  (pour  $j = 1, \dots, p$ ), on a

$$W_i(t_j) = X_i(t_j) + \delta_{ij},$$

où  $(\delta_{ij})_{i=1, \dots, n, j=1, \dots, p}$  est une suite de variables aléatoires indépendantes et identiquement distribuées représentant les erreurs (de mesures, . . .) faites en chaque point  $t_1, \dots, t_p$ , et telles que  $\mathbb{E}(\delta_{ij}) = 0$  et  $\mathbb{E}(\delta_{ij}^2) = \sigma_\delta^2$  pour tout  $i = 1, \dots, n$  et pour tout  $j = 1, \dots, p$ .

Ce modèle avec des erreurs dans les variables explicatives a été l'objet de nombreuses études dans le cadre multivarié (c'est-à-dire lorsque  $X_1, \dots, X_n$  sont des éléments de  $\mathbb{R}^p$ ). Par exemple, Fuller (1987)[30] donne pour ce modèle bruité une méthode par maximum de vraisemblance. Des résultats asymptotiques sont également donnés par Gleser (1981)[36]. Une méthode numérique importante, connue sous le nom de moindres carrés orthogonaux, a été notamment présentée par Golub et Van Loan (1980)[32], puis reprise et

développée dans un ouvrage de Van Huffel et Vandewalle (1991)[59]. L'idée de départ de cette méthode consiste à rajouter dans le problème de minimisation des moindres carrés la quantité

$$\frac{1}{n} \sum_{i=1}^n \|W_i - X_i\|^2,$$

où  $W_i$  et  $X_i$  (pour  $i = 1, \dots, n$ ) désignent les vecteurs de taille  $p$  de termes généraux respectifs  $W_i(t_j)$  et  $X_i(t_j)$ , pour tout  $j = 1, \dots, p$ , et  $\|\cdot\|$  désigne la norme vectorielle euclidienne usuelle (ici dans  $\mathbb{R}^p$ ). On détermine alors (voir Golub et Van Loan, 1980[32]) la solution du problème de minimisation, construisant ainsi un estimateur consistant de la moyenne conditionnelle.

Dans notre cadre fonctionnel, le cas de variables explicatives bruitées a déjà été considéré. Les méthodes envisagées sont généralement basées sur un débruitage de chaque courbe par des techniques de lissage (voir par exemple Chiou, Muller et Wang, 2003, Cardot, 2006[18]). dans ce travail, on propose de généraliser la méthode des moindres carrés orthogonaux à ce cadre fonctionnel, fournissant ainsi une méthode "globale" de débruitage, et non plus courbe par courbe. Après avoir expliqué le fonctionnement et le principe de résolution de cette méthode des moindres carrés orthogonaux dans le cas multivarié, l'objet de chapitre 2 sera de donner sa généralisation au cas d'une variable explicative fonctionnelle, et de s'intéresser au comportement asymptotique de l'estimateur construit par la méthode des moindres carrés orthogonaux. La méthode sera envisagée pour les splines de régression (c'est-à-dire dans le même contexte que dans les travaux de Cardot, Ferraty et Sarda, 2003[13]).

Concernant ce problème de variable explicative bruitée, une autre piste a commencée à être envisagée au cours de cet travail. En revenant à une méthode de débruitage courbe par courbe, elle consiste à considérer un lissage de chaque courbe bruitée (par exemple un lissage à noyau), puis de faire une

régression sur composantes principales fonctionnelle à l'aide de la technique utilisée par Kneip et Utikal (2001)[46] ou encore par Benko, Härdle et Kneip (2005)[7]. Comme cela a déjà été signalé, dans ces articles, lors de l'estimation de l'opérateur de covariance, on estime des produits scalaires entre les courbes observées plutôt que les courbes elles-mêmes. Cette méthode d'estimation a commencé à donner des résultats encourageants tant au niveau pratique sur des simulations que théorique avec la recherche de vitesses de convergence pour l'estimateur de  $\alpha$ . Ces premiers résultats seront également présentés dans chapitre 3.

## Perspectives

- Les modèles avec variables bruitées sont certainement amenés à se développer dans le futur, tant ils semblent en adéquation avec la réalité (à partir du moment où on mesure des variables, elles sont nécessairement entachées d'erreurs). Du coup, les perspectives de travaux futurs sur ce modèle défini par (2.1) sont multiples.

L'idée la plus naturelle pour supprimer le bruit de la courbe explicative est de la lisser (par exemple par un lissage à noyau). Cette méthode a commencé à être envisagée et des premiers travaux sont en cours. Le début de cette étude est présenté à la chapitre 3 dans cet travail . Les premiers résultats semblent encourageants.

Toujours concernant une variable explicative bruitée, il est peut-être envisageable de considérer l'estimation de quantiles conditionnels. La transposition de la méthode des moindres carrés orthogonaux à ce contexte n'est à priori pas évidente, principalement dû au fait que le problème de minimisation relatif à l'estimation de quantiles n'a pas de solution explicite.

Enfin, une autre perspective à envisager dans ce contexte de variable explicative bruitée propose un travail à plus long terme. L'idée serait de considérer un bruit à temps continu. Un problème se pose immédiatement : il n'existe pas de bruit blanc à temps continu. On peut alors envisager un bruit continu  $\delta(t)$  qui vérifierait par exemple une hypothèse de mélange.

- Ce début de travail donne des résultats assez encourageants. D'un point de vue théorique, il faut maintenant prouver un résultat de convergence concernant l'estimateur  $\hat{\alpha}_L$  de  $\alpha$  défini par

$$\hat{\alpha}_K = \frac{1}{n} \sum_{r=1}^K \sum_{i=1}^n \frac{Y_i}{\hat{\lambda}_r} \langle \widetilde{W}_i, \hat{\psi}_r \rangle \hat{\psi}_r,$$

ce qui devrait être raisonnablement envisageable au vu des résultats de convergence précédents concernant les valeurs propres et les vecteurs propres de l'estimateur de la matrice  $M$ .

D'un point de vue un peu plus appliqué, il reste à faire tout un travail sur le choix du nombre  $K$  de composantes principales. Une méthode par validation croisée peut à priori être envisagée, alors que Kneip et Utikal (2001)[46] proposent une procédure de test dans leur contexte d'estimation de densités. Enfin, il paraît assez intéressant de comparer cette procédure d'estimation avec celle des moindres carrés orthogonaux.



# Bibliographie

- [1] Ash and Gardner, (1975). Topics in Stochastic Processes. Academic Press, New York.
- [2] Aneiros-Perez, Cardot, Estevez-Perez and Vieu (2004). Maximum ozone concentration forecasting by functional nonparametric approaches. *Environmetrics*, 15, 675-685.
- [3] Berlinet, Biau and Rouvière (2005). Functional classification with wavelets. Preprint.
- [4] Besse, Cardot and Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27, 673-687.
- [5] Besse, Cardot and Stephenson (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27, 673-687.
- [6] Bosq, 2000. Linear processes in function spaces. *Lecture Notes in Statistics*, 149, Springer.
- [7] Benko, Härdle and Kneip (2005). Common functional principal components. SFB 649 Economic Risk Discussion Paper, 2006-010.
- [8] Breiman, Friedman, Olshen and Stone (1984). Classification and regression trees. *Wadsworth Statistics and Probability Series*, Wadsworth Advanced Books and Software, Belmont.
- [9] Besse and Ramsay (1986). Principal components analysis of sampled functions. *Psychometrika*, 51, 285-311.

- 
- [10] Besse and Cardot (2000). Autoregressive forecasting of some functional climatic variations. *Scandinavian Journal of Statistics*, 27, 673-687.
- [11] Benko, Müller and Stadtmüller (2005). Generalized functional linear models. *Annals of Statistics*, 33, 774-805.
- [12] Benhenni and Cambanis (1992). Sampling designs for estimating integrals of stochastic processes. *Annals of Statistics*, 20, 161-194.
- [13] Cardot, Ferraty and Sarda (1999, 2003). Functional linear model. *Statistic and Probability Letters*, 45, 11-22. Spline estimators for the functional linear model. *Statistica Sinica*, 13, 571-591.
- [14] Cuevas, Febrero and Fraiman (2002). Linear functional regression : the case of a fixed design and functional response. *Canadian Journal of Statistics*, 30, 285-300.
- [15] Chiou, Müller and Wang (2004). Functional response models. *Statistica Sinica*, 14, 675-693.
- [16] Cardot and Sarda (2005). Estimation in generalized linear models for functional data via penalized likelihood. *Journal of Multivariate Analysis*, 92, 24-41.
- [17] Chiou, Müller, Wang et Carey, 2003. A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica*, 13, 1119-1133.
- [18] Chiou, Müller and Wang, 2003, Cardot, 2006. Functional quasilikelihood regression models with smooth random effects. *Journal of the Royal Statistical Society, Series B*, 65, 405-423.
- [19] Crambes, 2005. Total least squares for functional data. Invited paper in *ASMDA 2005 Proceedings*, 619-626.
- [20] Cardot, 2000. Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *Nonparametric Statistics*, 12, 503-538.

- 
- [21] Chatelin (1983). Spectral approximation of linear operators. Academic Press, New-York.
- [22] Dunford and Schwarz, (1963). Linear operators. Interscience, New York.
- [23] Deville (1974).
- [24] Dauxois and Pousse (1976). Les analyses factorielles en calcul des probabilités et en statistique : essai d'étude synthétique. Thèse de doctodocorat, Université Paul Sabatier, Toulouse.
- [25] Dauxois, Pousse and Romain (1982). Asymptotic theory for the principal component analysis of a random vector function : some applications to statistical inference. *Journal of Multivariate Analysis*, 12, 136-154.
- [26] Devroye, Györfi et Lugosi, 1996. A probabilistic theory of pattern recognition. *Applications of Mathematics*, Springer, New York.
- [27] Frank et Friedman (1993). A statistical view of some chemometrics regression tools. *Technometrics*, 35, 109-135.
- [28] Ferraty and Vieu (2002, 2003). The functional nonparametric model and application to spectrometric data. *Computational Statistics*, 17, 545-564. Curves discrimination : a nonparametric functional approach. Special issue in honour of Stan Azen : a birthday celebration. *Computational Statistics and Data Analysis*, 44, 161-173.
- [29] Ferraty et Vieu (2006). Nonparametric functional data analysis : theory and practice. Springer, New York.
- [30] Fuller (1987). Measurement error models. Wiley, New York.
- [31] Gohberg and Krein, (1971). Introduction à la théorie des opérateurs linéaires non auto-adjoints dans un espace hilbertien. Dunod, Paris.
- [32] Golub and Van Loan, 1980. An analysis of the total least squares problem. *SIAM, Journal of Numerical Analysis*, 17, 883-893.
- [33] Golub and Van Loan, 1996. Matrix computations. Johns Hopkins University Press, Baltimore.

- 
- [34] Golub, Hansen and O'Leary (1999). Tikhonov regularization and total least squares. *SIAM, Journal of Matrix Analysis and Applications*, 21, 185-194.
- [35] Gasser, Sroka and Jennen-Steinmetz (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, 3, 625-633.
- [36] Gleser (1981). Estimation in a multivariate "variables" regression model : large sample results. *Annals of Statistics*, 9, 24-44.
- [37] Hoerl and Kennard, (1980). Ridge regression : advances, algorithms and applications. *American Journal of Mathematical Management Sciences*, 1, 5-83.
- [38] Helland, (1990). Partial least squares regression and statistical models. *Scandinavian Journal of Statistics*, 17, 97-114.
- [39] Hastie et Mallows (1993). Discussion of "A statistical view of some chemometrics regression tools." by Frank, I.E. and Friedman, J.H. *Technometrics*, 35, 140-143.
- [40] Hastie, Buja and Tibshirani (1995). Penalized discriminant analysis. *Annals of Statistics*, 23, 73-102.
- [41] Härdle (1991). Smoothing techniques with implementation in S. Springer, New-York.
- [42] Hall and Marron (1987). Estimation of integrated squared density derivatives. *Statistics and probability Letters*, 6, 109-115.
- [43] Huang and Fan (1999). Nonparametric estimation of quadratic regression functionals. *Bernoulli*, 5, 927-949.
- [44] Jones and Sheater (1991). Using non-stochastic terms to advantage in kernel-based estimation of integrated squared density derivatives. *Statistics and probability Letters*, 11, 511-514.
- [45] Kneip, Li, Mac Gibbon and Ramsay (2000). Curve registration by local regression. *Canadian Journal of Statistics*, 28, 19-29.

- 
- [46] Kneip and Utikal (2001). Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, 96, 519-542.
- [47] Koenker and Basseett (1978). Regression quantiles. *Econometrica*, 46, 33-50.
- [48] Müller (2005). Functional modeling and classification of longitudinal data. *Scandinavian Journal of Statistics*, 32, 223-240.
- [49] McCullagh and Nelder (1989). *Generalized linear models (Second Edition)*. Monographs on Statistics and Applied Probability, Chapman and Hall, London.
- [50] Müller and Stadtmüller (2005). Generalized functional linear models. *Annals of Statistics*, 33, 774-805.
- [51] Nadaraya (1964) and Watson (1964). On estimating regression. *Theory of Probability and its Applications*, 10, 186-190.
- [52] Rao (1958) and Tucker (1958). Some statistical methods for comparison of growth curves. *Biometrics*, 14, 1-17.
- [53] Ramsay (1982). When the data are functions. *Psychometrika*, 47, 379-396.
- [54] Ramsay and Silverman (2002 et 2005). *Applied functional data analysis*. Springer, New York. *Functional data analysis (Second Edition)*. Springer, New York.
- [55] Ramsay and Li (1998). Curve registration. *Journal of the Royal Statistical Society, Series B*, 60, 351-363.
- [56] Ramsay and Dalzell (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society, Series B*, 53, 539-572.
- [57] Ruppert, Sheater and Wand (1993). An effective bandwidth selector for local least squares regression. Working paper, 93-017.

- 
- [58] Sarda and Vieu (2000). Kernel regression. In *Smoothing and Regression : Approches, Computation and Application*, M.G. Schimek editor, Wiley Series in Probability and Statistics, 43-70.
- [59] Van Huffel and Vandewalle (1991). *The total least squares problem : computational aspects and analysis*. SIAM, Philadelphia.
- [60] Vieux, (1995). Order choice in nonlinear autoregressive models. *Statistics* 26, 307-328.
- [61] Wedderburn (1974). Quasi-likelihood functions, generalized linear models, and the Gauss-Newton method. *Biometrika*, 61, 439-447.
- [62] Yao, Müller and Wang (2005a,2005b). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100, 577-590. Functional linear regression analysis for longitudinal data. *Annals of Statistics*, 33, 2873-2903.