Un point de vue sur des approches factorielles et probabilistes de la covariance. Application à l'analyse locale du mouvement

Sullivan Hidot

▶ **To cite this version:**

Sullivan Hidot. Un point de vue sur des approches factorielles et probabilistes de la covariance. Application à l'analyse locale du mouvement. Traitement du signal et de l'image [eess.SP]. Université de La Rochelle, 2007. Français. NNT : . tel-00203241

HAL Id: tel-00203241

https://theses.hal.science/tel-00203241

Submitted on 9 Jan 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

présentée à

L' UNIVERSITÉ DE LA ROCHELLE
École Doctorale ED0374

en vue de l'obtention du grade et du titre de

DOCTEUR de l'Université de **LA ROCHELLE**

spécialité **INFORMATIQUE**

au cours de la soutenance publique du 7 décembre 2007, par

Sullivan HIDOT

Intitulé de la thèse :

Un point de vue sur des approches factorielles et probabilistes de la covariance

Application à l'analyse locale du mouvement

COMPOSITION DU JURY

Président :	J. PAGÈS	**Agrocampus Rennes**
Rapporteurs :	A. BACCINI	**Université Paul Sabatier, Toulouse III**
	G. GOVAERT	**Université de Technologie de Compiègne**
Directeur :	J.-Y. LAFAYE	**Université de La Rochelle**
Co-encadrant :	C. SAINT-JEAN	**Université de La Rochelle**

Pour ma mère

Pour Claire-Lise et notre fils Eliott,
né le 20 juillet 2006

Remerciements

Tout d'abord, je tenais à remercier chaleureusement Jean-Yves Lafaye et Christophe Saint-Jean pour m'avoir suivi et encadré sur ce sujet passionnant tout au long de ces trois années. Ils ont su me laisser la liberté dont j'avais besoin tout en sachant me mettre la pression - pour ne pas dire un coup de pied aux fesses - quand il le fallait. Leur disponibilité, leur franchise et leur rigueur scientifique m'auront été très précieux et, j'en suis convaincu, me serviront encore à l'avenir.

J'adresse également mes sincères remerciements à Alain Baccini, Gérard Govaert et Jérôme Pagès pour m'avoir fait l'honneur de participer au jury et à l'évaluation de cette thèse.

Un immense merci à Carl Frélicot et Michel Berthier pour leurs coups de main, ce genre de coup de main bien utile à la fin d'une thèse.

Merci aux matheux "d'en haut" pour leur disponibilité et leurs conseils suite à toutes mes questions : Jean-Christophe Breton, Eric Benoît et Guy Wallet.

Je n'oublie pas non plus toutes les personnes que j'ai eu la chance de rencontrer durant ces quatre dernières années passées à La Rochelle : Olivier, Elodie, Thierry, Benjamin de la Cervoise Saint Ambroise, Grand Thomas, Aymeric, Armelle, Karell, Renaud, Muriel, Guylain, Hoël, Thomas, Romain ...

Je pense aussi à tous mes proches et amis de Ch'Nord, qui m'accueillent toujours avec joie dès que j'ai un peu de temps pour remonter dans le plus beau coin de France : Virginie, Julien, Aurore, Matthieu, Sophie, Maxime, Guillaume, Murat ...

"Et s'ils s'y égaraient, cela aussi était justifié, car pour trouver il fallait tout d'abord se perdre."
Christian Charrière
Extrait de La forêt d'Iscambe

Résumé

Cette thèse s'intéresse à des approches factorielles et probabilistes de la covariance qui tient compte d'une connaissance exogène sur les observations. Nous adoptons un modèle qui décompose le signal en une fonction déterministe du temps caractérisant la tendance, et en un terme résiduel. Les méthodes factorielles sont consacrées à l'étude du terme tendanciel. Nous présentons le formalisme général de la covariance relationnelle ainsi que de nouvelles propriétés qui éclairent les interprétations et faisons le lien avec les notions déjà existantes. La covariance relationnelle s'intègre dans l'analyse en composantes principales (ACP), l'analyse factorielle d'opérateurs et l'analyse discriminante d'opérateurs. Nous montrons que l'ACP relationnelle est un cas particulier de l'ACP à noyaux et de l'ACP fonctionnelle, dont nous dressons les schémas de dualité correspondants. L'étude du terme résiduel est menée à l'aide d'approches probabilistes fondées sur la covariance. Dans un premier temps, ce terme est assimilé à un vecteur gaussien et nous introduisons une procédure de classification de matrices de covariance par la distribution de Wishart induite par l'hypothèse de gaussianité. En particulier, l'algorithme EM sur matrices de covariance est proposé. Dans un second temps, on procède à l'analyse fractale du terme résiduel, identifié par une trajectoire d'un processus autosimilaire. L'indice d'autosimilarité est estimé quelque soit l'échantillonnage et nous déterminons dans quelle mesure cette contrainte temporelle influe sur l'estimation. Nous appliquons les concepts présentés à l'analyse du mouvement : corpus de mouvements de danse contemporaine (méthodes factorielles et classification par Wishart), et données de biologie marine (segmentation par analyse fractale).

Mots-clés : Covariance relationnelle, Analyse en composantes principales, Analyse factorielle d'opérateurs, Distribution de Wishart, Reconnaissance des formes, Classification, Mouvement brownien fractionnaire, Analyse du mouvement.

Abstract

This thesis deals with factorial and probabilistic approaches of the covariance, which can take an extraneous knowledge on the observations into account. Dealing with temporal data, the model we adopt breaks the signal into a deterministic function depending on time and representing a trend and into a residual term. Factorial methods are devoted to studying the trend. We present a general formalism for the relational covariance as well as new properties that clarify the interpretation and the connexion with existing concepts. Relational covariance is embedded into various paradigms such as principal component analysis (PCA), operator factor analysis and operator discriminant factor analysis. We show that relational PCA is a special case of Kernel PCA and Functional PCA and we illustrate that property through duality diagrams. The study of the residuals is carried out within the scope of probabilistic models based on covariance. At a first step, the residual term is assumed to be multi-Gaussian, hence we use a Wishart distribution for the empirical covariance, followed by a clustering procedure (EM algorithms). At a second step, we apply fractal analysis to the residual term that is modeled as a self similar process. The Hurst parameter is estimated for several sampling protocols and we study how the protocol specificity may influence this estimation. All these concepts are applied to movement analysis e.g. : motion capture of contemporary dance movements (factorial methods and clustering of Wishart mixture models), as well as marine biology (segmentation and fractal analysis).

Keywords : Relational covariance, Principal component analysis, Operator factor analysis, Wishart distribution, Pattern recognition, clustering, Fractional Brownian motion, Movement analysis.

Table des matières

1 Introduction générale **1**

1.1 Problématique et contexte 1
1.2 Modèle 3
1.3 Plan de la thèse 5

2 Covariance Relationnelle **9**

2.1 Introduction 9
2.2 Modèle probabiliste 10
2.2.1 Connaissance endogène 10
2.2.2 Connaissance exogène 12
2.3 Covariance et espérance relationnelles 13
2.3.1 Covariance relationnelle 13
2.3.2 Espérance relationnelle 15
2.4 Discrétisation 18
2.5 Approches connexes 20
2.5.1 Autres applications du laplacien 20
2.5.2 Géostatistique 22
2.6 Conclusion 24

3 Approches liées à la statistique géométrique **27**

3.1 Introduction 27
3.2 ACP relationnelle 28
3.2.1 ACP discrète et schéma de dualité 28
3.2.2 Introduction du laplacien 31
3.3 ACP locale 31
3.4 ACP fonctionnelle 32
3.4.1 Présentation de la méthode 33
3.4.2 ACP relationnelle et ACP fonctionnelle 36
3.5 ACP à noyaux 37
3.5.1 Présentation de la méthode 37
3.5.2 ACP relationnelle et ACP à noyaux 42
3.6 Analyse factorielle d'opérateurs 43

3.6.1 Analyse d'opérateurs de covariance relationnelle . . . 46
3.6.2 Analyse discriminante d'opérateurs de covariance relationnelle . 50
3.7 Conclusion . . . 53

4 Approches probabilistes **55**
4.1 Introduction . . . 55
4.2 Distribution de Wishart : modèle proposé . . . 56
4.2.1 Définition et propriétés . . . 56
4.2.2 Classification supervisée . . . 60
4.2.3 Classification non supervisée : algorithme EM pour matrices de covariance . . . 62
4.2.4 Difficultés en classification . . . 65
4.2.5 Sélection du modèle . . . 68
4.2.6 Validation des méthodes par simulation . . . 69
4.3 Analyse fractale . . . 70
4.3.1 Mouvement brownien fractionnaire : définition, propriétés . . . 71
4.3.2 Simulation de trajectoires échantillonnées irrégulièrement . . . 73
4.3.3 Extensions multidimensionnelles du MBF . . . 75
4.3.4 Grande dimension et échantillonnage . . . 79
4.3.5 Estimation de l'autosimilarité avec un échantillonnage irrégulier 82
4.3.6 Étude expérimentale de la précision et du biais de l'EMV . . . 83
4.3.7 Application à la segmentation multidimensionnelle . . . 91
4.4 Conclusion . . . 97

5 Applications **99**
5.1 Introduction . . . 99
5.2 Mise en place du formalisme . . . 100
5.2.1 Expression de la connaissance exogène . . . 100
5.2.2 Covariance et espérance relationnelles : exemples simples . . . 100
5.2.3 ACP relationnelle : exemples simples . . . 102
5.3 ACP relationnelle . . . 103
5.3.1 Présentation des données . . . 103
5.3.2 Représentations graphiques . . . 108
5.3.3 Reconstruction partielle . . . 112
5.4 Analyse factorielle d'opérateurs . . . 114
5.4.1 Choix des paramètres et interprétations . . . 114
5.4.2 Représentations graphiques . . . 115
5.4.3 Classification supervisée . . . 122
5.5 Modèle de Wishart . . . 126
5.5.1 Suppression de la tendance locale : interprétations . . . 127
5.5.2 Choix des paramètres . . . 128

5.5.3 Classification supervisée . . . 130
5.5.4 Classification non supervisée . . . 134
5.6 Analyse fractale . . . 136
5.6.1 Présentation des données . . . 136
5.6.2 Expression de la connaissance exogène . . . 137
5.6.3 Choix des paramètres . . . 137
5.6.4 Segmentation multidimensionnelle . . . 138
5.7 Conclusion . . . 142

6 Conclusion générale **145**
6.1 Synthèse . . . 145
6.2 Rappel des contributions . . . 148
6.3 Perspectives . . . 149

A Chapitre 2 **153**
A.1 Preuve du sens de la définition 3 . . . 153
A.2 Démonstrations des propriétés . . . 154
A.2.1 Propriété 2 . . . 154
A.2.2 Propriété 3 . . . 155
A.2.3 Propriété 4 . . . 156
A.2.4 Propriété 5 . . . 156
A.2.5 Propriété 6 . . . 157
A.3 Éléments de théorie des graphes . . . 157
A.4 Éléments de théorie des processus . . . 159
A.5 Éléments de théorie de l'estimation . . . 162

B Chapitre 3 **165**
B.1 Obtention des composantes principales d'une ACP . . . 165

Liste des tableaux

3.1 ACP fonctionnelle. 35
3.2 ACP à noyaux. 40
3.3 Analyse d'opérateurs de covariance relationnelle. 49
3.4 Analyse discriminante d'opérateurs de covariance relationnelle. 52

4.1 Classification supervisée par la loi de Wishart. 63
4.2 Algorithme EM pour matrices de covariance. 66
4.3 Exemple d'estimation de H par Whittle avec irrégularité d'échantillonnage. 84
4.4 Calcul du nombre minimum de simulations en fonction de la taille du signal. 85
4.5 Estimations par EMV avec régularité d'échantillonnage. 88
4.6 Estimations par EMV avec Irr_S, comparaison avec le tableau 4.5. Les cellules en gris clair (resp. foncé) correspondent au non rejet de H_0 pour le test de Fisher (resp. Fisher et Student). 88
4.7 Estimations par EMV avec Irr_U, comparaison avec le tableau 4.5. . . . 88
4.8 Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas régulier. Les cellules grisées correspondent au non-rejet du test (p-values $> \frac{\alpha}{2}$). 90
4.9 Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas Irr_S. 90
4.10 Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas Irr_U. 90
4.11 Algorithme de Viterbi pour la segmentation de série temporelle. 95
4.12 Procédure de segmentation pour une série temporelle. 96

5.1 Mouvements-type de danse contemporaine et effectif par groupe. 105
5.2 Part d'inertie des 4 premiers axes pour le mouvement Marche. 112
5.3 Moyenne des normes de Hilbert-Schmidt des instances de chaque classe de mouvement pour les 60 mouvements et $\theta = max$ (échelle : 10^5). . . . 117
5.4 Leave-one out pour l'analyse d'opérateurs de covariance relationnelle. . 124
5.5 Leave-one out pour l'analyse discriminante d'opérateurs. 125
5.6 Analyse d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs non normés ($m = 15$ premières projections et $C = 14$ classes). 127

5.7 Analyse d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs normés ($m = 15$ premières projections et $C = 14$ classes). 128
5.8 Analyse discriminante d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs normés ($m = 15$ premières projections et $C = 14$ classes). 129
5.9 Leave-one out pour la classification supervisée par la loi de Wishart. . . 131
5.10 Classification supervisée en supprimant la tendance en fonction du niveau de lissage. 132
5.11 Classification supervisée avec modèle de Wishart : taux de reconnaissance par leave-one out en fonction di niveau de lissage θ ($C = 14$ classes). . . 133
5.12 Classification non supervisée : moyennes et écarts-types des différences symétriques en fonction de K et de C (50 itérations). 134

Table des figures

1.1 Le modèle $X_t = \varphi(t) + \epsilon_t$ illustré avec deux exemples de simulation. - (a) Le phénomène présente une tendance prédominante (analyse à l'aide de méthodes factorielles). - (b) Le terme résiduel contient également de l'information (analyse à l'aide d'approches probabilistes). 3

2.1 La valeur propre nulle associée au vecteur propre de coordonnées identiquement égales à l'unité implique que l'image du laplacien est incluse dans l'orthogonal de ce vecteur. 20

3.1 Schéma de dualité de l'ACP. 30
3.2 Schéma de dualité de l'ACP locale. 32
3.3 Schéma de dualité de l'ACP fonctionnelle. 35
3.4 ACP à noyaux (source [SSM96]) : projeter les individus dans un espace de dimension supérieure et se ramener à une ACP standard. 37
3.5 Schéma de dualité de l'ACP à noyaux. 39
3.6 (a) Tableau $\mathfrak{X}$ à trois entrées ; (b) Deux sous-tableaux $\mathbb{X}_k$ et $\mathbb{X}_{k'}$ de taille respective n_k et $n_{k'}$; (c) Tableau $\mathfrak{X}$ reconstitué en tableau bidimensionnel. 45
3.7 Modèle Tucker3. 45
3.8 Principe de l'analyse discriminante d'opérateurs de covariance relationnelle. 51

4.1 Simulation et représentation bidimensionnelle de matrices de Wishart - (a) $n = 6$, 7, 8, 9 10 ; (b) $n = 30$; (c) $n = 100$. 60
4.2 Tirage uniforme de 100 points à deux résolutions différentes. Y'a-t-il un ou plusieurs clusters ? 68
4.3 Matrices d'échelle estimées par EM pour l'exemple de simulation. . . . 70
4.4 Tracés des courbes du critère BIC et de $-2\log(L(K))$ en fonction du nombre de clusters, pour l'exemple de simulation. 71
4.5 Trajectoires browniennes fractionnaires avec $N = 500$ simulées par la méthode de Cholesky pour H=0.01, 0.25, 0.5, 0.7, 0.9 et 0.99. 76
4.6 Matrice de covariance avec $N = 500$ pour H=0.01 ($\simeq$ bruit blanc), 0.25, 0.5 (mouvement brownien), 0.7, 0.9 et 0.99. Les valeurs sont comprises entre 0 (couleur noire) et 1 (couleur blanche). 77

4.7 Simulation de 6 trajectoires avec H=0.4. La variance des variables aléatoires échantillonnées augmentent avec le temps. 78
4.8 Trajectoire d'un processus brownien multifractionnaire avec sa fonction associée $H(t) = arctan(t)$ (source [PLV95]). 78
4.9 Trajectoires browniennes fractionnaires bidimensionnelles de longueur 500 simulées par la méthode de Cholesky pour H=0.01, 0.25, 0.5, 0.7, 0.9 et 0.99. 81
4.10 Stratégie de tests pour la comparaison et l'étude du biais de l'estimateur de H par EMV - $Stud_2$: Test de Student (Intervalle de confiance) ; $Fish+$ $Stud_1$: Test de Fisher + Test de Student (Comparaison d'échantillons). 86
4.11 Trajectoires browniennes fractionnaires bidimensionnelles de longueur $N = 500$ simulées par la méthode de Cholesky pour différentes valeurs de H en abscisse et en ordonnée. 92

5.1 (a) Série temporelle X, $\sigma(X) = 89.59$ et $\sigma_g(X) = 18.14$; (b) Série $\tilde{X}$: permutation aléatoire de X, $\sigma(\tilde{X}) = 89.59$ et $\sigma_g(\tilde{X}) = 82.75$, avec $\theta = 5$. 101
5.2 Représentation graphique de deux variables aléatoires décorrélées pour la covariance relationnelle ($\theta = 5$) et non décorrélées pour la covariance. 102
5.3 Disposition des observations (petits cercles) avec deux vecteurs propres : l'un issu d'une ACP standard (G) et l'autre d'une ACP relationnelle (L) avec $\theta = 5$. 103
5.4 Exemple sur un ensemble de trois classes : ACP standard (G), ACP relationnelle (L) et analyse discriminante (D). 104
5.5 Danseurs équipés du dispositif de capture. 105
5.6 Cyberdôme du CNBDI. 106
5.7 Structure du graphe ($\theta = 3$) pour une observation fixée t_i. 107
5.8 Logiciel de reconstitution 3D de mouvements dansés. 109
5.9 ACP standard : projection des variables aléatoires sur le plan 1×2 (mouvement Marche), inertie : 99.85%. 110
5.10 ACP relationnelle avec $\theta = 5$: projection des variables aléatoires sur le plan 1×2 (mouvement Marche), inertie : 88.96%. 111
5.11 ACP relationnelle avec $\theta = 5$: projection des individus sur le plan 1×2 (mouvement Marche). 112
5.12 ACP relationnelle avec $\theta = 5$: projection des individus sur le plan 1×2 (mouvement Tour). 113
5.13 Mouvement Tour (vue de dessus). 113
5.14 Reconstruction partielle d'un Tour avec le premier axe (vue de dessus) : perte de la circularité du mouvement. 114
5.15 Norme au carré de la différence entre le tableau initial et la reconstruction partielle des 2 premiers axes en fonction des valeurs de θ, mouvement Tour (a) et Marche (b). 114

5.16 Somme cumulée des parts d'inertie des opérateurs principaux d'une analyse d'opérateurs de covariance relationnelle ; (a) Opérateurs non normés et $\theta = max$; (b) Opérateurs normés et $\theta = max$; (c) Opérateurs normés et $\theta = 10$. 116
5.17 Représentations graphiques des deux premiers plans principaux d'une analyse d'opérateurs non normés pour $\theta = max$. 117
5.18 Représentations graphiques des quatre premiers plans principaux d'une analyse d'opérateurs normés pour $\theta = max$. 118
5.19 De gauche à droite, capture d'une Marche (mouvement à dominante horizontale) ; capture d'un Saut attitude (mouvement intermédiaire) ; capture d'un Enroulé bassin (mouvement à dominante verticale). 119
5.20 De gauche à droite, Suspension : position initiale et finale (déplacement faible du centre de gravité) ; Chute allongée : position initiale et finale (déplacement fort du centre de gravité). 119
5.21 Représentations graphiques des quatre premiers plans principaux d'une analyse d'opérateurs normés pour $\theta = 10$. 121
5.22 Valeurs propres d'une analyse discriminante d'opérateurs ; (a) Opérateurs non normés et $\theta = max$; (b) Opérateurs normés et $\theta = max$; (c) Opérateurs normés et $\theta = 10$. 122
5.23 À gauche : plan 2×3 d'une analyse d'opérateurs normés ($\theta = 10$, inertie : 24%) ; à droite : plan 1×2 d'une analyse discriminante d'opérateurs normés ($\theta = 10$, inertie : 16%). 123
5.24 Suppression de la tendance locale d'une variable à 56 réalisations, pour $\theta = 1$ et $\theta = max$. 130
5.25 Tracés d'une variable et de son espérance relationnelle ($\theta = 10$). 132
5.26 Graphique de la fonction BIC en fonction du nombre de clusters pour le lissage $\theta = 10$. 135
5.27 Bassin aquacole. 136
5.28 Déplacement de bar analysé ($N = 852$ captures). 137
5.29 Résultat de l'estimation des autosimilarités locales en abscisse et en ordonnée. 139
5.30 De gauche à droite, résultat de la segmentation pour la somme puis le minimum des estimations marginales de H. 140
5.31 Tracés des portions de déplacement en fonction des résultats de la segmentation. 140
5.32 De gauche à droite, résultat de la segmentation pour la somme des estimations marginales de H avec une capture sur deux, puis une sur trois. 141
5.33 Résultat de la segmentation pour la somme des estimations marginales de H sur les composantes principales. 142

A.1 (a) Graphe simple - (b) Graphe simple orienté - (c) Symétrisé. 158

A.2 Graphe symétrique. 159

Chapitre 1

Introduction générale

1.1 Problématique et contexte

Depuis plusieurs décennies, les applications des méthodes factorielles et probabilistes se sont multipliées dans des domaines de recherche variés tels que la biologie, les sciences de l'environment, le traitement d'images . . . Ce succès est dû au fait qu'elles permettent la modélisation d'une structure complexe ainsi qu'une description synthétique de données à grande dimension. Cette thèse expose un point de vue sur les méthodes factorielles et probabilistes de la covariance dans la prise en compte d'une connaissance exogène donnée a priori. Ces considérations ne sont bien sûr pas un sujet nouveau et l'on recense un grand nombre d'approches connexes. On citera par exemple l'analyse d'un cube spatio-temporel [CSC93], qui généralise l'analyse en composantes principales (ACP) à un multi-tableau de dimension trois en considérant à la fois une contiguïté spatiale et temporelle sur les observations. L'analyse spatiale multivariée [TCC02] étudie la structure locale et globale des données en y intégrant une variable caractérisant une proximité spatiale. La géostatistique [Cre93] utilise des méthodes d'estimation pour analyser des données géologiques dont la distribution spatiale est sans structure a priori. Il paraît important de pouvoir donner un formalisme général permettant d'unifier l'ensemble de ces notions.

Le cadre théorique de la thèse a été motivé par une recherche appliquée dans le domaine de l'analyse du mouvement. L'objet d'intérêt est modélisé par un vecteur aléatoire $X = (X_1, \ldots, X_p)$ dont les réalisations sont dépendantes du temps. Le mouvement est alors assimilé à un changement de positions décrites par un ensemble de coordonnées évoluant au cours du temps. Dans la réalité, l'objet est naturellement lié à des caractéristiques de différentes natures : biologique, spatiale ou temporelle. La caractéristique

biologique d'une part, concerne les aspects morphologiques influant sur l'acquisition et l'exécution du mouvement. Imaginons par exemple un système de capteurs équipant un humain. Dans sa globalité, le système d'acquisition constitue un squelette dont les capteurs sont également contraints par la morphologie de l'exécutant. En particulier, un capteur relié au poignet et un autre au coude seront toujours à égale distance, quelque soit le mouvement exécuté. De plus, un humain, et donc le système d'acquisition, sont toujours soumis à certaines lois physiques (gravité). Le critère spatial fait référence à toutes les caractéristiques géométriques et topologiques de l'espace qui influent naturellement sur le déplacement de l'objet : les frontières de l'espace, sa structure, les stimulus extérieurs, les relations de proximité, ou encore les interactions avec d'autres objets. Enfin, le système de capture impose de fournir des données discrétisées, et la nature temporelle du phénomène soulève des questions liées à l'échantillonnage : prise en compte de l'ordre des dates de captures, valeurs manquantes, fréquence d'acquisition non constante ... Il s'agit de mettre en avant ces caractéristiques dans l'analyse et/ou d'en enrichir le modèle. Nous nous intéressons en effet à des méthodes capables de prendre en compte et de mettre en valeur ce type d'information. Les domaines applicatifs sont la réduction de dimension, la classification, la reconstruction et la segmentation de mouvement. Les méthodes factorielles et probabilistes sont adaptées pour traiter ce type de question.

En modélisation statistique, il est de règle d'identifier les données par des réalisations de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{T}, P)$. La covariance est alors un outil de référence pour analyser les liaisons entre variables aléatoires. Elle caractérise les moments croisés centrés d'ordre deux et s'interprète comme une mesure de l'existence et du sens de la liaison linéaire entre les variables aléatoires. En tout état de cause, la covariance classique n'est pas l'outil adapté puisque la plupart des caractéristiques évoquées précédemment ne peuvent être prises en compte. Pour donner quelques illustrations, la covariance classique est indépendante de toute permutation entre les réalisations. En particulier, lors de l'étude conjointe de séries temporelles avec la covariance, l'ordre des mesures n'influe pas dans l'analyse, ce qui devrait normalement en constituer une information majeure. Le calcul de la covariance ne met en jeu que la valeur des réalisations des variables aléatoires et non les instants sur lesquels elles sont échantillonnées. Autrement dit, la covariance suppose systématiquement que l'échantillonnage soit régulier, et fait par conséquent disparaître la dynamique.

De manière générale, l'ajout d'une connaissance exogène paraît donc indispensable pour mener une étude plus réaliste. Initialement introduits par Lebart [Leb69] pour l'analyse de la contiguïté, les graphes de relation binaire permettent de modéliser la plupart de ces informations. On construit ainsi les covariances et variances associées qui tiennent compte des dépendances des observations suivant le graphe défini sur $\Omega \times \Omega$. La

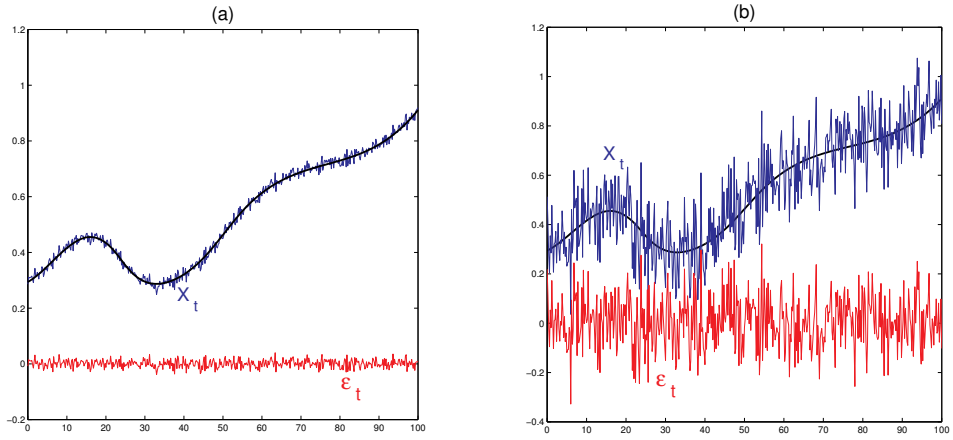

FIG. 1.1 – Le modèle $X_t = \varphi(t) + \epsilon_t$ illustré avec deux exemples de simulation. - (a) Le phénomène présente une tendance prédominante (analyse à l'aide de méthodes factorielles). - (b) Le terme résiduel contient également de l'information (analyse à l'aide d'approches probabilistes).

littérature scientifique recense un grand nombre de travaux sur ce sujet avec les notions suivantes : covariance partielle, locale, covariogramme croisé, semi-variogramme, variabilité ... Il s'agit de proposer une approche unificatrice reliant les méthodes factorielles et probabilistes. Le but de ce travail est également de déterminer dans quelle mesure l'ajout d'une connaissance exogène influe sur les principales approches.

1.2 Modèle

Nous menons notre étude en décomposant les données à l'aide du modèle suivant :

$$X_t = \varphi(t) + \epsilon_t$$

dans lequel $\varphi(t)$ est le terme *tendanciel*, correspondant à une fonction déterministe du temps et ϵ_t le terme *résiduel*, obtenu en retranchant le terme tendanciel du signal. Le modèle sera étudié selon deux points de vue. La figure 1.1 en donne une illustration avec deux exemples de simulation. En (a), le terme tendanciel $\varphi(t)$ est prédominant, le terme résiduel est quant à lui assimilé à du bruit. Dans ce cas, on analyse un phénomène avec tendance, et l'on traite le cas déterministe avec l'analyse par la covariance de $\varphi(t)$. L'approche employée est la statistique géométrique, qui s'appuie sur les méthodes factorielles et discriminantes. L'analyste souhaite en pratique avoir une représentation

synthétique et visible des données de manière à en faciliter l'interprétation. L'objectif des méthodes factorielles est de déterminer une représentation compacte de données multi-variées. Une des approches classiques de la statistique multidimensionnelle est l'analyse en composantes principales (ACP) [Jol86], qui extrait l'information pertinente des données en analysant la structure de la covariance à l'aide d'une décomposition en valeurs singulières. Parmi ses extensions, l'ACP fonctionnelle [RS97] s'applique dans le cas où les variables aléatoires sont continues et détermine une représentation synthétique des données à l'aide d'une base de fonctions choisies a priori. L'ACP à noyaux [SSM96], quant à elle, prend en compte les aspects non linéaires des liaisons entre les données, en projetant dans un espace de dimension supérieure pour ensuite se ramener à une ACP classique. La prise en compte dans l'ACP et ses variantes de n'importe quelle connaissance exogène modélisée par une relation binaire sera explicitée. L'analyse de tableaux à trois entrées est une extension naturelle des méthodes factorielles en dimension supérieure. L'une des applications est la reconstitution des données comme les modèles PARAFAC [Har70, Har72] et Tucker [KdL80], dans lesquels on décompose un tableau à trois dimensions à l'aide de tableaux bidimensionnels. L'analyse par extraction de signature considère la covariance comme une signature des données afin de les plonger dans le même espace et ainsi en étudier les relations existantes. Là encore, nous étudions comment intégrer la connaissance exogène en proposant des approches variantes de l'analyse factorielle d'opérateurs [Fou84] et de STATIS [Esc80, Esc85].

La figure 1.1 (b) présente un autre exemple où le terme ϵ_t contient de l'information difficilement interprétable avec les méthodes factorielles. Ce terme est assimilé à un vecteur gaussien dont l'étude sera menée à l'aide d'approches probabilistes. Dans un premier temps, le modèle sur la covariance est posé a priori, et l'on s'intéresse à l'estimation de la matrice de covariance. L'hypothèse de gaussianité de ϵ_t induit une distribution de Wishart pour la matrice de covariance. Cette distribution caractérise les matrices de covariance empiriques d'échantillons de vecteurs aléatoires gaussiens. La loi de Wishart sera utilisée dans le cadre de la classification. Dans le cas non supervisé, la procédure d'estimation par l'algorithme Expectation-Maximisation (EM) des paramètres des clusters sera présentée. Dans un second temps, on procède à l'analyse fractale du terme résiduel. ϵ_t s'identifie comme une trajectoire d'un mouvement brownien fractionnaire. On s'intéresse à l'estimation de l'autosimilarité par la covariance. En effet, la covariance tient compte de manière intrinsèque de n'importe quel type d'échantillonnage. L'indice d'autosimilarité peut alors être estimé quelque soit l'échantillonnage et il convient de déterminer dans quelle mesure cette contrainte temporelle influe sur l'estimation de l'autosimilarité.

1.3 Plan de la thèse

Cette thèse se découpe en six chapitres. Le chapitre 2 présente un formalisme général de la covariance qui tient compte de n'importe quel type de relation binaire intégrée a priori dans le modèle. Nous proposons pour cela l'appellation générique *covariance relationnelle*. La covariance relationnelle fait appel à la notion de graphe d'une relation binaire pour lequel nous définissons l'espérance relationnelle associée. Un certain nombre de propriétés liées à la covariance et l'espérance relationnelle seront introduites afin de faciliter et de mettre en valeur de nouvelles interprétations. Nous passons enfin en revue les différentes approches connexes : clustering spectral, géostatistique et réduction de dimension.

Le chapitre 3 explicite l'intégration de la covariance relationnelle dans l'analyse en composantes principales (ACP), et fait également le lien avec les méthodes à noyaux et l'ACP fonctionnelle dont l'ACP relationnelle est un cas particulier. Nous dressons les schémas de dualité associés qui éclairent une correspondance entre les espaces d'analyse. Nous étudions également l'ajout de la signature de covariance relationnelle dans les méthodes d'analyse de tableaux multiples : analyse d'opérateurs étendue également à l'analyse discriminante.

Nous passons au chapitre 4 à l'étude du terme ϵ_t sous l'hypothèse de gaussianité en adoptant deux points de vue. D'une part, nous présentons une nouvelle application de la loi de Wishart à la reconnaissance et la classification de matrices de covariance. Nous proposons de fait l'algorithme EM pour matrices de covariance et évoquons les contraintes du modèle. D'autre part, nous étudions l'influence de l'échantillonnage irrégulier dans l'estimation de l'autosimilarité pour des signaux de taille faible à l'aide de covariance. Nous proposons un protocole expérimental pour l'étude de l'estimateur associé. Une méthode de segmentation multidimensionnelle fondée sur la signature de l'autosimilarité est également proposée.

Les exemples applicatifs des techniques précédentes seront étudiés au chapitre 5. Les techniques proposées sont alors mises en œuvre en considérant différentes contraintes liées à la nature spatio-temporelle des données. Le contexte applicatif est consacré à l'analyse locale du mouvement. D'une manière générale, les données expérimentales sont exploitées à des fins exploratoires par des techniques de segmentation, de reconnaissance de formes, de reconstruction, de compression et/ou de classification. Le terme "local" signifie que l'on considère une relation binaire caractérisant les proximités temporelles entre les observations. Nous appliquons les méthodes factorielles et la classification par Wishart à un corpus de mouvements de danse contemporaine. L'analyse fractale du

terme résiduel est illustrée à l'aide de données de biologie marine. Nous étudions le déplacement 2D d'un poisson bar évoluant dans un bassin aquacole. Ce mouvement présente des valeurs manquantes avec un échantillonnage irrégulier.

Enfin, le chapitre 6 constitue la conclusion générale et les perspectives de ce travail.

Publications concernant ce travail

S. Hidot, C. Saint-Jean et J.-Y. Lafaye, *Étude expérimentale de l'influence d'un échantillonnage irrégulier dans l'estimation du paramètre de Hurst.* À paraître dans la **Revue de Statistique Appliquée**.

S. Hidot, C. Saint-Jean et J.-Y. Lafaye, *Classification de signaux multidimensionnels utilisant la distribution de Wishart : Application à la reconnaissance de mouvements.* Actes du Groupement de Recherche en Traitement du Signal et des Images **GRETSI**, Troyes, 2007.

S. Hidot, J.-Y. Lafaye et C. Saint-Jean, *Propriétés et interprétation de la covariance relationnelle en ACP.* **Revue Traitement du Signal**, 24(1) : 29-37, 2007.

S. Hidot, J.-Y. Lafaye et C. Saint-Jean, *Discriminant factor analysis for movement recognition : application to dance.* Proceedings of the International Conference on Computer Vision and Graphics **ICCVG'06**, Special issue in Journal of Machine Graphics and Vision, 15(3-4) : 391-399, 2006, Warsaw, Poland.

S. Hidot, J.-Y. Lafaye et C. Saint-Jean, *ACP relationnelle pour l'analyse du mouvement : Application à la danse.* Congrès Reconnaissance des Formes et Intelligence Artificielle **RFIA**, p.64, 9 pages, Tours, 2006.

S. Hidot, J.-Y. Lafaye et C. Saint-Jean, *Analyse factorielle d'opérateurs pour l'étude du mouvement : Application à la danse.* Atelier "Fouille de données", Journées francophones Extraction et Gestion des Connaissances **EGC**, pages 21-30, Lille, 2006.

Chapitre 2

Covariance Relationnelle

2.1 Introduction

En analyse statistique des données, on se place dans un espace probabilisé $(\Omega, \mathcal{T}, P)$ où les données sont modélisées par des réalisations de variables aléatoires réelles définies sur Ω. L'espace vectoriel des variables aléatoires est alors muni d'un produit scalaire particulier pour lequel les notions géométriques de base se traduisent par des interprétations statistiques : angle pour corrélation, projection pour régression ... À ce titre, la covariance permet d'exprimer l'information endogène sur les liaisons entre les variables. C'est l'un des outils les plus utilisés dans les méthodes statistiques. Mathématiquement, la covariance est un moment centré du second ordre et représente le produit scalaire entre les variables aléatoires centrées.

Dans la pratique, on peut imposer l'ajout d'une connaissance exogène qui influe sur la structure des données analysées. De manière générale, cette information exogène se ramène le plus souvent à l'expression d'une relation binaire entre les observations et l'on parle alors d'analyse de variables *instrumentales*, c'est à dire dépendantes d'une variable supplémentaire [Sab87]. En tout état de cause, cette connaissance a priori n'est pas prise en compte par la covariance. Dans ce chapitre nous étudions la covariance induite par une relation binaire entre les observations, que nous rebaptisons sous l'appellation générique *covariance relationnelle*. Le terme "relationnelle" a été choisi car nous considérons dans notre formalisme n'importe quelle relation binaire de référence. Dès lors, l'ajout d'une telle relation a nécessairement un impact sur le calcul de la covariance puisque celle-ci est modifiée en conséquence. Cette relation binaire est représentée par un graphe de connexité entre les observations. Nous proposons une formalisation afin de mettre en évidence les propriétés ayant de fortes implications dans la pratique et

notamment pour l'étude de données spatio-temporelles [HLSJ07], qui est à la base de ce travail. La covariance relationnelle induit la notion d'*espérance relationnelle*, analogue à l'espérance définie conditionnellement à la variable aléatoire du graphe de la relation binaire. Nous proposons ensuite un ensemble de réécritures des paramètres usuels manipulés tels que inertie, variance, covariance ou moyenne. Chaque formulation permet de mettre en évidence et d'appréhender une interprétation particulière de la quantité concernée. Ces interprétations utilisent successivement des considérations probabilistes ou géométriques.

Dans une première section, nous détaillons le modèle probabiliste dans lequel s'expriment les différentes notions apportées par la connaissance exogène, c'est à dire par la structure fixée a priori des observations. Dans un deuxième temps, nous définissons la covariance et l'espérance relationnelles et formulons l'ensemble des propriétés sous-jacentes. En section 3, nous précisons l'étape de discrétisation et évoquons l'emploi du laplacien dans les approches de réduction de dimension, clustering et analyse spectrale de graphe. En section 4, nous présentons les applications du laplacien en clustering spectrale et faisons un lien entre la covariance relationnelle et des approches connexes de type probabiliste et géométrique. Nous terminons par la conclusion en section 5.

2.2 Modèle probabiliste

2.2.1 Connaissance endogène

Nous disposons d'un espace probabilisé $(\Omega, \mathcal{T}, P)$, où Ω est un ensemble dont les éléments sont appelés *individus*, $\mathcal{T}$ une tribu sur Ω et P une mesure de probabilité sur Ω. Une *variable aléatoire réelle* est une application définie sur Ω :

$$X \ : \ \Omega \to \mathbb{R} \tag{2.1}$$

où les observations sont réalisées de manière aléatoire. Le produit scalaire pour la mesure P entre les variables aléatoires X et Y est défini de la manière suivante :

$$< X, Y >= \int_{\Omega} X(\omega)Y(\omega)dP(\omega) \tag{2.2}$$

Les variables aléatoires sont supposées de carré intégrable :

$$< X, X >= \|X\|^2 = \int_{\Omega} |X(\omega)|^2 dP(\omega) < +\infty \tag{2.3}$$

Pour $\omega \in \Omega$ fixé, $X(\omega)$ représente la réalisation sur Ω de la variable aléatoire X. On note par $\mathbf{1}$ la variable constante unitaire, *i.e.* $\mathbf{1}(\omega) = 1$ pour tout $\omega \in \Omega$. L'opérateur

linéaire de l'espérance appliquée à X est donnée par la formule :

$$E(X) =< X, \mathbf{1} >= \int_{\Omega} X(\omega) dP(\omega) \tag{2.4}$$

L'opération de centrage d'une variable aléatoire revient à lui retrancher son espérance :

$$X_c = X - E(X) \tag{2.5}$$

Définition 1 (Covariance) *Sans connaissance a priori sur l'espace Ω, la covariance entre deux variables aléatoires X et Y est donnée par le produit scalaire des variables centrées :*

$$Cov(X,Y) =< X_c, Y_c >= \int_{\Omega} (X(\omega) - E(X))(Y(\omega) - E(Y)) dP(\omega) \tag{2.6}$$

La variance et l'écart-type associés sont les suivants :

$$Var(X) =< X_c, X_c > \quad \text{et} \quad \sigma_X = \sqrt{Var(X)} = \|X_c\| \tag{2.7}$$

Une variable aléatoire *réduite* est obtenue en la divisant par son écart-type. La propriété suivante exprime la covariance en fonction d'une mesure produit sur $\Omega \times \Omega$:

Propriété 1 *La covariance entre deux variables aléatoires centrées X et Y se calcule en utilisant la formule quadratique :*

$$Cov(X,Y) = \frac{1}{2} \iint_{\Omega \times \Omega} (X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) dP(\omega) dP(\omega') \tag{2.8}$$

Cette expression de la covariance a déjà été proposée dans sa version discrétisée [LMP04]. Bien sûr, cette expression coïncide avec (2.6), d'après la formule de Huygens :

$$\begin{aligned} &\frac{1}{2} \iint_{\Omega \times \Omega} (X(\omega) - X(\omega'))(Y(\omega) - Y(\omega')) dP(\omega) dP(\omega') \\ &= \frac{1}{2} (2 < X, Y > -2E(X)E(Y)) =< X, Y > -E(X)E(Y) =< X_c, Y_c > \end{aligned} \tag{2.9}$$

La formule de la covariance est exprimée par une intégrale double définie sur $\Omega \times \Omega$, ce qui, comme on le verra, permet l'ajout d'une connaissance exogène dont l'expression générale concerne les couples d'individus.

2.2.2 Connaissance exogène

L'espace Ω est sans structure autre que probabiliste. Dans la réalité, nous disposons souvent d'informations complémentaires sur les liaisons entre les individus telles que proximité spatiale, temporelle, sociale ... Cette information est apportée par une variable "exogène" ordinale (pour une relation d'ordre totale ou de pré-ordre) ou catégorielle (comme une fonction indicatrice d'appartenance relativement à une partition des données). Une relation binaire est une forme générale pour représenter cette connaissance a priori. L'information relationnelle exogène s'exprime au travers d'un graphe orienté[1] défini de $\Omega \times \Omega$ à valeurs dans $\{0, 1\}$. Cette relation est directement introduite dans le modèle probabiliste initial à l'aide du son graphe qui exprime les liens entre les individus. Le graphe est induit par une variable g définie sur $\Omega \times \Omega$:

$$g(\omega, \omega') = \begin{cases} 1 & \text{si } \omega \text{ est connecté à } \omega' \\ 0 & \text{sinon} \end{cases} \tag{2.10}$$

Nous définissons le *degré*[2] comme une application de Ω dans $[0, 1]$ qui à $\omega \in \Omega$ associe la somme des poids des arêtes de g reliées à ω :

$$deg(\omega) = \int_{\Omega} g(\omega, \omega') dP(\omega') \leq P(\Omega) = 1 \tag{2.11}$$

La *connectivité* du graphe est la somme des degrés pour la mesure P :

$$\kappa = \int_{\Omega} deg(\omega) dP(\omega) \leq P(\Omega) = 1 \tag{2.12}$$

Soient $\mathcal{A}$ et $\mathcal{D}$ les opérateurs de $\mathcal{L}^2(\Omega)$ vers son dual $\mathcal{L}^2(\Omega)^*$:

$$\mathcal{A}(X).Y = \iint_{\Omega \times \Omega} X(\omega) Y(\omega') g(\omega, \omega') dP(\omega) dP(\omega') \tag{2.13}$$

$$\mathcal{D}(X).Y = \int_{\Omega} X(\omega) Y(\omega) deg(\omega) dP(\omega) \tag{2.14}$$

Définition 2 (Laplacien du graphe) *L'opérateur du laplacien du graphe Δ de $\mathcal{L}^2(\Omega)$ vers $\mathcal{L}^2(\Omega)^*$ est défini comme la différence entre les opérateurs $\mathcal{D}$ et $\mathcal{A}$ [Moh91] :*

$$\Delta = \mathcal{D} - \mathcal{A} \tag{2.15}$$

[1] On trouvera en Annexe de ce chapitre un rappel sur les notions de base de théorie des graphes.

[2] Habituellement, le degré d'un sommet est défini comme un entier égal au nombre d'arêtes entrantes et sortantes [Leb69]. Notre approche est analogue à Le Foll [Fol82] puisque nous rajoutons une pondération aux arêtes du graphe correspondant au produit des poids de leurs sommets.

L'opérateur Δ s'exprime alors de la manière suivante :

$$\Delta(X).Y = \mathcal{D}(X).Y - \mathcal{A}(X).Y \tag{2.16}$$

Comme on le verra dans la suite, cette notion peut s'utiliser dans le cas continu ou discret.

2.3 Covariance et espérance relationnelles

2.3.1 Covariance relationnelle

En calculant via le graphe g, la covariance sous sa forme quadratique s'exprime de la manière suivante :

$$Cov_g(X,Y) = \frac{1}{2\kappa}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega)dP(\omega') \tag{2.17}$$

Cette formule met en jeu les couples d'individus en relation pour le graphe. Notons que si le graphe de la relation binaire est complet, alors le résultat est la covariance classique puisque l'on retrouve la formule (2.8). Sinon, seuls interviennent dans la formule les couples d'individus en relation pour g. Le terme $dP(\omega)dP(\omega')$ est une pondération implicite de l'arête du graphe reliant ω et ω'. Par ailleurs, nous avons la relation :

$$\iint_{\Omega\times\Omega} dP(\omega)dP(\omega') = 1 \tag{2.18}$$

On introduit donc le facteur de normalisation induit par la connectivité du graphe qui permet d'avoir la somme des pondérations des arêtes du graphe égale à 1 :

$$\frac{1}{\kappa}\iint_{\Omega\times\Omega} g(\omega,\omega')dP(\omega)dP(\omega') = 1 \tag{2.19}$$

La covariance induite par g s'exprime de manière synthétique à l'aide du laplacien du graphe :

Définition 3 (Covariance relationnelle) *La covariance relationnelle entre deux variables aléatoires X et Y se calcule en utilisant la formule quadratique :*

$$Cov_g(X,Y) = \frac{1}{\kappa}\Delta(X).Y =: \frac{1}{\kappa} < X,Y >_\Delta \tag{2.20}$$

On montre[3] que si le graphe est symétrique, alors la définition a bien un sens.

À l'origine, la covariance relationnelle est plus connue sous le nom de covariance locale [Gea54, Mor48] dans le cadre de l'analyse de la contiguïté. De même, la covariance relationnelle s'apparente à la notion de covariance partielle [Rao64] qui exprime la covariance en tenant compte d'une variable aléatoire exogène à valeurs fixées. La définition ci-dessus présente un cadre général qui englobe un large ensemble de relations et permet de déduire un certain nombre de propriétés [HLSJ07]. Pour certaines d'entres elles, nous faisons référence aux auteurs qui en ont naturellement inspiré le formalisme.

Propriété 2 *La covariance relationnelle utilise implicitement une version symétrisée du graphe g.*

En effet, celle-ci n'agit pas sur un graphe particulier mais plutôt sur sa classe d'équivalence, où deux graphes g_1 et g_2 sont équivalents si les symétrisés coïncident :

$$g_1 \sim g_2 \Longleftrightarrow (g_1 + \widetilde{g}_1 = g_2 + \widetilde{g}_2) \tag{2.21}$$

où le graphe $\widetilde{g}$ est le *symétrique* de g :

$$\forall \omega, \omega' \in \Omega\,, \quad \widetilde{g}(\omega, \omega') = g(\omega', \omega) \tag{2.22}$$

Les covariances relationnelles d'un graphe quelconque et de son symétrique sont égales :

$$Cov_g(X, Y) = Cov_{\widetilde{g}}(X, Y) \tag{2.23}$$

On montre également que ces covariances coïncident avec celle du graphe valué symétrique g^* suivant :

$$g^*(\omega, \omega') = \frac{1}{2}\left(g(\omega, \omega') + \widetilde{g}(\omega, \omega')\right) \tag{2.24}$$

Ce graphe peut prendre trois valeurs : 0, 1 ou 1/2. Le cas où $g^*(\omega, \omega') = 0$ signifie que ω et ω' ne sont pas connectés pour g, celui où $g^*(\omega, \omega') = 1$ signifie que ω et ω' sont connectés pour g par deux arcs opposés, et $g^*(\omega, \omega') = 1/2$ signifie que ω et ω' sont connectés par un arc unique. Enfin, la formule (2.17) montre clairement que les arêtes réflexives n'interviennent pas dans le calcul de la covariance relationnelle. Cette propriété est fondamentale puisqu'elle permet de considérer un graphe de relation binaire symétrique, sans aucune perte de généralité.

[3] On trouvera en Annexe de ce chapitre la preuve du sens de cette définition ainsi que de l'ensemble des propriétés concernant la covariance et l'espérance relationnelles.

Propriété 3 *En notant par Δ_1 le laplacien du graphe complet, on obtient la décomposition suivante :*

$$< X, Y >_{\Delta_1} = < X, Y >_{\Delta} + < X, Y >_{\Delta_1 - \Delta} \tag{2.25}$$

Autrement dit :

$$Cov(X, Y) = \kappa Cov_g(X, Y) + (1 - \kappa) Cov_{\overline{g}}(X, Y) \tag{2.26}$$

où $\overline{g}$ est le graphe complémentaire défini par :

$$\forall \omega, \omega' \in \Omega\ , \quad \overline{g}(\omega, \omega') = 1 - g(\omega, \omega') \tag{2.27}$$

De la formule (2.26), on en déduit si $X = Y$:

$$Var(X) = \kappa Var_g(X) + (1 - \kappa) Var_{\overline{g}}(X) \tag{2.28}$$

La variance totale s'exprime en fonction de la variance relationnelle à laquelle s'ajoute le terme résiduel $(1-\kappa)Var_{\overline{g}}(X)$. Celui-ci s'interprète comme une variance via le graphe complémentaire de g. L'opérateur $\Delta_1 - \Delta$ induit un produit scalaire puisqu'il possède les mêmes propriétés que Δ.

2.3.2 Espérance relationnelle

Nous allons définir la notion d'espérance, conditionnellement à un graphe de connexité afin de donner une interprétation complémentaire de la covariance relationnelle.

Définition 4 (Espérance relationnelle) *Posons pour tout $\omega \in \Omega$:*

$$E(X|g)(\omega) = \frac{1}{deg(\omega)} \int_{\Omega} X(\omega') g(\omega, \omega') dP(\omega') \tag{2.29}$$

$E(X|g)(\omega)$ est appelée espérance relationnelle de la variable aléatoire X mesurée sur le voisinage de ω selon le graphe g.

L'espérance relationnelle $E(X|g)$ est une variable aléatoire de $\mathcal{L}^2(\Omega)$ qui associe à tout individu la moyenne de la variable aléatoire X calculée sur ses voisins selon g. Nous nous plaçons bien dans le cadre d'une espérance conditionnelle dont chaque valeur dépend des réalisations de la variable aléatoire g définie sur $\Omega \times \Omega$. Il est clair que $E(X|g)$ est bien définie si $deg(\omega)$ est non nul pour tout ω, ce qui impose que tout individu possède

au moins un arc sortant. Dans ce cas, on vérifie aisément l'existence d'une telle variable aléatoire en utilisant l'inégalité de Cauchy-Schwarz :

$$\begin{aligned} E(X|g)(\omega) &\leq |E(X|g)(\omega)| \\ &\leq \frac{1}{deg(\omega)} \left(\int_\Omega |X(\omega')|^2 dP(\omega') \right)^{1/2} \left(\int_\Omega |g(\omega,\omega')|^2 dP(\omega') \right)^{1/2} \\ &< +\infty \end{aligned} \tag{2.30}$$

De manière générale, $E(.|g)$ ne peut s'interpréter comme une espérance ordinaire par défaut d'idempotence :

$$E(E(X|g)|g) \neq E(X|g) \tag{2.31}$$

Ainsi, l'application linéaire $E(.|g)$ n'est pas un projecteur et ne vérifie pas la propriété d'orthogonalité d'une espérance conditionnelle ordinaire, et l'on a en général :

$$< E(X|g), X - E(X|g) > \neq 0 \tag{2.32}$$

Par contre, si g est une relation d'équivalence (*i.e.* : partition), chaque individu est connecté pour le graphe à l'ensemble des éléments de sa classe, alors l'espérance relationnelle coïncide exactement avec l'espérance conditionnelle et la covariance (resp. variance) relationnelle s'interprète comme la covariance (resp. variance) intra-classe.

Propriété 4 *Soit $< ., . >_Q$ le produit scalaire défini par la densité de probabilité Q :*

$$Q(\omega) = \frac{P(\omega) deg(\omega)}{\kappa} \tag{2.33}$$

La covariance relationnelle entre X et Y est égale au produit scalaire pour Q de X avec l'écart de Y à son espérance relationnelle :

$$Cov_g(X,Y) = \frac{1}{\kappa} < X, Y >_\Delta = < X, Y - E(Y|g) >_Q \tag{2.34}$$

où

$$< X, Y >_Q = \int_\Omega X(\omega) Y(\omega) dQ(\omega) = \frac{1}{\kappa} \int_\Omega X(\omega) Y(\omega) deg(\omega) dP(\omega) \tag{2.35}$$

Par linéarité, (2.34) implique que dans le cas où la variable aléatoire X est Q-centrée, la variance de X pour Q se décompose comme suit :

$$Var_Q(X) = Var_g(X) + < X, E(X|g) >_Q \tag{2.36}$$

La variance totale d'une variable aléatoire s'exprime comme la somme de sa variance relationnelle et de sa variabilité globale (appelée aussi *autocovariance spatiale*). Pour

un graphe définissant une proximité, l'autocovariance spatiale mesure la similarité entre les individus voisins. Le terme $< X, E(X|g) >_Q$ est le produit scalaire entre la variable aléatoire X et sa moyenne locale dans le graphe et permet de caractériser la dispersion locale de X. Ce terme n'est pas considéré comme une variance car il peut éventuellement présenter des valeurs négatives. Thioulouse *et al.* [TCC02] appliquent cette notion dans le cas discret pour analyser la structure locale et globale de dispersions d'oiseaux sur des zones géographiques données. Avec cette formule appliquée à des données adaptées, il est possible de mesurer l'indépendance spatiale en utilisant l'indice de Geary [Gea54] et de Moran [Mor48]. L'indépendance spatiale signifie que les individus ne sont soumis à aucune liaison spatiale. En particulier, cela garantit de pouvoir utiliser les tests statistiques usuels où l'on suppose l'indépendance des tirages.

L'indice de Moran ou *autocorrélation spatiale* s'apparente à un coefficient de corrélation et se définit comme le quotient entre variabilité globale et variance totale :

$$I_{Moran} = \frac{< X, E(X|g) >_Q}{Var_Q(X)} \in [0, 1] \tag{2.37}$$

Il permet de mesurer la ressemblance entre les individus en relation pour le graphe. L'indice de Geary, complémentaire, est le rapport entre variance locale et variance totale :

$$I_{Geary} = \frac{Var_g(X)}{Var_Q(X)} = 1 - I_{Moran} \tag{2.38}$$

Si l'indice de Geary est proche de 1, alors les individus n'ont aucune relation pour le graphe. Ils sont disposés de manière aléatoire, sans aucune structure dans l'espace.

Propriété 5 *L'opérateur $E(.|g)$ est auto-adjoint pour $< ., . >_Q$:*

$$< E(X|g), Y >_Q = < X, E(Y|g) >_Q \tag{2.39}$$

La covariance relationnelle étant symétrique, la formule (2.34) s'applique de la même manière à $Cov_g(X, Y)$ ou $Cov_g(Y, X)$ et l'on en déduit :

$$< X, Y - E(Y|g) >_Q = < X - E(X|g), Y >_Q \tag{2.40}$$

Le produit scalaire pour Q entre l'espérance relationnelle de la variable X (resp. Y) et Y (resp. X) donne le même résultat. Cette relation est similaire à celle de la covariance qui s'exprime avec le centrage d'une seule variable :

$$\begin{aligned} Cov(X, Y) &= < X, Y - E(Y) > = < X - E(X), Y > \\ &= < X - E(X), Y - E(Y) > \end{aligned} \tag{2.41}$$

Avec l'usage d'un graphe, nous avons l'analogie, mais avec le centrage par rapport à l'espérance relationnelle, sans pour autant avoir l'égalité avec le produit scalaire des variables toutes deux centrées, d'après (2.32). Autrement dit :

$$\begin{aligned} Cov_g(X,Y) &= < X, Y - E(Y|g) >_Q = < X - E(X|g), Y >_Q \\ &\neq < X - E(X|g), Y - E(Y|g) >_Q \end{aligned} \tag{2.42}$$

Propriété 6 *La moyenne de l'espérance relationnelle d'une variable aléatoire est égale à son espérance totale pour Q :*

$$E_Q(E(X|g)) = E_Q(X) \tag{2.43}$$

Cette propriété est analogue au théorème de l'espérance totale qui met en jeu l'espérance d'une variable aléatoire conditionnellement à une autre. Dans notre cas, l'espérance est liée à la pondération des individus ainsi qu'à leur liaison pour le graphe.

2.4 Discrétisation

Pour mettre en œuvre les techniques précédentes, il convient d'obtenir un échantillon des variables aléatoires analysées. Nous passons ainsi à une formulation discrète des questions évoquées plus haut dans un cadre continu. Dans ce cas, l'espace Ω est un ensemble fini de taille n et définit les variables aléatoires discrétisées sur $\mathcal{L}^2(\{\omega_1, ..., \omega_n\})$, identifiable à $\mathbb{R}^n$. Soient $(X_1, \ldots, X_p)$ un vecteur aléatoire à p dimensions, $\mathbb{X}$ le tableau $n \times p$ contenant un n-échantillon du vecteur aléatoire, et P une mesure associant une pondération $(P_{\omega_i})_{i=1}^n$ aux individus. On note par D la matrice diagonale $n \times n$ contenant les poids des individus. Cette matrice constitue une métrique sur l'espace engendré par les variables aléatoires. L'expérience aléatoire conduit à la fois à l'obtention des n réalisations du vecteur aléatoire et à celle du graphe de la relation g. On note par G la *matrice d'adjacence* de taille $n \times n$ associée au graphe g :

$$G_{ij} = \begin{cases} 1 & \text{si } \omega_i \text{ est connecté à } \omega_j \\ 0 & \text{sinon} \end{cases} \tag{2.44}$$

Les opérateurs $\mathcal{A}$ et $\mathcal{D}$ qui conduisent à l'expression du laplacien Δ ont des formes matricielles symétriques $(n \times n)$ et conservent les propriétés générales précédemment

étudiées. $\mathcal{A}$ est la *matrice d'adjacence pondérée* de G exprimant la relation binaire induite par la matrice G :

$$\mathcal{A}_{ij} = P_{\omega_i} P_{\omega_j} G_{ij} \ , \quad 1 \leq i, j \leq n \tag{2.45}$$

La *matrice des degrés* $\mathcal{D}$ est diagonale et contient la somme des éléments de chaque ligne de $\mathcal{A}$:

$$\mathcal{D}_{ii} = \sum_{j=1}^{n} \mathcal{A}_{ij} = P_{\omega_i}.deg(\omega_i) \tag{2.46}$$

où $deg(\omega_i)$ est la somme discrétisée des poids des individus dont les individus sont connectés à i pour G :

$$deg(\omega_i) = \sum_{j=1}^{n} P_{\omega_j} G_{ij} \tag{2.47}$$

La formule (2.17) s'écrit sous sa version discrétisée :

$$Cov_G(X, Y) = \frac{1}{2\kappa} \sum_{i,j=1}^{n} (X(\omega_i) - X(\omega_j))(Y(\omega_i) - Y(\omega_j)) G_{ij} P_{\omega_i} P_{\omega_j} \tag{2.48}$$

où la connectivité s'exprime dans sa version discrétisée :

$$\kappa = \sum_{i=1}^{n} P_{\omega_i} deg(\omega_i) = trace(\mathcal{D}) \tag{2.49}$$

À l'aide des versions discrétisées de $\mathcal{A}$ et $\mathcal{D}$, nous exprimons la covariance relationnelle sous forme matricielle [Leb69] :

$$V_G = \frac{1}{\kappa} {}^t\mathbb{X}(\mathcal{D} - \mathcal{A})\mathbb{X} = \frac{1}{\kappa} {}^t\mathbb{X}\Delta\mathbb{X} \tag{2.50}$$

Δ est une matrice symétrique semi-définie positive d'ordre n et de rang $n-1$ si le graphe est connexe [Bol98]. En général, si le graphe a k composantes connexes, alors le rang du laplacien du graphe sera $n-k$. De plus, $\mathcal{K}er(\Delta)$ est l'espace engendré par la variable aléatoire constante unitaire. Il est donc inutile de centrer préalablement les variables aléatoires pour déterminer la matrice de covariance relationnelle :

$$Cov_g(X_c, Y_c) = Cov_g(X, Y) \tag{2.51}$$

Dans la propriété 4, le Q-centrage n'a d'importance que pour le terme de droite qui met en jeu $< ., . >_Q$. Toutes les variables aléatoires sont donc projetées dans l'orthogonal[4] de la variable aléatoire constante unitaire (cf fig. 2.1) :

$$\Delta\mathbf{1} = (\mathcal{D} - \mathcal{A})\mathbf{1} = \mathcal{D}\mathbf{1} - \mathcal{A}\mathbf{1} = 0 = 0\mathbf{1} \tag{2.52}$$

Par ailleurs, les variables aléatoires correspondant aux vecteurs propres de Δ forment une base orthonormale de l'espace des individus et donc, la somme des réalisations des vecteurs propres non constants de Δ est toujours nulle, par orthogonalité avec le vecteur propre constant égal à 1.

[4]Dans l'espace euclidien $\mathbb{R}^n$ muni du produit scalaire usuel.

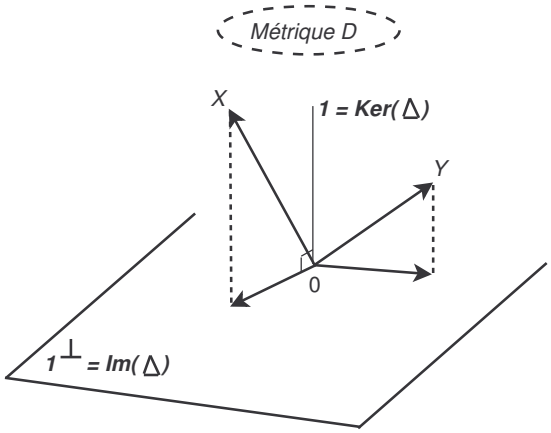

FIG. 2.1 – La valeur propre nulle associée au vecteur propre de coordonnées identiquement égales à l'unité implique que l'image du laplacien est incluse dans l'orthogonal de ce vecteur.

2.5 Approches connexes

La prise en compte de relations topologiques ou géographiques par exemple entre les individus a donné lieu à de nombreux développements. Dans cette section, nous proposons de préciser les applications du laplacien du graphe (réduction de dimension et clustering spectral) et d'évoquer les liens avec le domaine de la géostatistique[5].

2.5.1 Autres applications du laplacien

La métrique du laplacien a été appliquée efficacement dans des contextes variés tels que le clustering et le partitionnement de graphe. Dans tous les cas, on fait appel à sa décomposition spectrale et l'on trouvera à ce sujet un ensemble de propriétés dans [AM85, Moh91]. L'analyse lissée [Esc89] consiste à supprimer les variations locales en considérant le barycentre des individus préalablement équipondérées et reliées pour le graphe. Mathématiquement, cela revient à analyser le tableau $\mathcal{D}^{-1}\mathcal{A}\mathbb{X}$ au lieu de $\mathbb{X}$. L'analyse des différences locales [Esc89] se rapproche de l'analyse locale en supprimant le barycentre des individus où les individus sont connectées, et considère donc le tableau $\mathbb{X} - \mathcal{D}^{-1}\mathcal{A}\mathbb{X}$ au lieu de $\mathbb{X}$. L'analyse lissée et l'analyse des différences locales sont deux approches complémentaires qui généralisent l'analyse intra et inter-classe.

[5]Voir en Annexe de ce chapitre pour les rappels concernant la théorie des processus et de l'estimation.

Fiedler [Fie73] étudie les propriétés de connexité d'un graphe par l'intermédiaire de la plus petite valeur propre non nulle du laplacien, qu'il appelle *connectivité algébrique*. Il montre que ce nombre est strictement positif si et seulement si le graphe est connexe. Plus récemment [ST96], la connectivité algébrique est appelée *valeur de Fiedler* et son vecteur propre associé *vecteur de Fiedler*. La valeur de Fiedler permet de déterminer la segmentation optimale par coupe d'un graphe, en le scindant en deux sous-graphes disjoints. Pour cela, on détermine une certaine valeur de coupe du graphe. En faisant correspondre les sommets du graphe avec les éléments du vecteur de Fiedler, on partitionne le graphe suivant que ces éléments sont inférieurs ou non à cette valeur.

Dans la méthode laplacian eigenmaps [BN02], le laplacien approxime l'opérateur généralisé de Laplace Beltrami, qui permet un plongement dans un espace de dimension réduite $m < n$ des individus. La méthode est robuste, simple dans sa mise en place puisqu'elle ne nécessite au final qu'une décomposition spectrale, et permet également la préservation de la géométrie locale des points d'intérêt. Le plongement est caractérisé par un n-échantillon d'un vecteur aléatoire $Y = (Y_1, \ldots, Y_m)$ de dimension m. Le vecteur Y est fondé sur un problème de minimisation d'une fonctionnelle mettant en jeu la matrice d'adjacence. Les variables aléatoires du plongement Y minimisent la relation $\frac{1}{2}\sum_{i,j=1}^{n}(Y_{ki} - Y_{kj})^2\mathcal{A}_{ij}$ [BN03]. En développant, on retrouve l'expression matricielle de la covariance relationnelle. Pour k allant de 1 à m :

$$\begin{aligned}
\frac{1}{2}\sum_{i,j=1}^{n}(Y_{ki} - Y_{kj})^2\mathcal{A}_{ij} &= \frac{1}{2}\left(\sum_{i,j=1}^{n} Y_{ki}^2\mathcal{A}_{ij} + \sum_{i,j=1}^{n} Y_{kj}^2\mathcal{A}_{ij} - 2\sum_{i,j=1}^{n} Y_{ki}Y_{kj}\mathcal{A}_{ij}\right) \\
&= \frac{1}{2}\left(\sum_{i=1}^{n} Y_{ki}^2\sum_{j=1}^{n}\mathcal{A}_{ij} + \sum_{j=1}^{n} Y_{kj}^2\sum_{i=1}^{n}\mathcal{A}_{ij} - 2\sum_{i,j=1}^{n} Y_{ki}Y_{kj}\mathcal{A}_{ij}\right) \\
&= \frac{1}{2}\left(\sum_{i=1}^{n} Y_{ki}^2\mathcal{D}(i,i) + \sum_{j=1}^{n} Y_{kj}^2\mathcal{D}(j,j) - 2\sum_{i,j=1}^{n} Y_{ki}Y_{kj}\mathcal{A}_{ij}\right) \\
&= {}^tY_k\mathcal{D}Y_k - {}^tY_k\mathcal{A}Y_k \\
&= {}^tY_k\Delta Y_k
\end{aligned} \tag{2.53}$$

Avec une contrainte de normalisation ${}^tY_kDY_k = 1$, le problème est donc équivalent à déterminer les vecteurs propres de $\Delta Y_k = \lambda DY_k$ correspondant aux m plus petites valeurs propres λ [BN03]. Par ailleurs :

$$\Delta Y_k = \mathcal{D}Y_k - \mathcal{A}Y_k = \lambda\mathcal{D}Y_k \Longleftrightarrow \mathcal{A}Y_k = (1-\lambda)\mathcal{D}Y_k\,, \quad 1 \leq k \leq m \tag{2.54}$$

Cela revient donc à calculer les vecteurs propres de $\mathcal{D}^{-1}\mathcal{A}$ correspondant aux m plus grandes valeurs propres. Deux variantes de graphe sont proposées. L'une est fondée sur une proximité spatiale contrôlée par un paramètre ϵ, et l'autre sur une indicatrice d'appartenance aux k plus proches voisins :

$$G_{ij}^{\epsilon} = \mathbb{1}_{\{\|x_i - x_j\|^2 < \epsilon\}}(i,j) \tag{2.55}$$

ou bien

$$G_{ij}^{k} = \begin{cases} 1 & \text{si } i \text{ est l'un des } k \text{ plus proches voisins de } j \\ 0 & \text{sinon} \end{cases} \tag{2.56}$$

Le premier graphe est analogue à la transformation de Sammon [Sam69] et attribue un poids d'autant plus important aux distances faibles. Pour la matrice d'adjacence, deux variantes sont également proposées. Soit une relation simple sans pondération :

$$\mathcal{A}_{ij} = G_{ij} \tag{2.57}$$

ou soit définie par un noyau gaussien :

$$\mathcal{A}_{ij} = \exp\left\{ -\frac{\|x_i - x_j\|^2}{\sigma^2} \right\} \tag{2.58}$$

où σ est l'écart-type à fixer contrôlant l'étendue de la relation. La méthode laplacian eigenmaps considère une pondération des arêtes du graphe de la forme $P_{\omega_i\omega_j}$ et non plus $P_{\omega_i}P_{\omega_j}$. Il s'agit en fait d'un cas général pour lequel la notion de degré définie en (2.11) et donc d'espérance relationnelle n'a plus de sens puisque il n'est plus possible d'attribuer un poids aux sommets du graphe. Par conséquent, seules subsistent les propriétés 2 et 3 concernant le choix arbitraire d'un graphe symétrique et la décomposition de la variance, le laplacien du graphe restant quant à lui symétrique semi-défini positif par construction.

De la même manière, le clustering spectral [Wei99, NJW01] consiste à segmenter l'ensemble des individus à l'aide d'un partitionnement de graphe. Le problème est de déterminer les éléments propres d'une matrice caractérisant une proximité spatiale. Le principe est le même que la méthode laplacian eigenmaps, et la matrice d'adjacence est similaire à celle de la formule (2.58). La seule différence réside dans la définition du graphe, qui est moins restrictive puisqu'on impose uniquement la non réflexivité des arêtes :

$$G_{ij} = 1 - \delta_{ij} \tag{2.59}$$

En ajoutant un lien de proximité spatiale, les résultats sont meilleurs que ceux obtenus avec la méthode des centres mobiles appliquée aux données brutes.

2.5.2 Géostatistique

La géostatistique, développée par Matheron [Mat65], est le domaine se situant entre les statistiques et la géologie. Le but est d'analyser les corrélations spatiales entre variables aléatoires modélisées par un processus aléatoire gaussien $\{Z(s)/s \in \Omega \subset \mathbb{R}^d\}$ et

à prédire certaines valeurs à partir de points où les valeurs de mesure sont connues. En pratique, d prend trois valeurs : 1, 2 ou 3 suivant la dimension du "terrain" étudié. Les variables analysées sont dites *régionalisées*, ce qui sous-entend que les valeurs de mesure ne sont pas indépendantes de leur localisation spatiale. Par exemple, il est fort probable que la température prélevée sur un terrain après écoulement de lave soit sensiblement la même à un point qu'à ses points voisins. Les différentes étapes que le géostatisticien doit respecter dans son analyse et dans son protocole sont les suivantes [Cre93] :

- Trouver une zone d'échantillonnage.
- Faire un résumé qualitatif ou quantitatif des données sur l'échantillon, c'est à dire effectuer les mesures.
- Détecter et admettre l'hypothèse de stationnarité pour et par l'emploi du variogramme.
- En déduire, par krigeage, les valeurs des mesures en des points inconnus (n'appartenant pas à la zone d'échantillonnage).

En géostatistique, nous mesurons les propriétés structurales d'une série par l'intermédiaire du variogramme. C'est l'outil de base et également la statistique la plus utilisée dans le cadre de l'estimation. Il donne une indication sur la variabilité spatiale du phénomène étudié. Son emploi nécessite l'*hypothèse intrinsèque*, fondamentale en géostatistique, à savoir que les accroissements du processus sont stationnaires en espace de moyennes nulles. Ainsi, la variance des accroissements ne dépend que de l'écart entre les variables aléatoires. Formellement, cette hypothèse se formule par deux assertions :

$$\forall h \in \Omega\,, \quad E(Z_{s+h} - Z_s) = 0 \quad \text{et} \quad var(Z_{s+h} - Z_s) = 2\gamma(h) \tag{2.60}$$

Les quantités 2γ et γ sont appelées respectivement *variogramme* et *semivariogramme*. Le variogramme peut être de type *isotropique* ou *anisotropique*, suivant qu'il dépende de la norme (cas isotropique) et en plus de la direction de h (cas anisotropique). Les analyses locales fondées sur le variogramme [Mat65, Kri51] établissent les prémisses des méthodes factorielles dédiées à l'analyse de données géologiques. Comme pour les fonctions à noyaux, plusieurs appellations peuvent être employées suivant la définition de γ : linéaire, sphérique, exponentiel, ... Si le processus Z est observé sur un ensemble discret $\{s_1, ..., s_n\}$ régulièrement espacé de h, Matheron [Mat65] propose une estimation classique du variogramme. En définissant $G_h = \{(s_i, s_j) \in \Omega \times \Omega | s_i - s_j = h\}$, cet estimateur s'exprime sous la forme :

$$\widehat{\gamma}(h) = \frac{1}{2|G_h|} \sum_{i,j=1}^{n} (Z(s_i) - Z(s_j))(Z(s_i) - Z(s_j))G_h(i,j) \tag{2.61}$$

où $|G_h|$ désigne le nombre d'éléments distincts de G_h. Pour h fixé, on retrouve l'expression de la variance relationnelle empirique sans pondération calculée à partir de

la formule (2.48). L'ensemble G_h est un graphe non symétrique qui fait du semivariogramme un cas particulier de la variance relationnelle empirique avec équipondération des individus. L'estimateur de $\gamma(h)$ est asymptotiquement non biaisé et sensible à la présence de données aberrantes [HC84].

On peut également citer d'autres outils de la géostatistique liés à la covariance relationnelle. Si Z est un processus stationnaire, la fonction d'autocovariance (ou *covariogramme*) C se définit par :

$$C(h) = cov(Z_{s+h}, Z_s) \tag{2.62}$$

La formule suivante relie le variogramme et le covariogramme [Cre93, p.67] :

$$\gamma(h) = C(0) + C(h) \tag{2.63}$$

Sous l'hypothèse intrinsèque, cette formule est analogue à (2.36) concernant la décomposition de la variance totale. Dans le cadre de l'estimation, ce résultat devient faux :

$$\widehat{\gamma}(h) \neq \widehat{C}(0) + \widehat{C}(h) \tag{2.64}$$

Néanmoins, on démontre que la différence entre les deux termes est d'autant plus petite que $|G_h|/n$ est proche de 1 [Cre93]. Le variogramme est utilisé dans le krigeage [Cre93], méthode d'interpolation spatiale qui estime la valeur d'un phénomène naturel en des sites non observés. Notre approche est différente car elle n'utilise pas le variogramme comme un outil mais comme un objet d'étude apportant de l'information sur l'interaction et la dispersion des phénomènes.

2.6 Conclusion

Dans ce chapitre, nous avons présenté le formalisme de la covariance relationnelle qui permet de prendre en compte une information exogène de type relationnel dans l'étude de la liaison entre variables aléatoires. La covariance relationnelle est invariante quelque soit le centrage préalable des variables aléatoires. C'est un produit scalaire dans l'orthogonal de l'espace des variables constantes. Elle utilise implicitement une version symétrisé du graphe de connexité, qui autorise donc de travailler sans perte de généralité avec le graphe de la relation binaire symétrique, et sous-entend également la notion d'espérance relationnelle dont la définition nécessite des contraintes de connexité du graphe. Sans avoir la totalité de ses propriétés, cet opérateur présente de fortes similitudes avec la notion classique d'espérance conditionnelle. Au delà des propriétés générales, valables quelle que soit la relation étudiée, des propriétés spécifiques

peuvent être proposées lorsque la relation exogène étudiée est d'une forme particulière. En particulier, le choix d'une relation binaire caractérisant une partition entre les individus amène naturellement à considérer, comme en analyse discriminante, les notions de covariances intra et inter-classes. Quelque soit la relation étudiée, il est clair que la covariance relationnelle présente un cadre très général en incluant des outils statistiques connus. C'est une covariance à part entière qui s'utilise donc dans n'importe quelle méthode d'analyse du second ordre.

Chapitre 3

Approches liées à la statistique géométrique

3.1 Introduction

Dans ce chapitre, nous considérons l'introduction naturelle du laplacien dans l'analyse en composantes principales (ACP) [Jol86] et dans les méthodes d'analyse de tableaux multiples fondées sur la covariance. L'ACP standard se place dans le contexte d'un espace probabilisé, et optimise des critères d'inertie pour rendre compte des covariances entre variables et des distances mutuelles entre individus dans des espaces de dimension réduite. Nous proposons l'appellation ACP *relationnelle* qui analyse un vecteur aléatoire conditionnellement à la donnée d'une relation binaire sur les individus. Formellement, le concept est le même que l'ACP où l'on adopte le laplacien du graphe de la relation comme la métrique sur l'espace des variables. Les extensions dans le cas fonctionnel (ACP fonctionnelle [RS97]) et dans celui d'une non linéarité entre les liaisons (ACP à noyaux [SSM96]) seront également vues en détail.

Dans le cadre de l'analyse de tableaux multiples, nous introduisons également deux méthodes d'analyse de tableaux à trois entrées, appelées respectivement *analyse d'opérateurs de covariance relationnelle* [HLSJ06b] et *analyse discriminante d'opérateurs de covariance relationnelle* [HLSJ06c]. L'analyse de tableaux multiples considère deux approches de traitement : l'une pratique une décomposition synthétique des données, comme les modèles PARAFAC [Har70, Har72] et Tucker [KdL80], l'autre mesure les liaisons entre les sous-tableaux par extraction de signature, comme l'analyse factorielle multiple [EP98], l'analyse canonique généralisée [Hor61], STATIS [Esc80, Esc85] ou l'analyse procrustéenne [Tuc58, Gow75]. Les deux méthodes que nous proposons s'ins-

crivent dans cette seconde approche et peuvent être considérées comme variantes de la méthode STATIS appliquée à un ensemble d'opérateurs de covariance relationnelle.

En section 2, nous présentons le formalisme de l'ACP relationnelle et montrons en section 3 comment cette méthode peut être vue comme un cas particulier d'ACP fonctionnelle et d'ACP à noyaux, dont nous dressons le schéma de dualité correspondant, ce qui permet d'établir le lien avec l'ACP standard. Nous introduisons en section 4 les deux méthodes d'analyse factorielle multi-tableaux tenant compte d'une relation binaire exogène. Nous terminons par la conclusion en section 5.

3.2 ACP relationnelle

3.2.1 ACP discrète et schéma de dualité

Du fait de sa simplicité et de sa facilité d'implémentation, l'ACP [Hot33, Jol86] est l'approche la plus connue en analyse factorielle de données. On détermine une nouvelle base orthonormale pour laquelle la géométrie des individus et des variables est représentée de façon optimale. Dans ce paragraphe, nous présentons un inventaire des adaptations de l'ACP suivant la nature des données traitées : cas fini (ACP discrète), continu (ACP fonctionnelle) et non linéaire (ACP à noyaux).

Soit $\mathbb{X} = (X_1, \ldots, X_p)$ un n-échantillon d'un vecteur aléatoire de dimension p, représenté par un tableau de taille $n \times p$ dont les colonnes sont les réalisations des p variables aléatoires et les lignes les n individus x_i représentés dans la base canonique de $\mathbb{R}^p$. Dans le modèle, on introduit une matrice M qui représente la métrique sur l'espace des individus et qui peut s'interpréter comme un système de pondérations des variables aléatoires et de leurs interactions d'ordre deux. De même, l'espace des variables est muni de la métrique diagonale des poids des individus notée D. Les variables $X_1, \ldots, X_p$ sont supposées centrées pour D :

$$ {}^t\mathbb{X} D \mathbf{1} = 0 \in \mathbb{R}^p \tag{3.1}$$

Le cas centré est le cas général, bien que dans certaines situations, on puisse décider et justifier de réaliser une ACP non centrée. On analyse alors les moments d'ordre deux et non plus la covariance. L'ACP permet une dualité en construisant d'une part de nouvelles variables aléatoires synthétiques appelées *composantes principales* (CP), qui maximisent les covariances entre les variables aléatoires. D'autre part, l'ACP maximise l'inertie du nuage des individus projetés à l'aide des *axes principaux*.

L'ACP du triplet statistique $(\mathbb{X}, M, D)$ s'exprime de la façon suivante. La matrice de covariance empirique V entre les variables aléatoires est donnée par la relation :

$$V = {}^t\mathbb{X}D\mathbb{X} \in \mathbb{R}^{p\times p} \tag{3.2}$$

La matrice VM est l'*opérateur de covariance* du vecteur $\mathbb{X}$. Soient U et Λ respectivement les vecteurs propres M-normés à l'unité et les valeurs propres associées de VM. Les vecteurs propres sont portés par les axes principaux. On a les relations suivantes[1] :

$$VMU = U\Lambda \quad \text{et} \quad C = \mathbb{X}MU \tag{3.3}$$

Les CP sont D-orthogonales et D-normés à $\Lambda^{1/2}$. Les vecteurs propres U forment une base de l'espace des individus et les CP une base de l'espace des variables aléatoires. Par construction, la projection M-orthogonale d'un individu x_i sur le $j$ème vecteur propre est égale à la $i$ème observation de la $j$ème CP :

$$< x_i, U_j >_M = {}^t x_i M U_j = (\mathbb{X}MU)_{ij} = C_{ij} = C_j(\omega_i) \tag{3.4}$$

La coordonnée de la variable aléatoire X_i sur la CP réduite $C_j/\sqrt{\lambda_j}$ est égale au produit de la $i$ème coordonnée du $j$ème vecteur propre par la racine carrée de la valeur propre associée :

$$\begin{aligned} \frac{1}{\sqrt{\lambda_j}} < X_i, C_j >_D &= \frac{1}{\sqrt{\lambda_j}} < \mathbb{X}e_i, \mathbb{X}MU_j >_D = \frac{1}{\sqrt{\lambda_j}} {}^t e_i {}^t\mathbb{X}D\mathbb{X}MU_j \\ &= \frac{1}{\sqrt{\lambda_j}} {}^t e_i VMU_j = \frac{1}{\sqrt{\lambda_j}} {}^t e_i \lambda_j U_j = \sqrt{\lambda_j}\, {}^t e_i U_j \\ &= \sqrt{\lambda_j} U_{ij} \end{aligned} \tag{3.5}$$

où e_i est le $i$ème vecteur de la base canonique de $\mathbb{R}^n$. D est considérée comme la matrice d'une application linéaire définie par $(D(Y))(X_i) =< Y, X_i >_D$, pour Y un vecteur fixé de $\mathbb{R}^n$. La matrice M définit l'application $(M(y))(x_i) =< y, x_i >_M$, pour y un vecteur fixé de $\mathbb{R}^p$.
En multipliant la première égalité de (3.3) par M, on obtient :

$$MV(MU) = (MU)\Lambda \tag{3.6}$$

On voit donc que $U^* = MU$ est l'ensemble des vecteurs propres de MV associés aux mêmes valeurs propres que VM. Ces vecteurs, appelés *facteurs principaux*, sont M^{-1}-orthonormés :

$${}^tU^*M^{-1}U^* = {}^t(MU)M^{-1}MU = {}^tUMU = I_p \tag{3.7}$$

Remarquons que $\mathbb{X}U^* = \mathbb{X}MU = C$. Par dualité, on a également les deux égalités suivantes :

$$WDC = C\Lambda \quad \text{et} \quad U = {}^t\mathbb{X}DC\Lambda^{-1} \tag{3.8}$$

[1]Voir en Annexe de ce chapitre pour la solution du problème d'obtention des CP.

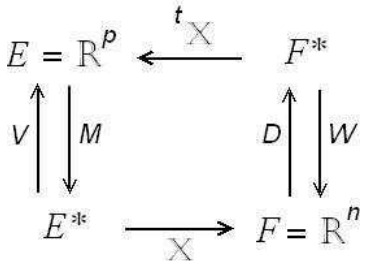

FIG. 3.1 – Schéma de dualité de l'ACP.

où :

$$W = \mathbb{X} M {}^t\mathbb{X} \in \mathbb{R}^{n \times n} \tag{3.9}$$

Multiplions la première égalité par D :

$$DW(DC) = (DC)\Lambda \tag{3.10}$$

De même, $C^* = DC$ est la matrice des vecteurs propres de DW de valeurs propres Λ. Ces vecteurs sont D^{-1}-normés. On appelle les vecteurs de C^* les *cofacteurs principaux*. Le contexte de l'ACP du triplet $(\mathbb{X}, M, D)$ peut être représenté par le *schéma de dualité* de la figure 3.1 [CP76].

L'ACP connaît beaucoup de variantes dont voici les principales. L'ACP canonique consiste à prendre la métrique canonique sur les individus. Comme $M = I_p$, l'inertie se confond alors avec la variance des variables aléatoires. L'ACP normée analyse les corrélations en normant les variables aléatoires par leur écart-type. L'ACP pondérée [Fol82] insiste sur l'introduction d'une pondération des couples d'individus parallèlement à la pondération classique. L'ACP lissée [BE90] applique préalablement un filtrage de type moyenne mobile ou différences premières des variables analysées. On citera aussi l'ACP de Wartenberg [War85] qui considère une contiguïté spatiale tout en levant la contrainte de matrice définie positive. Wartenberg interprète les valeurs propres négatives comme un poids supplémentaire à l'interaction spatiale (covariance) plutôt qu'à la structure (variance). Plusieurs autres méthodes factorielles sont considérées comme des cas particuliers de l'ACP munie d'une métrique spécifique. L'analyse des correspondances peut être vue comme une double-ACP (sur lignes et colonnes) avec la métrique du χ^2. L'analyse discriminante est une ACP sur les centres de gravité des classes avec la métrique de Mahalanobis. Nous retenons de toutes ces déclinaisons (pour lesquelles un exposé théorique est donnée dans [Leb69]) qu'elles ont en commun d'appliquer la démarche classique de l'ACP, en introduisant dans le modèle initial une information de type relationnel. Toutes ces méthodes apparaissent comme des cas particuliers de l'analyse canonique ou de l'ACP avec variables instrumentales, qui recherche des combinaisons linéaires des variables aléatoires sous des contraintes issues de variables exogènes.

3.2.2 Introduction du laplacien

Il apparaît que la spécificité de l'analyse relationnelle réside effectivement dans les propriétés de la covariance relationnelle et non pas dans celles de l'ACP [HLSJ07]. Plus précisément, une fois que l'on s'est convaincu que la covariance relationnelle est une covariance à part entière, l'ACP s'applique clairement de façon classique au triplet statistique $(\mathbb{X}, M, \frac{1}{\kappa}\Delta)$ avec toutes ses propriétés usuelles. L'ACP relationnelle est une ACP classique dans laquelle on introduit le laplacien Δ en tant que métrique sur l'espace des variables aléatoires. Les aspects formels sont identiques mais les interprétations sont spécifiques.

Notons que dans l'espace des individus, le critère au sens des moindres carrés des distances calculées avec la métrique M implique que l'on peut passer des axes principaux d'une ACP relationnelle à ceux d'une ACP standard par une simple transformation (naturellement non orthogonale). Ce passage du global au relationnel met en jeu les coordonnées respectives des vecteurs propres (les lignes de la matrice U) dans la base canonique.

Nous passons maintenant en revue les variantes de l'ACP pour lesquelles nous précisons les liens avec la covariance relationnelle et/ou la manière d'introduire la métrique du laplacien dans le modèle. Précisément, nous explicitons l'ACP locale, l'ACP fonctionnelle et l'ACP à noyaux, dont l'ACP relationnelle s'avère être un cas particulier.

3.3 ACP locale

L'ACP locale, introduite par Aluja-Banet *et al.* [ABNT91, ABL84], est similaire à l'ACP relationnelle puisqu'elle est également basée sur un graphe d'une relation binaire entre les individus. On considère la matrice, notée T, qui croise les arcs avec les sommets du graphe :

$$T(i,j) = \begin{cases} 1 & \text{si } j \text{ est un sommet origine de l'arc } i \\ -1 & \text{si } j \text{ est un sommet final de l'arc } i \\ 0 & \text{sinon} \end{cases} \tag{3.11}$$

La matrice T est de taille $n^2 \times n$ car le graphe complet possède n^2 arcs pour n sommets. Soit maintenant L la métrique sur $\mathbb{R}^{n^2}$ définie comme le produit de Kronecker de la

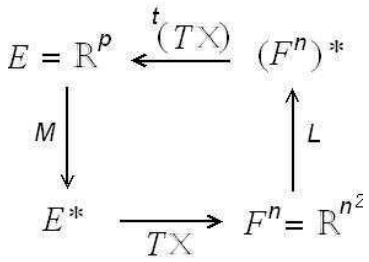

FIG. 3.2 – Schéma de dualité de l'ACP locale.

matrice D des poids des individus par elle-même :

$$L = D \otimes D = \begin{pmatrix} D(1,1)D & \cdots & D(1,n)D \\ \vdots & \ddots & \vdots \\ D(n,1)D & \cdots & D(n,n)D \end{pmatrix} \in \mathbb{R}^{n^2 \times n^2} \tag{3.12}$$

L'ACP locale consiste à faire une décomposition en valeurs singulières de $T\mathbb{X}$ relativement aux métriques M et L, ce qui conduit naturellement à la diagonalisation de la matrice ${}^t(T\mathbb{X})L(T\mathbb{X})M$. Pour une équipondération des individus et la métrique M identité, cette matrice coïncide avec celle de la covariance relationnelle établie à la formule (2.50) [Leb69]. Dans le cas d'un graphe, nous ajoutons la pondération des individus à la formule établie dans [ABNT91] mettant en jeu la matrice T et le laplacien. En notant par $diag(D)$ le vecteur colonne correspondant à la diagonale de D, on a la formule :

$$\Delta = \left(\frac{1}{2}{}^tTT\right) \circ diag(D)\,{}^t diag(D) \tag{3.13}$$

où $\circ$ désigne le produit matriciel de Hadamard (terme à terme). L'ACP locale travaille avec le schéma de dualité illustré à la figure 3.2.

Comme l'ACP relationnelle, l'ACP locale exploite un graphe d'une relation binaire entre les individus. Néanmoins, son usage n'est pas commode en pratique, puisqu'elle nécessite de manipuler des matrices d'ordre n^2. Néanmoins, on peut travailler par dualité dans l'espace $\mathbb{R}^p$.

3.4 ACP fonctionnelle

Habituellement, les données dites fonctionnelles sur lesquelles on pratique l'ACP se modélisent par p variables aléatoires pour lesquelles chaque réalisation est une courbe et non plus un nombre réel comme l'ACP discrète [RS97]. Dans le cas présent, nous

considérons que les variables aléatoires et non plus les individus sont des fonctions continues. L'approche classique [RS97] de l'ACP fonctionnelle formule une discrétisation du problème par l'intermédiaire de la décomposition dans une base de fonctions $< \phi_1, ..., \phi_m >$. Soient :

- $\mathbf{X} = (X_1, \ldots, X_p)$ le vecteur aléatoire de p variables aléatoires centrées.
- $\phi = (\phi_1, ..., \phi_m)$ le vecteur colonne des m fonctions de base.
- $\mathcal{W}_\phi$ la matrice carrée d'ordre m des produits scalaires des éléments de ϕ.
- $\mathbf{C}$ la matrice $p \times m$ des coefficients des décompositions des X_i dans ϕ.

3.4.1 Présentation de la méthode

Le problème consiste à trouver un ensemble de p fonctions, notées f_j $(1 \leq j \leq p)$ et appelées *fonctions propres* (FP), qui maximisent la somme des carrés des covariances entre les FP et les X_i :

$$\sum_{i=1}^{p} < f_j, X_i >^2 = \sum_{i=1}^{p} \left(\int_\Omega f_j(\omega).X_i(\omega) dP(\omega) \right)^2 \tag{3.14}$$

Les FP forment une base orthonormale du sous-espace vectoriel engendré par les X_i :

$$< f_j, f_{j'} > = \int_\Omega f_j(\omega).f_{j'}(\omega) dP(\omega) = \delta_{jj'} \tag{3.15}$$

En tant qu'éléments de l'espace des fonctions, les FP se décomposent également en somme pondérée de fonctions de base. Soit $u' = {}^t(u'_1, ..., u'_m)$ les coefficients de la décomposition d'une FP f_j dans ϕ :

$$\forall \omega \in \Omega\,, \quad f_j(\omega) = \sum_{k=1}^{m} u'_k \phi_k(\omega) = \phi(\omega) u' \tag{3.16}$$

où $\phi(\omega)$ dénote le vecteur des images de ω par les fonctions de base :

$$\phi(\omega) = (\phi_1(\omega), \ldots, \phi_m(\omega)) \tag{3.17}$$

La contrainte de normalisation sur les FP impose :

$$\int_\Omega |f_j(\omega)|^2 dP(\omega) = \int_\Omega {}^t(\phi(\omega)u')(\phi(\omega)u') dP(\omega) = \int_\Omega {}^t u' {}^t\phi(\omega)\phi(\omega)u' dP(\omega) = {}^t u' \mathcal{W}_\phi u' = 1 \tag{3.18}$$

On considère la fonction définie sur $\Omega \times \Omega$ qui à (ω, ω') associe le produit scalaire usuel entre les individus associés :

$$w(\omega, \omega') = \mathbf{X}(\omega) {}^t\mathbf{X}(\omega') = \sum_{i=1}^{p} X_i(\omega) X_i(\omega') \tag{3.19}$$

Soit l'opérateur de $\mathcal{L}^2(\Omega)$ dans $\mathcal{L}^2(\Omega)$ associé :

$$W(f) = \int_\Omega w(.,\omega')f(\omega')dP(\omega') \tag{3.20}$$

Par définition des f_j, on peut écrire :

$$W(f_j)(\omega) = \int_\Omega w(\omega,\omega')f_j(\omega')dP(\omega') = \lambda f_j(\omega) \tag{3.21}$$

où λ est la valeur propre associée à f_j. Supposons maintenant que chaque X_i s'exprime comme une somme de m fonctions de base ϕ_k :

$$\mathbf{X} = \phi\mathbf{C} \tag{3.22}$$

On peut donc écrire $W(f_j)$ de la façon suivante :

$$\begin{aligned} W(f_j)(\omega) &= \int_\Omega w(\omega,\omega')f_j(\omega')dP(\omega') \\ &= \int_\Omega \mathbf{X}(\omega)^t\mathbf{X}(\omega')f_j(\omega')dP(\omega') \\ &= \int_\Omega (\phi(\omega)\mathbf{C})^t(\phi(\omega')\mathbf{C})(\phi(\omega')u')dP(\omega') \\ &= \phi(\omega)\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi u' = \lambda f_j(\omega) = \lambda\phi(\omega)u' \end{aligned} \tag{3.23}$$

Ce résultat est vrai pour tout ω et la formule 3.21 est donc équivalente à :

$$\phi\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi u' = \lambda\phi u' \tag{3.24}$$

que l'on simplifie car ϕ est de rang maximal :

$$\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi u' = \lambda u' \tag{3.25}$$

Le problème de l'ACP fonctionnelle se résume donc à trouver les coefficients de la décomposition des FP dans ϕ, c'est à dire à diagonaliser la matrice $\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi$. D'après (3.18), en posant $u = \mathcal{W}_\phi^{1/2}u'$, alors ${}^tuu = 1$ et (3.25) devient :

$$\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi u' = \mathbf{C}^t\mathbf{C}\mathcal{W}_\phi^{1/2}u = \lambda u' = \lambda\mathcal{W}_\phi^{-1/2}u \tag{3.26}$$

Ce qui est équivalent à :

$$\mathcal{W}^{1/2}\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi^{1/2}u = \lambda u \tag{3.27}$$

On calcule la matrice de projection $\mathbf{F}$ de taille $p \times p$ des X_i sur les FP. Si $\mathbf{B} = \mathcal{W}_\phi^{-1/2}\mathbf{U}$ est la matrice $m \times m$ des projections des FP dans la base ϕ :

$$\begin{aligned} \mathbf{F}_{ij} &= \int_\Omega X_i(\omega)f_j(\omega)d\omega = [{}^t(\phi\mathbf{C})(\phi\mathbf{B})]_{ij} = (\mathbf{C}\mathcal{W}_\phi\mathbf{B})_{ij} \\ &= (\mathbf{C}\mathcal{W}_\phi^{1/2}\mathcal{W}_\phi^{1/2}\mathbf{B})_{ij} = (\mathbf{C}\mathcal{W}_\phi^{1/2}\mathbf{U})_{ij} \end{aligned} \tag{3.28}$$

ACP fonctionnelle

Entrées :
- $\mathbf{X} = (X_1, \ldots, X_p)$ les p variables aléatoires centrées.
- $\phi = (\phi_1, ..., \phi_m)$ le vecteur colonne des m fonctions de base.
- $\mathcal{W}_\phi$ la matrice carrée d'ordre m des produits scalaires des éléments de ϕ.
- $\mathbf{C}$ la matrice $p \times m$ des coefficients des décompositions des X_i dans ϕ.

Sortie :
- Matrice $\mathbf{F}$ des projections des X_i sur les fonctions propres.

Procédure :

1. Calcul des vecteurs propres normés $\mathbf{U}$ de la matrice $\mathcal{W}_\phi^{1/2}\mathbf{C}^t\mathbf{C}\mathcal{W}_\phi^{1/2}$.
2. Calcul des fonctions propres $f(\omega) = \phi(\omega)\mathbf{B}$ où $\mathbf{B} = \mathcal{W}_\phi^{-1/2}\mathbf{U}$.
3. Calcul des parts d'inertie des valeurs propres correspondantes.
4. Calcul des projections des X_i sur les fonctions propres $\mathbf{F} = \mathbf{C}\mathcal{W}_\phi^{1/2}\mathbf{U}$.

TAB. 3.1 – ACP fonctionnelle.

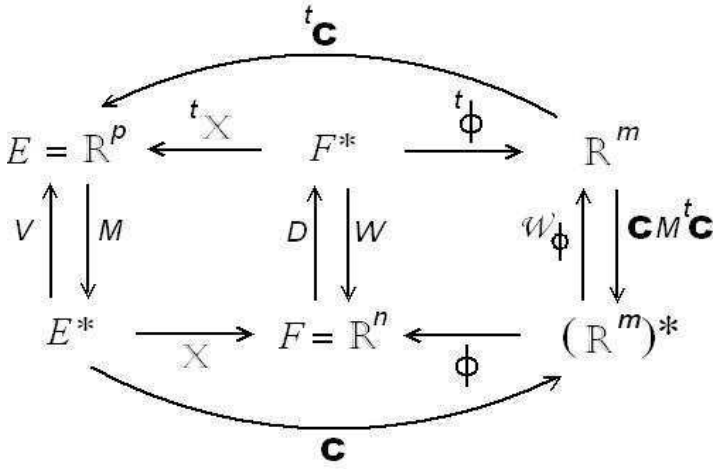

FIG. 3.3 – Schéma de dualité de l'ACP fonctionnelle.

$\mathbf{F}_{ij}$ est la valeur de la projection de X_i sur la $j$ème FP. La table 3.1 donne un récapitulatif de la méthode. La figure 3.3 présente le schéma de dualité de l'ACP fonctionnelle auquel est rajouté la correspondance avec le schéma de l'ACP discrète (partie droite sur la figure). Le schéma de dualité se situe dans un contexte expérimental, où les fonctions sont discrétisées. ϕ est alors une matrice de taille $n \times m$ et contient en colonne les n images de chacune des m fonctions de base. On exprime alors la matrice de covariance des fonctions de base de la même manière :

$$\mathcal{W}_\phi = {}^t\phi D \phi \tag{3.29}$$

La matrice de covariance V s'exprime à la fois dans sa version discrétisée :

$$V = {}^t\mathbb{X} D \mathbb{X} \tag{3.30}$$

et également à l'aide des coefficients de leur décomposition dans l'espace des fonctions de base :

$$V = {}^t\mathbb{X} D \mathbb{X} = {}^t(\phi\mathbf{C}) D (\phi\mathbf{C}) = {}^t\mathbf{C}({}^t\phi D \phi)\mathbf{C} = {}^t\mathbf{C}\mathcal{W}_\phi\mathbf{C} \tag{3.31}$$

3.4.2 ACP relationnelle et ACP fonctionnelle

Pour montrer que l'ACP relationnelle est une ACP fonctionnelle particulière, on se place dans l'espace vectoriel engendré par les variables aléatoires X_i, que l'on suppose centrées et linéairement indépendantes. Pour un graphe symétrique, le laplacien dans sa version discrétisée est une matrice symétrique semi-définie positive. D'après un théorème établi dans [Woo66], il existe une unique matrice racine carrée de Δ, notée $\Delta^{1/2}$:

$$\Delta = \left(\Delta^{1/2}\right)^2 \tag{3.32}$$

Cette matrice est également symétrique semi-définie positive. On décompose alors la matrice de covariance relationnelle de la manière suivante :

$$\begin{aligned} V_G &= \frac{1}{\kappa}{}^t\mathbb{X}\Delta\mathbb{X} = \frac{1}{\kappa}{}^t\mathbf{C}({}^t\phi\Delta\phi)\mathbf{C} \\ &= {}^t\mathbf{C}{}^t\left(\frac{1}{\sqrt{\kappa}}D^{-1/2}\Delta^{1/2}\mathbb{X}\right) D \left(\frac{1}{\sqrt{\kappa}}D^{-1/2}\Delta^{1/2}\mathbb{X}\right)\mathbf{C} \end{aligned} \tag{3.33}$$

Si l'on pose $\phi = \frac{1}{\sqrt{\kappa}}D^{-1/2}\Delta^{1/2}\mathbb{X}$, ces $m = p$ fonctions forment une base de l'espace $< X_1, \ldots, X_p >$. De cette façon, on se ramène à l'ACP classique sur les fonctions de base. Ainsi, la covariance entre les fonctions propres coïncide avec la covariance relationnelle entre les variables aléatoires associées :

$$\begin{aligned} Cov(\phi_i, \phi_j) &= < \phi_i, \phi_j > = < \frac{1}{\sqrt{\kappa}}D^{-1/2}\Delta^{1/2}X_i, \frac{1}{\sqrt{\kappa}}D^{-1/2}\Delta^{1/2}X_j > \\ &= \frac{1}{\kappa}{}^tX_i\Delta X_j = Cov_G(X_i, X_j) \end{aligned} \tag{3.34}$$

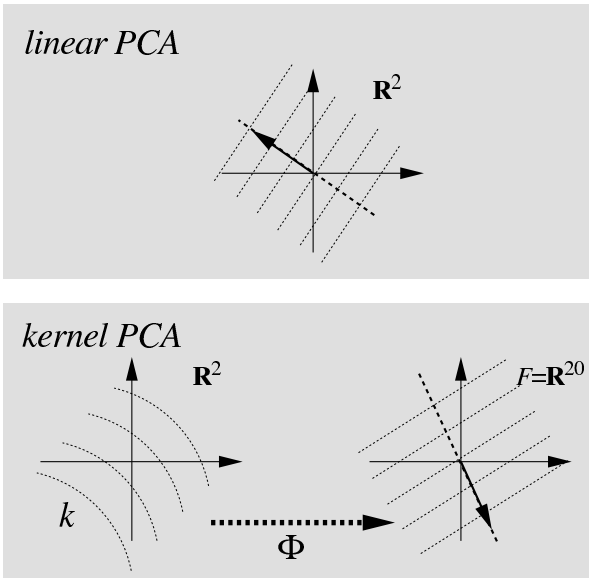

FIG. 3.4 – ACP à noyaux (source [SSM96]) : projeter les individus dans un espace de dimension supérieure et se ramener à une ACP standard.

3.5 ACP à noyaux

L'ACP à noyaux [SSM96] est appliquée dans le cas où les liaisons entre les variables aléatoires sont non linéaires. Comme illustré à la figure 3.4, le principe est de plonger les individus dans un espace de dimension supérieure de manière à linéariser les liaisons entre les variables aléatoires. On se ramène ensuite à une ACP standard sur cet espace. Nous commençons par la description de la méthode puis nous mettons en évidence comment introduire dans le modèle une relation binaire sur les individus à l'aide du schéma de dualité.

3.5.1 Présentation de la méthode

On suppose l'existence d'une application Φ qui plonge les individus de $\mathbb{R}^p$ vers un espace $\mathcal{E}$, appelé *espace caractéristique* :

$$\begin{array}{rccc} \Phi & : & \mathbb{R}^p & \rightarrow & \mathcal{E} \simeq \mathbb{R}^s \\ & & x & \mapsto & \Phi(x) \end{array} \tag{3.35}$$

La dimension s de l'espace caractéristique $\mathcal{E}$ peut être très grande, voire infinie. En notant par $\mathbb{Y}$ le tableau de taille $n \times s$ contenant l'ensemble des images par Φ des

individus, l'ACP à noyaux consiste à faire une ACP du triplet statistique $(\mathbb{Y}, I_s, \frac{1}{n}I_n)$. La matrice de covariance est donc donnée par :

$$V_{\mathcal{E}} = \frac{1}{n}{}^t\mathbb{Y}\mathbb{Y} = \frac{1}{n}\sum_{i=1}^{n} {}^t\Phi(x_i)\Phi(x_i) \in \mathbb{R}^s \times \mathbb{R}^s \tag{3.36}$$

On cherche à déterminer les éléments propres de $V_{\mathcal{E}}$. Un vecteur propre v constituerait un élément d'une base du sous-espace vectoriel de $\mathcal{E}$ engendré par les $\Phi(x_i)$. C'est donc une combinaison linéaire pondérée par les éléments du vecteur unique $\alpha = {}^t(\alpha_1, \ldots, \alpha_n)$:

$$v = \sum_{i=1}^{n} \alpha_i \Phi(x_i) \tag{3.37}$$

La connaissance a priori de l'application Φ n'est pas indispensable puisque l'on a besoin uniquement de la valeur du produit scalaire dans $\mathcal{E}$ entre les $\Phi(x_i)$. Ce produit scalaire, ou *noyau*, est défini à partir des individus de $\mathbb{R}^p$ et agit comme une mesure de similarité non linéaire sur $\mathbb{R}^p$:

$$\begin{array}{rcccl} k & : & \mathbb{R}^p \times \mathbb{R}^p & \rightarrow & \mathbb{R} \\ & & (x, x') & \mapsto & k(x, x') =< \Phi(x), \Phi(x') >_{\mathcal{E}} \end{array} \tag{3.38}$$

où $< ., . >_{\mathcal{E}}$ est le produit scalaire usuel dans $\mathcal{E}$. Les valeurs du noyau sont stockées dans la matrice de Gram $W_{\mathcal{E}}$ des $\Phi(x_i)$:

$$(W_{\mathcal{E}})_{i,j} = (\mathbb{Y}{}^t\mathbb{Y})_{ij} =< \Phi(x_i), \Phi(x_j) >_{\mathcal{E}}, \quad 1 \leq i, j \leq n \tag{3.39}$$

L'ACP dans $\mathcal{E}$ impose le centrage des variables aléatoires pour la métrique $\frac{1}{n}I_n$, soit $\sum_i \Phi(x_i) = 0 \in \mathbb{R}^s$. Ne connaissant pas l'application Φ, cette contrainte est prise en compte en double-centrant la matrice $W_{\mathcal{E}}$:

$$\overline{W}_{\mathcal{E}} = W_{\mathcal{E}} - \mathbf{1}_n W_{\mathcal{E}} - W_{\mathcal{E}}\mathbf{1}_n + \mathbf{1}_n W_{\mathcal{E}} \mathbf{1}_n \tag{3.40}$$

où $\mathbf{1}_n$ est la matrice carrée d'ordre n de termes constants égaux à $1/n$:

$$\mathbf{1}_n = \frac{1}{n}\begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix} \in \mathbb{R}^n \times \mathbb{R}^n \tag{3.41}$$

La matrice $\overline{W}_{\mathcal{E}}$ est donc la matrice de Gram telle que $\sum_i \Phi(x_i) = 0$. Pour tout k fixé entre 1 et n, on a :

$$< \Phi(x_k), V_{\mathcal{E}} v >_{\mathcal{E}} = \lambda < \Phi(x_k), v >_{\mathcal{E}} \tag{3.42}$$

où λ est la valeur propre associée à v. En remplaçant les termes à l'aide des formules 3.36 et 3.37, on obtient :

$$\frac{1}{n}\overline{W}_{\mathcal{E}}^2 \alpha = \lambda \overline{W}_{\mathcal{E}} \alpha \tag{3.43}$$

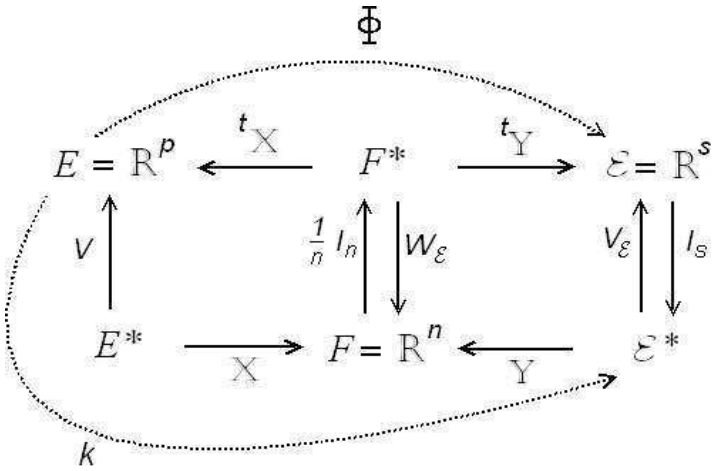

FIG. 3.5 – Schéma de dualité de l'ACP à noyaux.

que l'on simplifie, puisque $\overline{W}_{\mathcal{E}}$ est une matrice de Gram inversible :

$$\frac{1}{n}\overline{W}_{\mathcal{E}}\alpha = \lambda\alpha \tag{3.44}$$

Par conséquent, α est un vecteur propre de $\overline{W}_{\mathcal{E}}/n$ et le raisonnement est le même pour les autres vecteurs propres de $V_{\mathcal{E}}$. La projection de $\Phi(x)$ sur l'axe v est donnée par la formule :

$$< \Phi(x), v >_{\mathcal{E}} = \sum_{i=1}^{n} \alpha_i < \Phi(x), \Phi(x_i) >_{\mathcal{E}} = \sum_{i=1}^{n} \alpha_i k(x, x_i) \tag{3.45}$$

Par ailleurs, on impose que v soit un vecteur normé, ce qui implique la condition suivante pour le vecteur α :

$$\begin{aligned} < v, v >_{\mathcal{E}} &= < \sum_{i=1}^{n} \alpha_i \Phi(x_i), \sum_{j=1}^{n} \alpha_j \Phi(x_j) >_{\mathcal{E}} = \sum_{i,j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) \\ &= {}^t\alpha \overline{W}_{\mathcal{E}} \alpha = \lambda n {}^t\alpha\alpha = 1 \end{aligned} \tag{3.46}$$

Par conséquent, α doit être normé en le divisant par la racine carrée du produit λn. L'encadré 3.2 donne les différentes étapes de l'ACP à noyaux et la figure 3.5 présente son schéma de dualité. La partie gauche du diagramme est le schéma de dualité classique de l'ACP dans l'espace E et la partie droite correspond au schéma de l'ACP dans $\mathcal{E}$. La matrice $W_{\mathcal{E}}$ de la matrice de Gram entre les $\Phi(x_i)$ permet d'établir la correspondance entre les deux schémas. Les liaisons linéaires entre les variables aléatoires étant présentes dans $\mathcal{E}$ mais pas dans E, il n'est pas possible d'exprimer $W_{\mathcal{E}}$ par un produit de matrices du schéma de l'ACP dans l'espace initial E. En revanche, cette matrice s'exprime matriciellement dans $\mathcal{E}$ par la composition $W_{\mathcal{E}} = \mathbb{Y}\left(\frac{1}{s}I_s\right){}^t\mathbb{Y}$. La métrique

ACP à noyaux

Entrées :
- Tableau $\mathbb{X}$.
- Noyau $k(x, y) =< \Phi(x), \Phi(y) >_{\mathcal{E}}$.

Sortie :
- Représentation des $\Phi(x_i)$ dans les axes principaux.

Procédure :

1. Calcul de la matrice de Gram $(W_{\mathcal{E}})_{ij} = k(x_i, x_j)$.
2. Double-centrer la matrice de Gram $\overline{W}_{\mathcal{E}} = W_{\mathcal{E}} - \mathbf{1}_n W_{\mathcal{E}} - W_{\mathcal{E}}\mathbf{1}_n + \mathbf{1}_n W_{\mathcal{E}}\mathbf{1}_n$.
3. Calcul des éléments propres (α, λ) de $\frac{1}{n}\overline{W}_{\mathcal{E}}$.
4. Normer les vecteurs propres par la racine carrée du produit de la valeur propre avec n.
5. Calcul des projections des $\Phi(x_i)$ sur les axes par le produit scalaire entre le vecteur propre correspondant et la $i$ème ligne de $\overline{W}_{\mathcal{E}}$.

TAB. 3.2 – ACP à noyaux.

identité I_s induit naturellement un isomorphisme entre $\mathcal{E}$ et son dual par la relation :

$$\begin{array}{rccc} I_s & : & \mathcal{E} & \rightarrow & \mathcal{E}^* \\ & & u & \mapsto & < u, . >_{\mathcal{E}} \end{array} \tag{3.47}$$

Les liaisons induites par l'application Φ et k ont été tracées en pointillés car ces transformations ne sont pas linéaires. Néanmoins, si l'on définit le noyau comme une application de $\mathbb{R}^p$ vers $\mathcal{E}^*$ de la manière suivante :

$$\begin{array}{rccc} k & : & \mathbb{R}^p & \rightarrow & \mathcal{E}^* \\ & & x & \mapsto & < \Phi(x), . >_{\mathcal{E}} \end{array} \tag{3.48}$$

alors on établit d'après le schéma de dualité la composition $k = I_s \circ \Phi$:

$$\begin{array}{rccccc} k & : & \mathbb{R}^p & \rightarrow & \mathcal{E} & \rightarrow & \mathcal{E}^* \\ & & x & \mapsto & \Phi(x) & \mapsto & < \Phi(x), . >_{\mathcal{E}} \end{array} \tag{3.49}$$

Il existe différents type de noyau, mais très peu sont conformes au théorème de Mercer qui assure la représentation d'un noyau comme un produit scalaire. Les noyaux conformes les plus utilisés sont les suivants :

Le noyau polynomial :

$$k(x, y) =< x, y >^d , \quad d \geq 1 \tag{3.50}$$

Le noyau sigmoïde :

$$k(x, y) = \tanh(\beta_0 < x, y > +\beta_1) \tag{3.51}$$

Le noyau à base radiale :

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\gamma^2}\right) \tag{3.52}$$

Les constantes β_0, β_1 et γ sont choisies a priori par l'utilisateur. Les noyaux polynômiaux et à base radiales sont toujours conformes au théorème de Mercer. Le noyau sigmoïde quant à lui, vérifie le théorème uniquement pour certaines valeurs de β_0 et β_1 [Hay99]. Pour un noyau polynomial avec $d = 1$, on retrouve le produit scalaire usuel entre les individus et donc l'ACP standard.

Notons que, contrairement à l'ACP classique, l'ACP à noyaux ne permet pas une reconstruction exacte des données dans l'espace initial. Ceci est dû à la non connaissance de l'application Φ, qui n'est en outre pas nécessairement bijective. Cependant, il est toujours possible de trouver un vecteur $\widetilde{v}$ de $\mathbb{R}^p$ dont l'image par Φ approxime un vecteur propre v de $V_{\mathcal{E}}$ [SSM96] :

$$v \approx \Phi(\widetilde{v}) \tag{3.53}$$

Il suffit pour cela d'écrire la distance dans $\mathcal{E}$ en fonction du noyau :

$$\begin{aligned} d(\Phi(x),\Phi(x'))^2_{\mathcal{E}} &= \|\Phi(x)-\Phi(x')\|^2_{\mathcal{E}} =< \Phi(x)-\Phi(x'),\Phi(x)-\Phi(x') >_{\mathcal{E}} \\ &= < \Phi(x),\Phi(x) >_{\mathcal{E}} -2 < \Phi(x),\Phi(x') >_{\mathcal{E}} + < \Phi(x'),\Phi(x') >_{\mathcal{E}} \\ &= k(x,x) - 2k(x,x') + k(x',x') \end{aligned} \tag{3.54}$$

Le vecteur $\widetilde{v}$ est obtenu en minimisant la distance $d(\Phi(\widetilde{v}),v)_{\mathcal{E}}$ par rapport à $\widetilde{v}$:

$$\widetilde{v} = \underset{u}{\operatorname{argmin}} \|v-\Phi(u)\|^2_{\mathcal{E}} \tag{3.55}$$

En exprimant la distance à l'aide du noyau et d'après (3.37), on en déduit :

$$\|v-\Phi(u)\|^2_{\mathcal{E}} = \sum_{i=1}^{n} \alpha_i^2 k(x_i,x_i) - 2\sum_{i=1}^{n} \alpha_i k(x_i,u) + k(u,u) \tag{3.56}$$

Après suppression des termes indépendants de u, on aboutit à la formule suivante :

$$\widetilde{v} = \underset{u}{\operatorname{argmin}} \|v-\Phi(u)\|^2_{\mathcal{E}} = \underset{u}{\operatorname{argmin}} \left[k(u,u) - 2\sum_{i=1}^{n} \alpha_i k(x_i,u) \right] \tag{3.57}$$

La reconstruction des données est donc toujours possible mais de qualité variable.

En résumé, l'ACP à noyaux est une méthode efficace pour réduire la dimension de données présentant une structure non linéaire. Son implémentation est aussi simple que l'ACP standard et n'impose pas d'optimisation de fonction. Elle présente néanmoins l'inconvénient lié aux courbes de projections constantes "incurvées" dans l'espace de départ, qui rendent les interprétations difficiles [SSM96]. D'autre part, la résolution spectrale de l'ACP à noyaux impose de travailler avec un nombre raisonnable d'individus.

3.5.2 ACP relationnelle et ACP à noyaux

Comme le noyau définit un produit scalaire dans l'espace caractéristique, cette application est donc symétrique définie positive. Par conséquent, on ne peut établir l'expression générale du noyau muni d'une relation binaire entre les individus puisque le laplacien est semi-défini positif. Toutefois, la relation peut être incluse de façon matricielle à l'aide du schéma de dualité. Dans le diagramme, si l'on remplace la métrique $\frac{1}{n}I_n$ par celle du laplacien, le modèle tient compte de la relation sans aucun biais méthodologique dans la mesure où l'on considère les images $\Phi(x_i)$ comme des individus dans l'espace caractéristique. En effet, nous supposons que la relation binaire induite par les

individus est directement "retransmise" dans $\mathcal{E}$. Si $k(.,.)$ est un noyau quelconque, nous proposons d'introduire le *noyau relationnel* de la manière suivante :

$$\mathcal{K} = \frac{1}{n}\overline{W}_{\mathcal{E}}.\Delta \tag{3.58}$$

Dans ce cas, l'ACP à noyaux est un cas général de l'ACP relationnelle et la symétrie du graphe n'est pas imposée. La matrice $\mathcal{K}$ place naturellement dans le modèle la relation binaire entre les individus projetés. Les projections sur les axes principaux sont ensuite calculées de manière classique avec la formule 3.45.

3.6 Analyse factorielle d'opérateurs

L'analyse des tableaux à trois entrées est un domaine assez ancien qui a débuté au milieu des années 60 avec les travaux de Tucker [Tuc64]. On peut distinguer un ensemble de variantes des modèles d'analyse de tableaux multiples et l'on trouvera une analyse détaillée et comparative des principaux modèles dans la thèse de Bro [Bro98]. Des applications récentes en ont montré l'efficacité dans des thèmes de recherche utilisant des données à structure complexe : analyse environmentale [SWM+04], système de communication [RV00], traitement d'image [AOVE05, YZFY04, Vis05]. Globalement, les méthodes d'analyse de tableaux multiples se définissent comme une extension des méthodes factorielles classiques à des données tridimensionnelles, dont la représentation naturelle est une matrice 3D contenant la superposition de N sous-tableaux :

$$\mathfrak{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_N) \tag{3.59}$$

La première approche de l'analyse de tableaux multiples a pour but de déterminer une décomposition synthétique du tableau $\mathfrak{X}$. Parmi les méthodes classiques, l'ACP bidimensionnelle [YZFY04] est une généralisation de l'ACP standard pour un tableau à trois entrées. Plutôt que d'appliquer directement l'ACP sur le tableau $\mathfrak{X}$ bidimensionnel (figure 3.6(c)), l'ACP bidimensionnelle construit une matrice de projection commune à tous les sous-tableaux. Cela permet d'éviter un pré-traitement ainsi que des contraintes numériques liées à la grande dimension des données. Introduite dans le cadre de la reconnaissance de visages, l'ACP bidimensionnelle implique nécessairement que les sous-tableaux aient la même dimension. On peut montrer que les colonnes de la matrice de projection sont les vecteurs propres associés aux premières valeurs propres de la matrice de covariance généralisée :

$$V_{ACP2D} = \frac{1}{N}\sum_{k=1}^{N} {}^t(\mathbb{X}_k - \overline{\mathbb{X}})(\mathbb{X}_k - \overline{\mathbb{X}}) \in \mathbb{R}^p \times \mathbb{R}^p \tag{3.60}$$

où $\overline{\mathbb{X}}$ est le sous-tableau "moyen" :

$$\overline{\mathbb{X}} = \frac{1}{N} \sum_{k=1}^{N} \mathbb{X}_k \tag{3.61}$$

Les modèles PARAFAC (PARAllel FACtor analysis) [Har70] sont des méthodes de décomposition matricielle du tableau $\mathfrak{X}$. Récemment, une application a été proposée à la séparation de sources de données tridimensionnelles en combinant les modèles PARAFAC avec l'analyse en composantes indépendantes [Cas06]. Les modèles PARAFAC proposent la reconstruction dans le cas de données tridimensionnelles et en supposant que les sous-tableaux sont de même dimension. On cherche à déterminer trois matrices P^1, P^2 et P^3 vérifiant la formule de reconstruction suivante :

$$\widehat{\mathfrak{X}}_{ijk} = \sum_{l=1}^{L} P^1_{il} P^2_{jl} P^3_{kl} \tag{3.62}$$

Plus L est grand, et plus la reconstruction est proche du tableau $\mathfrak{X}$. Les matrices P^1, P^2 et P^3 sont respectivement de taille $n \times L$, $p \times L$ et $N \times L$. PARAFAC généralise au cas 3D la décomposition en valeurs singulières d'un sous-tableau :

$$\widehat{\mathbb{X}}_{ij} = \sum_{l=1}^{L} P^1_{il} P^2_{jl} \tag{3.63}$$

Les solutions du modèle PARAFAC sont uniques et déterminées en minimisant l'erreur quadratique :

$$E_{parafac} = \|\mathfrak{X}_{ijk} - \sum_{l=1}^{L} P^1_{il} P^2_{jl} P^3_{kl}\|^2 \tag{3.64}$$

Dorénavant, nous considérons sans perte de généralité que le nombre d'individus n_k varie suivant l'indice du sous-tableau, bien que par dualité les méthodes décrites soient adaptables dans le cas où il s'agit du nombre p de variables, et non n_k, qui soit non constant. Le tableau multiple $\mathfrak{X}$ contient l'empilement de N sous-tableaux $\mathbb{X}_k$ de taille respective $p \times n_k$, où p est le nombre de variables, commun à tous les sous-tableaux, et n_k la taille de l'échantillon correspondant à $\mathbb{X}_k$. Ainsi, $\mathfrak{X}_{ijk}$ est l'élément ijk de $\mathfrak{X}$, *i.e.* la $i$ème réalisation de la $j$ème variable du sous-tableau $\mathbb{X}_k$. La figure 3.6 donne une illustration graphique du tableau $\mathfrak{X}$: en (a) est représenté le tableau entier $\mathfrak{X}$, en (b) un exemple de deux sous-tableaux avec k et k' fixés, et en (c) le tableau $\mathfrak{X}$ re-dimensionné en tableau bidimensionnel.

Les modèles PARAFAC2 [Har72] généralisent quant à eux PARAFAC en adaptant le modèle à des sous-tableaux de taille différente. Pour cela, on procède à une étape préalable en identifiant chacun des sous-tableaux par la signature ${}^t\mathbb{X}_k\mathbb{X}_k$ de taille commune $p \times p$. On se ramène ensuite à appliquer la méthode standard PARAFAC à ces

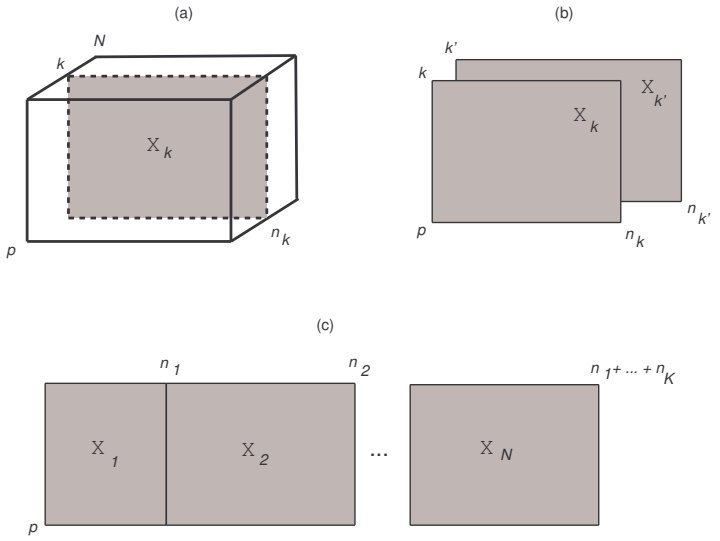

FIG. 3.6 – (a) Tableau $\mathfrak{X}$ à trois entrées ; (b) Deux sous-tableaux $\mathbb{X}_k$ et $\mathbb{X}_{k'}$ de taille respective n_k et $n_{k'}$; (c) Tableau $\mathfrak{X}$ reconstitué en tableau bidimensionnel.

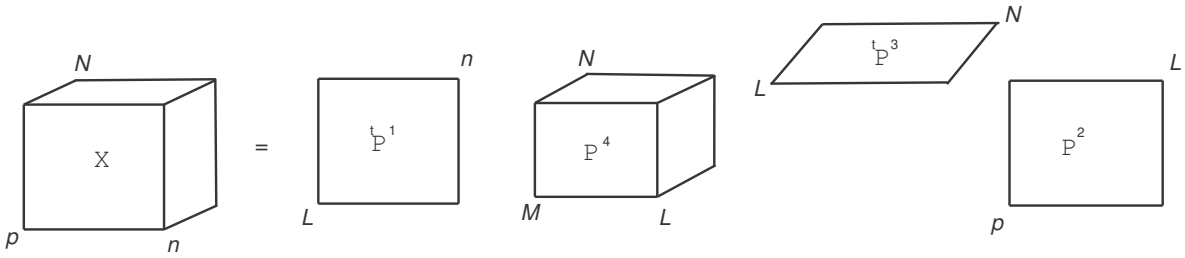

FIG. 3.7 – Modèle Tucker3.

signatures. On peut montrer que les solutions de PARAFAC2 sont uniques sous certaines contraintes liées au rang et à la non-négativité des matrices [HL96, BKdL96]. On citera enfin le modèle Tucker3 [KdL80], connu aussi sous le nom d'*ACP à trois modes*. La figure 3.7 donne une illustration du modèle. Il s'agit d'un modèle de régression dans lequel on estime également les valeurs d'une matrice tridimensionnelle, notée P^4, non nécessairement cubique et vérifiant la formule suivante :

$$\widehat{\mathfrak{x}}_{ijk} = \sum_{l=1}^{L} \sum_{m=1}^{M} \sum_{n=1}^{N} P^1_{il} P^2_{jm} P^3_{kn} P^4_{lmn} \tag{3.65}$$

Il est clair que PARAFAC est un cas particulier du modèle Tucker3 où P^4 est la matrice cubique identité d'ordre L. On montre que les solutions du modèle Tucker3 sont uniques à une rotation près [Bro98].

La seconde approche ne s'intéresse pas à une décomposition synthétique mais plutôt à l'extraction de signatures, et considère chacun des sous-tableaux comme un point dans un même espace. Les signatures attribuées sont appelées *opérateur* et sont déterminées à partir de critères de type distance ou produit scalaire. Ainsi, ces méthodes permettent d'étudier les ressemblances et les dissemblances mutuelles entre les sous-tableaux puisque les signatures sont plongées dans un même espace. Les plus classiques sont l'analyse factorielle multiple [EP98], l'analyse canonique généralisée [Hor61], STATIS (Structuration des Tableaux À Trois Indices de la Statistique) [Esc80, Esc85] et l'analyse procrustéenne [Tuc58, Gow75]. L'analyse factorielle multiple consiste à appliquer une ACP standard sur $\mathfrak{X}$ en adoptant la disposition bidimensionnelle de $\mathfrak{X}$. Afin de ne pas privilégier les variables présentant une forte corrélation mutuelle, chaque sous-tableau est préalablement normé par la plus grande valeur propre d'une ACP spécifique effectuée sur les variables de ce sous-tableau. L'analyse canonique généralisée résume l'information de chaque sous-tableau à l'aide d'une seule variable, puis en déduit une variable artificielle la plus liée en moyenne. Cette approche admet comme cas particuliers toutes les méthodes classiques d'analyse factorielle d'un seul sous-tableau : analyse canonique, analyse discriminante et ACP. La méthode STATIS identifie chaque sous-tableau par sa matrice de Gram entre individus, et en recherche une configuration qui représente la part d'information "moyenne" apportée par les sous-tableaux. L'analyse procrustéenne, proposée à l'origine par Tucker [Tuc58] puis généralisée par Gower [Gow75] au cas $N > 2$, cherche à déterminer le sous-tableau constituant l'information "moyenne" entre les $\mathbb{X}_k$, à l'aide d'ajustements géométriques de chacun de ces sous-tableaux (homothétie et rotation). L'analyse procrustéenne est plus complexe dans sa mise en œuvre que STATIS puisqu'il s'agit d'une méthode d'optimisation par une heuristique où l'on recherche les solutions du problème de minimisation :

$$E_{procrus} = \sum_{k=1}^{N} trace({}^t(O_1 - \rho_k \mathbb{X}_k P_k)(O_1 - \rho_k \mathbb{X}_k P_k)) \tag{3.66}$$

où ρ_k et P_k sont respectivement le coefficient de normalisation et la matrice de rotation de $\mathbb{X}_k$. Si les n_k sont non constants, une astuce consiste à les compléter par des 0 de manière à obtenir une taille commune à tous les sous-tableaux.

3.6.1 Analyse d'opérateurs de covariance relationnelle

Dans ce travail, nous nous intéressons à l'analyse par extraction de signature en considérant le modèle suivant. Une fois définie l'ACP relationnelle, nous proposons l'extension naturelle à l'analyse d'opérateurs de covariance relationnelle (AO) pour l'étude statistique de tableaux à trois entrées [HLSJ06b]. La méthode proposée est une variante dérivée de l'analyse factorielle d'opérateurs [Fou84]. Le but est d'étudier les

liens entre les sous-tableaux en adoptant le même principe que l'ACP : rechercher les sous-espaces *principaux* préservant au mieux la distance entre les opérateurs. Ces sous-espaces sont engendrés par une base orthonormée issue d'une combinaison linéaire des opérateurs de référence. Précisément, on identifiera un sous-tableau par son opérateur de covariance. Cette identification est réalisée sans perte de généralité car dans le cas où il s'agit du nombre de variables et non celui des individus qui soit non constant, alors la signature naturelle de chacun des sous-tableaux est la matrice de Gram des produits scalaires entre les individus. Les interprétations sont alors différentes dans le sens où l'on analyse l'interstructure des opérateurs à l'aide des liaisons entre les individus et non plus les variables des sous-tableaux. L'algorithme se déroule en trois étapes. La première étape consiste à calculer à partir de $\mathfrak{X}$ les opérateurs à analyser. L'AO cherche à construire une base orthonormale $(O_1, \ldots, O_N)$ de l'espace des opérateurs $< \Gamma_1, \ldots, \Gamma_N >$. Ayant choisi un graphe G_k d'une relation binaire et une métrique M_k sur les observations du sous-tableau $\mathbb{X}_k$, l'opérateur Γ_k se définit de la manière suivante :

$$\Gamma_k = \frac{1}{\kappa_k} {}^t\mathbb{X}_k \Delta_k \mathbb{X}_k M_k \ , \quad 1 \leq k \leq N \tag{3.67}$$

où κ_k et Δ_k désignent respectivement la connectivité et le laplacien du graphe G_k. Les opérateurs Γ_k sont les opérateurs de covariance relationnelle. Dans ce sens, l'AO peut être considérée comme une extension de l'ACP relationnelle dans le cadre de l'analyse de tableaux multiples. L'opérateur Γ_k est autoadjoint pour le produit scalaire $< ., . >_{M_k}$:

$$\forall x, y \in \mathbb{R}^p \ , \quad < x, \Gamma_k.y >_{M_k} = < \Gamma_k.x, y >_{M_k} \tag{3.68}$$

La seconde étape de la méthode définit les relations entre les opérateurs comme le font l'analyse factorielle multiple [EP98] ou l'analyse multiple sous contrainte de voisinages [CSC93]. Pour cela, on utilise le produit scalaire de Hilbert-Schmidt (HS) :

$$< \Gamma_k, \Gamma_{k'} >_{HS} = trace(\Gamma_k^* \Gamma_{k'}) = trace(\Gamma_k \Gamma_{k'}) \tag{3.69}$$

Le produit scalaire HS est équivalent au produit scalaire usuel sur $\mathbb{R}^{p^2}$ des matrices mises sous forme vectorielle. Lorsque ce produit scalaire est normalisé, cela coïncide avec le coefficient RV de Robert et Escoufier [RE76] utilisé dans la méthode STATIS :

$$< \Gamma_k, \Gamma_{k'} >_{RV} = \frac{< \Gamma_k, \Gamma_{k'} >_{HS}}{\sqrt{< \Gamma_k, \Gamma_k >_{HS} \times < \Gamma_{k'}, \Gamma_{k'} >_{HS}}} \tag{3.70}$$

En supposant qu'il n'y ait aucune relation de proportionnalité entre les Γ_k, la matrice des produits scalaires $\mathcal{H} = (< \Gamma_k, \Gamma_{k'} >_{HS})_{1 \leq k, k' \leq N}$ est symétrique définie positive. On détermine ensuite les éléments propres $(\mathcal{U}, \mathcal{L})$ de la matrice $\mathcal{H}$:

$$\mathcal{H}\mathcal{U} = \mathcal{U}\mathcal{L} \tag{3.71}$$

La troisième étape consiste à déterminer les opérateurs principaux[2] en utilisant la relation :

$$\widetilde{O} = \Gamma\mathcal{U} \tag{3.72}$$

Les vecteurs propres de $\mathcal{H}$ forment une base orthonormée de $\mathbb{R}^N$ et on vérifie que $\widetilde{O}$ est une base orthogonale de l'espace des opérateurs :

$$\begin{aligned} < \widetilde{O}_k, \widetilde{O}_{k'} >_{HS} &= < \sum_{i=1}^{N} \Gamma_i \mathcal{U}_{ik}, \sum_{j=1}^{N} \Gamma_j \mathcal{U}_{jk'} >_{HS} \\ &= \sum_{i,j=1}^{N} \mathcal{U}_{ik} < \Gamma_i, \Gamma_j >_{HS} \mathcal{U}_{jk'} = {}^t\mathcal{U}_k \mathcal{H} \mathcal{U}_{k'} \\ &= \mathcal{L}_{k'} {}^t\mathcal{U}_k \mathcal{U}_{k'} = \mathcal{L}_{k'} \delta_{kk'} \end{aligned} \tag{3.73}$$

Pour obtenir une base orthonormale de $< \Gamma_1, \ldots, \Gamma_N >$, on normalise les éléments de $\widetilde{O}$:

$$O = (O_1, \ldots, O_N) = \widetilde{O}\mathcal{L}^{-1/2} \tag{3.74}$$

Le compromis O_1 caractérise l'information moyenne apportée par les opérateurs. Cet élément, appelé *compromis*, exprime une représentation synthétique de l'ensemble des sous-tableaux dont STATIS mesure l'intrastructure des observations en effectuant une ACP sur cet opérateur. En revanche, O_1 n'apporte aucune information sur les relations entre les opérateurs et nous proposons pour cela de nous intéresser aux autres éléments de la base O. La projection de l'opérateur Γ_k sur l'opérateur principal $O_{k'}$ normée est obtenue en multipliant le $k$ème élément du vecteur $\mathcal{U}_{k'}$ par la racine carrée de la valeur propre associée $\mathcal{L}_{k'}$:

$$\begin{aligned} < \Gamma_k, O_{k'} >_{HS} &= < \Gamma_k, \frac{\widetilde{O}_{k'}}{\sqrt{\mathcal{L}_{k'}}} >_{HS} = < \Gamma_k, \frac{\Gamma\mathcal{U}_{k'}}{\sqrt{\mathcal{L}_{k'}}} >_{HS} \\ &= \frac{1}{\sqrt{\mathcal{L}_{k'}}} < \Gamma_k, \sum_{i=1}^{N} \mathcal{U}_{ik'} \Gamma_i >_{HS} = \frac{1}{\sqrt{\mathcal{L}_{k'}}} \sum_{i=1}^{N} \mathcal{H}(k,i) \mathcal{U}_{ik'} \\ &= \frac{1}{\sqrt{\mathcal{L}_{k'}}} (\mathcal{L}_{k'}\mathcal{U})_{k,k'} = \sqrt{\mathcal{L}_{k'}} \mathcal{U}_{k,k'} \end{aligned} \tag{3.75}$$

Grâce à cette méthode, nous caractérisons donc la totalité de l'information et la connaissance exogène est prise en compte par l'introduction du laplacien à la formule 3.67 dans chacun des sous-tableaux. STATIS est une AO particulière sur les matrices de Gram, où l'on impose tous les graphes G_k complets. L'encadré suivant donne une synthèse de la méthode.

[2]La démonstration de l'obtention des opérateurs principaux est rigoureusement la même que pour celle des CP (cf Annexe de ce chapitre).

Analyse d'opérateurs de covariance relationnelle

Entrées :
- Tableau $\mathfrak{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_N)$.
- Graphes de relation binaire G_k de chaque sous-tableau $\mathbb{X}_k$.
- Métriques M_k sur les individus de $\mathbb{X}_k$.

Sorties :
- Éléments propres $(\mathcal{U}, \mathcal{L})$ de $\mathcal{H}$.
- Opérateurs Γ.
- Opérateurs principaux O.

Procédure :

1. Calcul des opérateurs $\Gamma = (\Gamma_1, \ldots, \Gamma_N)$ avec $\Gamma_k = \frac{1}{\kappa_k} {}^t\mathbb{X}_k \Delta_k \mathbb{X}_k M_k$ et avec l'étape éventuelle de normalisation $\Gamma_k = \frac{\Gamma_k}{\|\Gamma_k\|_{HS}}$.
2. Calcul de la matrice $\mathcal{H} = (< \Gamma_k, \Gamma_{k'} >_{HS})_{1 \leq k,k' \leq N}$ des produits scalaires de Hilbert-Schmidt entre les opérateurs.
3. Calcul des éléments propres $(\mathcal{U}, \mathcal{L})$ de $\mathcal{H}$.
4. Calcul de la base orthonormée $O = \Gamma \mathcal{U} \mathcal{L}^{-1/2}$ des opérateurs principaux.
5. Calcul des projections $\mathcal{U}\mathcal{L}^{1/2}$ des Γ_k sur la base O.

TAB. 3.3 – Analyse d'opérateurs de covariance relationnelle.

3.6.2 Analyse discriminante d'opérateurs de covariance relationnelle

Maintenant que nous avons décrit le contexte de l'analyse d'opérateurs de covariance relationnelle, nous proposons dans cette section une extension à l'analyse factorielle discriminante [Fou84]. L'analyse discriminante d'opérateurs de covariance relationnelle (ADO) [HLSJ06c] consiste à appliquer de manière successive une analyse d'opérateurs, puis une analyse discriminante linéaire (ADL) sur les projections dans la base orthonormale obtenue. Il est naturellement possible de procéder directement à l'ADL sur les opérateurs mis sous forme vectorielle. Mais, dans un souci de réduire la complexité de calcul et de palier au mauvais conditionnement de la matrice à diagonaliser, nous choisissons de procéder à l'enchaînement ACP-ADL. Cette technique a déjà été préconisée dans le cadre des approches discriminantes appliquées à la reconnaissance de visages [BHK97, SW96].

Le principe général de l'ADO est illustrée à la figure 3.8. Nous considérons les premières (disons m) projections des opérateurs sur la base O. Ces projections sont contenues dans le tableau $\mathcal{Q}$ de taille $N \times m$, qui correspond en fait aux m premières colonnes de la matrice $\mathcal{U}\mathcal{L}^{1/2}$. La colonne i de $\mathcal{Q}$ est donc la projection de tous les Γ_k sur le $i$ème opérateur principal O_i. Par ailleurs, on se donne également un ensemble de K classes qui constitue une partition des Γ_k. On note par $\mathcal{G}$ la matrice $K \times m$ des centres de gravité des classes d'opérateurs, identifiés par leur projection, et par $\mathcal{P}$ la matrice de leur pondération. Précisément, $\mathcal{P}$ est une matrice diagonale proportionnelle aux effectifs de chaque classe. Après centrage des colonnes de $\mathcal{Q}$ pour la métrique $\mathcal{P}$, la matrice de covariance $\mathcal{V}$ entre les opérateurs projetés est donnée par la formule :

$$\mathcal{V} = \frac{1}{N} {}^t\mathcal{Q}\mathcal{Q} \in \mathbb{R}^{m \times m} \tag{3.76}$$

L'ADL détermine les axes pour lesquels l'inertie inter-classe, c'est à dire la mesure de dispersion des centres de gravité des classes, est maximisée. L'ADL du tableau $\mathcal{Q}$ est équivalente à appliquer l'ACP centrée du triplet statistique $(\mathcal{G}, \mathcal{V}^{-1}, \mathcal{P})$ [Sap90]. La matrice $\mathcal{V}^{-1}$ correspond à la métrique de Mahalanobis dans le sous-espace de $\mathbb{R}^m$ des centres de classes. Les axes discriminants sont donc les vecteurs propres $\mathfrak{U}$ associés aux valeurs propres $\mathfrak{L}$ de l'ACP considérée :

$$({}^t\mathcal{G}\mathcal{P}\mathcal{G}\mathcal{V}^{-1})\mathfrak{U} = \mathfrak{U}\mathfrak{L} \tag{3.77}$$

On a au plus $K - 1$ valeurs propres non nulles. D'après la formule (3.72), les opérateurs Γ_k peuvent être reconstruits à l'aide d'une combinaison linéaire des O_k pondérée par les valeurs des projections correspondantes. Avec cette remarque, on en déduit la

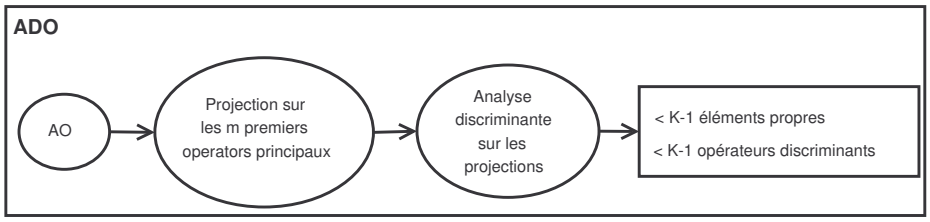

FIG. 3.8 – Principe de l'analyse discriminante d'opérateurs de covariance relationnelle.

construction des opérateurs discriminants :

$$\mathcal{D}_k = \sum_{i=1}^{m} O_i \mathfrak{U}_{ik} \in \mathbb{R}^{p\times p}\,, \quad 1 \leq k \leq K \tag{3.78}$$

Si le nombre de projections est inférieur au nombre de sous-tableaux ($m < N$), alors la construction de ces opérateurs discriminants est *partielle* puisque l'on omet dans l'analyse la part d'information fournie par les projections sur les $N - m$ derniers axes. De la même manière, on détermine la projection des opérateurs discriminants partiels sur la base des opérateurs principaux :

$$\begin{aligned} < \mathcal{D}_k, O_{k'} >_{HS} &= < \mathcal{D}_k, \frac{\widetilde{O}_{k'}}{\sqrt{\mathcal{L}_{k'}}} >_{HS} = \frac{1}{\sqrt{\mathcal{L}_{k'}}} < \sum_{i=1}^{N} \mathfrak{U}_{ik}\Gamma_i, \sum_{j=1}^{N} \mathcal{U}_{jk'}\Gamma_j >_{HS} \\ &= \frac{1}{\sqrt{\mathcal{L}_{k'}}} \sum_{i,j=1}^{N} \mathfrak{U}_{ik}\mathcal{H}(i,j)\mathcal{U}_{jk'} = \frac{1}{\sqrt{\mathcal{L}_{k'}}} {}^t\mathfrak{U}_k \mathcal{H}\mathcal{U}_{k'} = \sqrt{\mathcal{L}_k}\, {}^t\mathfrak{U}_k \mathcal{U}_{k'} \end{aligned} \tag{3.79}$$

L'encadré suivant donne les différentes étapes de la procédure.

Analyse discriminante d'opérateurs de covariance relationnelle

Données :
- Tableau $\mathfrak{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_N)$.
- Graphes de relation binaire G_k de chaque sous-tableau $\mathbb{X}_k$.
- Métriques M_k sur les individus de $\mathbb{X}_k$.
- Nombre m de premières projections et nombre K de classes.

Sorties :
- Éléments propres $(\mathfrak{U}, \mathfrak{L})$ de ${}^t\mathcal{G}\mathcal{P}\mathcal{G}\mathcal{V}^{-1}$.
- Opérateurs discriminants $\mathcal{D}$.
- Matrice de covariance $\mathcal{V}$ entre les opérateurs projetés.
- Opérateurs Γ.
- Opérateurs principaux O.

Procédure :

1. $[\mathcal{U}, \mathcal{L}, \Gamma, O] = AO(\mathfrak{X})$.
2. Calcul de la matrice des m premières projections $\mathcal{Q} = \mathcal{U}\mathcal{L}^{1/2}(:, 1:m)$.
3. Calcul des centres de gravité des classes $\mathcal{G}$.
4. Calcul de la matrice de covariance des projections $\mathcal{V} = \frac{1}{N}{}^t\mathcal{Q}\mathcal{Q}$.
5. Calcul de la matrice diagonale des fréquences des effectifs des classes $\mathcal{P}$.
6. Calcul des éléments propres $(\mathfrak{U}, \mathfrak{L})$ de ${}^t\mathcal{G}\mathcal{P}\mathcal{G}\mathcal{V}^{-1}$.
7. Calcul des opérateurs discriminants $\mathcal{D}_k = \sum_{i=1}^{m} O_i \mathfrak{U}_{ik}\ , \quad 1 \leq k \leq K$.
8. Calcul des projections $\mathcal{L}^{1/2}{}^t\mathfrak{U}\mathcal{U}$ des Γ_k sur les opérateurs discriminants.

TAB. 3.4 – Analyse discriminante d'opérateurs de covariance relationnelle.

3.7 Conclusion

Dans ce chapitre, nous avons présenté la covariance relationnelle comme une signature employée dans les méthodes factorielles du second ordre. L'ACP relationnelle est un cas général de l'ACP standard et considère la métrique du laplacien pour tenir compte de n'importe quelle structure extra-probabiliste sur l'espace des individus. L'ACP relationnelle est un cas particulier de l'ACP fonctionnelle et de l'ACP à noyaux. Pour l'ACP fonctionnelle, l'information relationnelle est introduite en appliquant aux fonctions de base l'opérateur racine carrée du laplacien, qui possède par ailleurs les mêmes propriétés. On se ramène ensuite à procéder à une ACP fonctionnelle standard avec cette nouvelle base. L'ACP à noyaux construit des projections non linéaires en appliquant une ACP standard dans un espace de dimension supérieure. Dans ce cas, on établit à l'aide du schéma de dualité que l'information exogène est directement "transmise" aux observations projetées. Le noyau relationnel s'exprime sous la forme matricielle mais ne peut s'exprimer sous la forme d'une application par perte de certaines propriétés induisant un produit scalaire.

L'analyse d'opérateurs de covariance relationnelle est une méthode relevant de l'analyse factorielle, qui étudie à des fins descriptives un tableau à trois entrées, où chaque sous-tableau est muni d'une relation binaire agissant sur ses observations. L'analyse d'opérateurs de covariance relationnelle est une variante de STATIS et recherche une base d'opérateurs principaux de l'espace engendré par les opérateurs de covariance relationnelle pondérés. Une extension naturelle de cette méthode est l'analyse discriminante des opérateurs de covariance relationnelle dans laquelle nous appliquons une analyse factorielle discriminante standard sur les projections des opérateurs sur les axes principaux, et ce afin d'éviter des calculs lourds et un mauvais conditionnement des matrices.

Chapitre 4

Approches probabilistes

4.1 Introduction

Nous venons d'exposer des techniques factorielles de réduction de dimension fondées sur l'inertie et appliquées à la classification de données multidimensionnelles. Implicitement, ce travail concerne l'étude des termes dans la décomposition commune d'un modèle linéaire gaussien $X_t = \varphi(t) + \epsilon_t$, où X_t se divise en un terme tendanciel $\varphi(t)$ et un autre résiduel ϵ_t. Le terme tendanciel a été étudié aux chapitres précédents avec les concepts d'ACP et d'analyse factorielle d'opérateurs fondées sur la covariance relationnelle. Dans ce chapitre, nous proposons une analyse complémentaire de celle menée sur $\varphi(t)$, en nous intéressant à l'étude du terme supposé gaussien centré $\epsilon_t = X_t - \varphi(t)$ (*i.e.* de X_t privé de sa tendance), à l'aide d'approches probabilistes. L'hypothèse de gaussianité, habituellement admise [RW06], est fondamentale puisqu'elle permet de résumer l'ensemble des caractéristiques de ϵ_t à l'aide de la covariance.

Le présent chapitre se divise en deux parties. D'une part, nous considérons ϵ_t comme un vecteur gaussien centré, ce qui induit une distribution de Wishart pour la covariance empirique. Nous étudions une approche bayésienne en complément des méthodes géométriques déjà évoquées. Nous proposons une méthode de classification paramétrique supervisée ou non supervisée, de matrices de covariance à partir de la loi de Wishart [HSJL07]. Dans les deux approches, chaque classe (ou cluster dans le cas non supervisé) est modélisée par une loi de Wishart de paramètres à estimer. Dans l'approche supervisée, on dispose d'un échantillon de matrices de covariance étiqueté par des classes. La seconde non supervisée utilise une approche de type EM. La deuxième partie de ce chapitre est consacrée à l'analyse fractale et au caractère autosimilaire que peut présenter ϵ_t, ce qui impose à nouveau de supposer une forme particulière pour les matrices de

covariance analysées. Nous proposons un protocole expérimental afin d'estimer l'autosimilarité sur des données soumises à des contraintes de grande dimension et d'irrégularité d'échantillonnage. Pour la première contrainte, nous étudions comment mesurer l'indice d'autosimilarité sur une courbe plongée dans un espace à plusieurs dimensions et paramétrée par le temps. Nous proposons de ce fait une méthode de segmentation multidimensionnelle basée sur le paramètre de Hurst.

Dans la première partie, nous rappelons les définitions et les propriétés de base de la distribution de Wishart. Nous explicitons ensuite les deux méthodes de classification de matrices de covariance en évoquant les contraintes dans le choix des paramètres et la sélection du modèle. Nous terminons par l'étude d'un exemple simple de simulation permettant de valider les méthodes proposées. La seconde partie du chapitre fait un rappel des notions de base des processus autosimilaires puis nous énonçons une méthode de simulation de signaux autosimilaires irréguliers dans le temps. Ensuite, nous vérifions en détail les propriétés asymptotiques de l'estimateur et son comportement pour des signaux de faible longueur et ce, avec différents types d'irrégularité temporelle [HSJL]. Enfin, nous proposons une méthode de segmentation adaptée au traitement de l'échantillonnage irrégulier et décrivons un algorithme *ad hoc* d'estimation par maximum de vraisemblance.

4.2 Distribution de Wishart : modèle proposé

4.2.1 Définition et propriétés

La distribution de Wishart [Wis28] est une version multidimensionnelle de la loi du χ^2 qui caractérise les estimateurs des matrices de covariance de lois gaussiennes multivariées :

Définition 5 (Loi de Wishart centrée) *Soit ε un n-échantillon constitué de réalisations indépendantes d'un vecteur aléatoire gaussien de dimension p, de moyenne nulle et de matrice de covariance Σ. Alors ${}^t\varepsilon\varepsilon$ suit une loi de Wishart de degré de liberté n et de matrice d'échelle Σ. On notera ${}^t\varepsilon\varepsilon \sim W_p(n, \Sigma)$.*

Dans ce qui suit, nous considérons que le nombre de degrés de liberté (ddl) est supérieur à la dimension du vecteur. Sinon, on perd l'hypothèse d'indépendance des réalisations puisque la taille de l'échantillon est plus petite que le nombre de variables. Cette singularité fait référence aux matrices *anti-Wishart* dont une étude a été récemment menée

[YZ02]. Parmi les distributions multidimensionnelles associées, on peut citer la loi du T^2 de Hotelling et la loi Λ de Wilks, généralisant respectivement la loi de Student et de Fisher-Snedecor [Sap90]. La distribution de Wishart est une loi s'appliquant à des matrices calculées à partir d'échantillons du même vecteur gaussien. Elle possède comme densité :

$$f(M;\Sigma,n) = \frac{|M|^{\frac{n-p-1}{2}} \exp\left\{-\frac{1}{2}trace(\Sigma^{-1}M)\right\}}{2^{\frac{np}{2}}\pi^{\frac{p(p-1)}{4}}|\Sigma|^{\frac{n}{2}}\prod_{j=1}^{p}\Gamma\left(\frac{1}{2}(n-j+1)\right)} \tag{4.1}$$

où Γ est la fonction définie par :

$$\Gamma(x) = \int_0^{+\infty} t^{x-1}e^{-t}dt\ , \quad x>0 \tag{4.2}$$

L'espérance d'une loi de Wishart à pour un nombre de ddl fixé est la suivante :

$$E(M) = n\Sigma \tag{4.3}$$

Dans la littérature, on trouve de nombreux travaux théoriques sur la distribution de Wishart dont l'inverse est extensivement utilisé comme modèle a priori pour les méthodes bayésiennes [Row03]. Cependant, les applications en reconnaissance des formes sont récentes et peu nombreuses [SNC03]. Voici quelques propriétés relatives à la distribution de Wishart [GN00] :

Propriété 7 (Transformation) *Pour toute matrice $\mathcal{M}$ de dimension $p \times p'$ de rang p', on a :*

$$W \sim W_p(n,\Sigma) \Longrightarrow {}^t\mathcal{M}W\mathcal{M} \sim W_{p'}(n, {}^t\mathcal{M}\Sigma\mathcal{M}) \tag{4.4}$$

Cette propriété caractérise le passage d'une distribution de Wishart à une autre par simple produit matriciel. Comme on le verra, cela peut être commode dans le cadre de la simulation et en particulier lorsqu'une matrice d'échelle est l'identité.

Propriété 8 (Additivité) *Si $W_1 \sim W_p(n_1,\Sigma)$ et $W_2 \sim W_p(n_2,\Sigma)$ sont des variables aléatoires de Wishart indépendantes, alors :*

$$W_1 + W_2 \sim W_p(n_1+n_2,\Sigma) \tag{4.5}$$

L'indépendance entre deux matrices de Wishart signifie que les épreuves de tirage des échantillons correspondants ont été réalisées de manière indépendante, conformément à la procédure de l'échantillonnage aléatoire simple [Sap90].

Propriété 9 (Quadratique) *Si $W \sim W_p(n, \Sigma)$, alors pour tout vecteur x de taille p, la norme au carré de x pour le produit scalaire associé à W est une réalisation d'une loi du χ^2 à n ddl :*

$$\frac{{}^t xWx}{{}^t x\Sigma x} \sim \chi^2_n \tag{4.6}$$

Cette propriété montre que la loi du χ^2 est un cas particulier de la distribution de Wishart. En effet, si $p = 1$, on retrouve l'expression de la densité de probabilité du χ^2 :

$$f(\mathrm{m}; n, \sigma^2) = \frac{\mathrm{m}^{\frac{n-2}{2}} \exp\left\{-\frac{\mathrm{m}}{2\sigma^2}\right\}}{2^{\frac{n}{2}} \sigma^n \Gamma\left(\frac{n}{2}\right)} \tag{4.7}$$

Propriété 10 (dite de Cholesky) *Soient $W \sim W_p(n, I_p)$ et T la matrice triangulaire inférieure dans la décomposition de Cholesky de W. Alors les éléments diagonaux au carré T^2_{ii} sont des réalisations de loi du χ^2 à $n-i+1$ ddl. Les éléments extra-diagonaux inférieurs sont des réalisations de variables aléatoires gaussiennes centrées réduites.*

La loi du χ^2 normalisée est la distribution d'une variance empirique calculée sur des réalisations indépendantes de lois normales centrées réduites, ce qui correspond aux éléments diagonaux de $T{}^tT$. On remplit donc la diagonale de T à l'aide de cette loi en supprimant successivement un ddl puisque la trace de $T{}^tT$ est constante. Comme on a fixé n ddl, l'élément diagonal T_{ii} est donné par la racine carrée d'une réalisation de la loi du χ^2 à $n - i + 1$ ddl :

$$(T_{ii})^2 \sim \chi^2_{n-i+1}\,, \quad 1 \leq i \leq p \tag{4.8}$$

Les éléments extra-diagonaux inférieurs sont supposés suivre une loi normale centrée réduite :

$$T_{ij} \sim \mathcal{N}(0, 1)\,, \quad 1 \leq i < j \leq p \tag{4.9}$$

Les éléments diagonaux de $T{}^tT$ suivent une loi du χ^2 conformément à la caractérisation des variances empiriques. Par définition de la loi de Wishart centrée, les éléments extra-diagonaux sont des covariances empiriques entre des n-échantillons de variables gaussiennes centrées.

En pratique, on simule une matrice de Wishart à partir de la loi du χ^2 en combinant les propriétés 7 et 10 citées plus haut. Supposons donnée une matrice d'échelle Σ de

dimension p et le nombre n de ddl. Pour simuler une matrice de Wishart, on procède à la simulation de la matrice T décrite précédemment :

$$T = \begin{pmatrix} \chi^2_n & 0 & \cdots & 0 \\ \mathcal{N}(0,1) & \chi^2_{n-1} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \mathcal{N}(0,1) & \cdots & \mathcal{N}(0,1) & \chi^2_{n-p+1} \end{pmatrix} \in \mathbb{R}^p \times \mathbb{R}^p \tag{4.10}$$

D'après la propriété 10, on a donc :

$$T\,{}^tT \sim W_p(n, I_p) \tag{4.11}$$

Nous appliquons ensuite la propriété 7 en calculant la racine carrée de Σ que nous choisissons comme valeur pour $\mathcal{M}$. On en déduit une matrice de Wishart, notée W :

$$W = {}^t\Sigma^{1/2}(T\,{}^tT)\Sigma^{1/2} \sim W_p(n, {}^t\Sigma^{1/2} I_p \Sigma^{1/2}) = W_p(n, \Sigma) \tag{4.12}$$

En guise d'illustration, un ensemble de matrices de covariance sont simulées avec $p = 2$ issues de la loi de Wishart de matrice d'échelle :

$$\Sigma = \begin{pmatrix} 0.2 & 0.3 \\ 0.3 & 2.0 \end{pmatrix}$$

Les valeurs propres de Σ sont égales à 2.049 et 0.151. Chacune des matrices simulées est normalisée par le ddl correspondant de manière à obtenir une matrice de covariance. La figure 4.1 donne une représentation des matrices de covariance pour $p = 2$ à l'aide d'ellipses tracées sur un plan. Chaque ellipse est centrée à l'origine et son orientation est donnée par la base des vecteurs propres de la matrice de covariance. Les longueurs des axes sont égales à la racine carrée de la valeur propre associée. Sur la figure, l'ellipse en pointillé identifie Σ. En (a), cinq matrices de Wishart sont simulées avec un ddl allant de 6 à 10. En (b) est présenté cinq autres matrices avec un ddl constant égal à 30. Enfin, les cinq matrices illustrées en (c) ont pour ddl égal à 100. Plus le ddl de la matrice simulée est grand, plus on se rapproche du cas asymptotique, c'est à dire que l'échantillon devient représentatif du vecteur aléatoire. Par conséquent, la matrice de covariance empirique tend vers la matrice théorique Σ. Les matrices simulées avec un faible ddl donnent quant à elles une mauvaise simulation.

Lorsqu'il existe des liens entre les observations (supposées indépendantes) de l'échantillon, autrement dit, si ${}^t\mathbb{X}\mathbb{X}$ suit une loi de Wishart, peut-on affirmer qu'il en est de même pour ${}^t\mathbb{X}S\mathbb{X}$ pour S matrice symétrique ? En fait, ce problème a été traité dans un cadre plus général [MW03] où Masaro *et al.* ont énoncé un théorème d'équivalence sur la nature Wishart de ${}^t\mathbb{X}S\mathbb{X}$, si S est une matrice symétrique quelconque. Le théorème est basé sur une décomposition matricielle de Σ où sont imposées de nombreuses

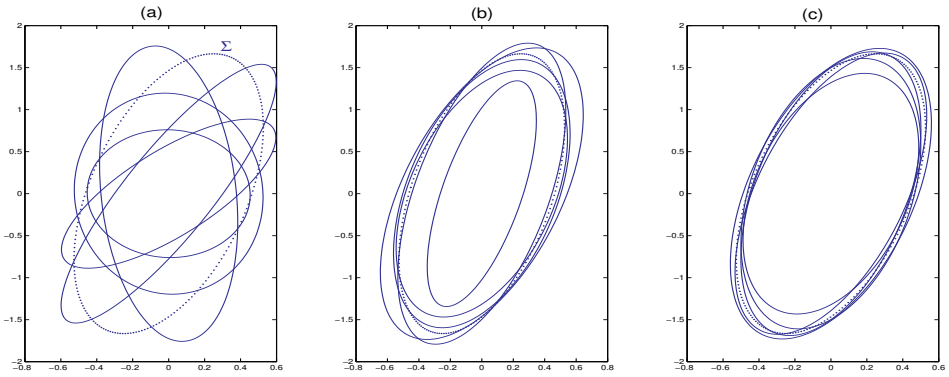

FIG. 4.1 – Simulation et représentation bidimensionnelle de matrices de Wishart - (a) $n = 6,\ 7,\ 8,\ 9\ 10$; (b) $n = 30$; (c) $n = 100$.

conditions d'équivalence qui, de manière générale, ne sont pas vérifiées expérimentalement. Par conséquent, nous considérons dans ce qui suit les matrices de la forme $\mathbb{XX}$, conformes à la définition classique. En particulier, on ne supposera pas ici l'existence d'une relation binaire agissant sur les observations ($S = \frac{1}{\kappa}\Delta$).

4.2.2 Classification supervisée

Nous proposons un modèle probabiliste de classification supervisée basé sur la distribution de Wishart. Comme pour l'analyse factorielle d'opérateurs vue au chapitre précédent, nous reprenons le tableau $\mathfrak{X}$ constitué de l'empilement des N sous-tableaux $\mathbb{X}_i$ de taille respective $n_i \times p$. On notera par ε_i le sous-tableau $\mathbb{X}_i$ obtenu après suppression de la tendance locale des variables et ce afin d'être le plus possible conforme à la gaussianité des variables aléatoires. D'une manière générale, on fixe C le nombre de classes avec lesquelles on regroupe les N sous-tableaux ε_i. On attribue ensuite à chacune des classes une matrice d'échelle à estimer. La signature Γ_i de chaque sous-tableau est obtenue en fixant la métrique M_i et le graphe de la relation binaire de la manière suivante :

$$M_i = n_i I_p \quad \text{et} \quad g_i(\omega, \omega') = 1\,, \quad \forall \omega, \omega' \in \Omega \tag{4.13}$$

Ce qui nous donne comme signature :

$$\Gamma_i = \frac{1}{\kappa_i} {}^t\varepsilon_i \Delta_i \varepsilon_i M_i \tag{4.14}$$

Puisque le graphe considéré est complet, la covariance relationnelle coïncide avec la covariance et donc :

$$\Gamma_i = \left(\frac{1}{n_i} {}^t\varepsilon_i \varepsilon_i\right)(n_i I_p) = {}^t\varepsilon_i \varepsilon_i \in \mathbb{R}^{p \times p} \tag{4.15}$$

La classe k est constituée des N_k opérateurs[1] $(\Gamma_{ik})_{i=1}^{N_k}$ supposés suivre une loi de Wishart de ddl n_{ik} et de matrice d'échelle Σ_k :

$$\Gamma_{ik} = {}^t\varepsilon_{ik}\varepsilon_{ik} \sim W_p(n_{ik}, \Sigma_k) \quad \text{avec} \quad \varepsilon_{ik} \sim \mathcal{N}_p(0, \Sigma_k) \tag{4.16}$$

Il est important de noter que le nombre d'individus n_{ik} des sous-tableaux est donné, ce qui détermine les ddl des lois de Wishart considérées. Les valeurs des n_{ik} ne sont donc pas à être estimées dans le modèle. Une estimation de la matrice d'échelle Σ_k quant à elle est calculée en maximisant le logarithme de la vraisemblance et en admettant que les ε_{ik} sont indépendants :

$$\begin{aligned} L(\Sigma_k) &= P(\Gamma_{1k}, \ldots, \Gamma_{N_k k}|\Sigma_k) = \prod_{i=1}^{N_k} f(\Gamma_{ik}; \Sigma_k, n_{ik}) \\ \text{et} \quad \log L(\Sigma_k) &= \sum_{i=1}^{N_k} \log f(\Gamma_{ik}; \Sigma_k, n_{ik}) \end{aligned} \tag{4.17}$$

où f est la densité de la loi de Wishart. On a :

$$\begin{aligned} \log f(M; \Sigma, n) &= \frac{n-p-1}{2}\log|M| - \frac{1}{2}trace(\Sigma^{-1}M) \\ &- \frac{np}{2}\log(2) - \frac{p(p-1)}{4}\log(\pi) - \frac{n}{2}\log|\Sigma| - \sum_{j=1}^{p}\log\Gamma\left(\frac{1}{2}(n-j+1)\right) \end{aligned} \tag{4.18}$$

En utilisant les formules de dérivées matricielles suivantes [PP06] et puisque les matrices sont symétriques :

$$\frac{\partial}{\partial\Sigma}trace(\Sigma^{-1}\Gamma) = -{}^t(\Sigma^{-1}\Gamma\Sigma^{-1}) = -\Sigma^{-1}\Gamma\Sigma^{-1} \quad \text{et} \quad \frac{\partial}{\partial\Sigma}\log|\Sigma| = {}^t\Sigma^{-1} = \Sigma^{-1} \tag{4.19}$$

On obtient :

$$\begin{aligned} \frac{\partial}{\partial\Sigma_k}\log f(\Gamma_{ik}; \Sigma_k, n_{ik}) &= \frac{\partial}{\partial\Sigma_k}\left(-\frac{1}{2}trace(\Sigma_k^{-1}\Gamma_{ik}) - \frac{n_{ik}}{2}\log|\Sigma_k|\right) \\ &= \frac{1}{2}\Sigma_k^{-1}\Gamma_{ik}\Sigma_k^{-1} - \frac{n_{ik}}{2}\Sigma_k^{-1} \end{aligned} \tag{4.20}$$

La condition nécessaire vérifiée par le maximum de vraisemblance nous donne :

$$\frac{\partial\log L(\Sigma_k)}{\partial\Sigma_k} = \sum_{i=1}^{N_k}\frac{\partial}{\partial\Sigma_k}\log f(\Gamma_{ik}; \Sigma_k, n_{ik}) = \sum_{i=1}^{N_k}\left(\frac{1}{2}\Sigma_k^{-1}\Gamma_{ik}\Sigma_k^{-1} - \frac{n_{ik}}{2}\Sigma_k^{-1}\right) = 0 \tag{4.21}$$

En simplifiant par $\frac{1}{2}\Sigma_k^{-1}$, on aboutit à :

$$\widehat{\Sigma}_k = \frac{\sum_{i=1}^{N_k}\Gamma_{ik}}{\sum_{j=1}^{N_k}n_{jk}} = \frac{\sum_{i=1}^{N_k}n_{ik}\left(\frac{1}{n_{ik}}{}^t\varepsilon_{ik}\varepsilon_{ik}\right)}{\sum_{j=1}^{N_k}n_{jk}} \tag{4.22}$$

[1] L'indice k que l'on ajoute à Γ_i signifie que l'opérateur appartient à la classe k.

Ce résultat est naturel puisque l'estimée $\widehat{\Sigma}_k$ est égale à la somme des matrices de covariance empirique pondérée par les tailles normalisées des échantillons. D'un point de vue géométrique, $\widehat{\Sigma}_k$ correspond au barycentre des éléments de la classe k. Une fois les matrices d'échelle de chaque classe estimée, la procédure de classification d'un sous-tableau ε de taille $n \times p$ consiste à rechercher sa probabilité d'appartenance à la classe k, notée $P(k|\varepsilon)$. D'après la règle de Bayes :

$$P(k|\varepsilon) \propto P(k)P(\varepsilon|k) \tag{4.23}$$

où $P(\varepsilon|k)$ est la probabilité d'observer ε dans la classe k, *i.e.* que l'opérateur ${}^t\varepsilon\varepsilon$ associé à ε soit conforme à une densité de Wishart $W_p\left(n, \widehat{\Sigma}_k\right)$:

$$P(\varepsilon|k) = f({}^t\varepsilon\varepsilon; \widehat{\Sigma}_k, n) \tag{4.24}$$

La probabilité $P(k)$ est le poids de la classe k. Puisque les fréquences dans l'échantillon ne sont pas nécessairement représentatives, l'équiprobabilité des classes est considérée plutôt que la fréquence empirique de leur effectif :

$$P(k) = \frac{1}{C}, \quad 1 \leq k \leq C \tag{4.25}$$

Le critère d'affectation pour ε consiste à calculer l'indice de la classe pour lequel le logarithme de la probabilité $P(k|\varepsilon)$ est maximum :

$$\underset{k=1,\ldots,C}{\operatorname{argmax}} \log P(k|\varepsilon) = \underset{k=1,\ldots,C}{\operatorname{argmax}} \log P(\varepsilon|k) \tag{4.26}$$

La classification supervisée est résumée à la table 4.1.

4.2.3 Classification non supervisée : algorithme EM pour matrices de covariance

Dans le cadre de la classification de matrices de covariance, des études [FFPL01, LGA+99] ont proposé une approche non supervisée par k-means appliquée à des images polarimétriques. Dans ce paragraphe, nous présentons une méthode de classification non supervisée de matrices de covariance à l'aide de l'algorithme Expectation-Maximisation (EM) [DLR77]. Il s'agit d'une procédure itérative de maximisation d'une vraisemblance sur des données observées. Comme la vraisemblance est optimisée à chaque itération, cela permet de conclure à la convergence de l'algorithme [Dan98]. Reigber *et al.* [RMNFF07] ont récemment formulé la procédure dans un cas très restrictif avec équipondération des clusters et sans prise en compte dans le modèle des ddl des lois

Classification supervisée par la loi de Wishart

Entrées :
- Sous-tableau sans tendance ε de taille n.
- Nombre C de classes.
- Base de données : opérateurs $(\Gamma_{ik})_{i=1}^{N}$ répartis a priori dans les C classes.
- Effectifs $(N_k)_{k=1}^{C}$ des classes.

Sortie :
- Classe de ε et vraisemblance associée.

Procédure :

Estimation de la matrice d'échelle Σ_k associée à la loi de Wishart de la classe k :

$$\widehat{\Sigma}_k = \frac{\sum_{i=1}^{N_k} \Gamma_{ik}}{\sum_{j=1}^{N_k} n_{jk}}, \quad 1 \leq k \leq C$$

Attribution d'une classe à ε :

$$Class(\varepsilon) = \underset{k=1,\ldots,C}{\operatorname{argmax}} \left(\log f({}^t\varepsilon\varepsilon; \widehat{\Sigma}_k, n) \right)$$

TAB. 4.1 – Classification supervisée par la loi de Wishart.

de Wishart associées. Nous proposons un cadre général dans lequel la totalité des paramètres est estimée : probabilités d'appartenance, poids des clusters, matrice de covariance des lois et vraisemblance. Néanmoins, cela ne va pas sans poser des problèmes pratiques dans l'estimation et nous y reviendrons plus tard dans la partie consacrée aux applications.

Dans le cas non supervisé, on détermine a posteriori quels éléments constituent chacun des clusters. Ces derniers sont construits à l'aide de la procédure EM. On choisit K un nombre de clusters $Cl_1, \ldots, Cl_K$ à estimer pour la classification et on dispose d'un nombre N d'opérateurs $(\Gamma_i)_{i=1}^N$ sans tendance correspondant aux sous-tableaux ε_i de taille $n_i \times p$. Pour le cluster Cl_k, on pose $\Theta_k = (\pi_k, \Sigma_k)$, où π_k est le poids du cluster et Σ_k la matrice de covariance de la distribution de Wishart associée. On introduit dans le modèle une indicatrice d'appartenance d'un sous-tableau à un cluster :

$$z_{ik} = \begin{cases} 1 & \text{si } \Gamma_i \in Cl_k \\ 0 & \text{sinon} \end{cases} \tag{4.27}$$

La vraisemblance est donnée par l'expression suivante où l'on introduit de façon naturelle les termes z_{ik} :

$$L(\Theta) = \prod_{i=1}^{N} \prod_{k=1}^{K} (\pi_k P(\Gamma_i|\Theta_k))^{z_{ik}} \tag{4.28}$$

$$\log L(\Theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} z_{ik} \log(\pi_k P(\Gamma_i|\Theta_k)) \tag{4.29}$$

La quantité $P(\Gamma_i|\Theta_k)$ est la probabilité d'appartenance de $\mathbb{X}_i$ à Cl_k, connaissant les paramètres Θ_k :

$$P(\Gamma_i|\Theta_k) = f(\Gamma_i; \Sigma_k, n_i) \tag{4.30}$$

La procédure EM consiste à maximiser l'espérance du logarithme de la vraisemblance :

$$E \log L(\Theta) = \sum_{i=1}^{N} \sum_{k=1}^{K} E(z_{ik}) \log(\pi_k P(\Gamma_i|\Theta_k)) \tag{4.31}$$

Le terme $E(z_{ik})$ est estimé par $\widehat{z}_{ik}$. À chaque étape et ce jusqu'à convergence, on réestime les paramètres z_{ik} (étape 'Expectation') puis π_k et Σ_k (étape 'Maximisation') de la manière suivante. Les estimés de π_k sont calculés en maximisant le logarithme de la vraisemblance de la formule (4.29) avec la contrainte lagrangienne $\sum_{k=1}^{K} \pi_k = 1$:

$$\widehat{\pi}_k = \max_{\pi_k} \left(\log L(\Theta) - \lambda \left(\sum_{k=1}^{K} \pi_k - 1 \right) \right) \tag{4.32}$$

Pour k fixé, on obtient en dérivant :

$$\begin{aligned} \frac{\partial}{\partial \pi_k}\left(\log L(\Theta) - \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right) &= \sum_{i=1}^{N} z_{ik}\frac{\partial \log(\pi_k P(\Gamma_i|\Theta_k))}{\partial \pi_k} - \lambda \\ &= \sum_{i=1}^{N} z_{ik}\frac{1}{\pi_k} - \lambda = 0 \end{aligned} \tag{4.33}$$

Donc :

$$\pi_k = \frac{\sum_{i=1}^{N} z_{ik}}{\lambda} \tag{4.34}$$

La dérivée de l'expression (4.32) par rapport à λ nous permet d'écrire :

$$\sum_{k=1}^{K}\pi_k = 1 = \frac{\sum_{k=1}^{K}\sum_{i=1}^{N} z_{ik}}{\lambda} = \frac{N}{\lambda} \tag{4.35}$$

donc $\lambda = N$ et au final :

$$\widehat{\pi}_k = \frac{\sum_{i=1}^{N} z_{ik}}{N} \tag{4.36}$$

On établit la formule donnant l'estimée de Σ_k de façon analogue à celle obtenue pour la classification supervisée où interviennent de plus les probabilités d'appartenance z_{ik} :

$$\widehat{\Sigma}_k = \frac{\sum_{i=1}^{N} z_{ik}\Gamma_i}{\sum_{j=1}^{N} z_{jk} n_j} \tag{4.37}$$

La probabilité z_{ik} caractérise l'appartenance de l'opérateur Γ_i à Cl_k en tenant compte du poids du cluster. Celle-ci est obtenue en normalisant par la somme de toutes les probabilités :

$$\widehat{z}_{ik} = \frac{\pi_k P(\Gamma_i|\Theta_k)}{\sum_{l=1}^{K}\pi_l P(\Gamma_i|\Theta_l)} \tag{4.38}$$

La table 4.2 donne une synthèse de la procédure EM pour les matrices de covariance.

4.2.4 Difficultés en classification

Dans ce type d'approche, on rencontre classiquement un certain nombre de problèmes, communs à tous les algorithmes de classification. Ces contraintes sont principalement liées à la nature même des données traitées :

- Initialisation de l'algorithme.
- Dimensionnalité des données.
- Valeurs manquantes.

Algorithme EM pour matrices de covariance

Entrées :
- Seuil de convergence η.
- Opérateurs $(\Gamma_i)_{i=1}^N$ des ε_i.
- Tailles $(n_i)_{i=1}^N$ des ε_i.
- Nombre de clusters K à estimer.
- Valeurs initiales $\Theta^{(0)}$ des Θ_k.

Sortie :
- Probabilités d'appartenance estimées $\widehat{z}_{ik}$.

Procédure :

$t \leftarrow 0$

E-step : Estimation des $z_{ik}^{(t)}$

$$\widehat{z}_{ik}^{(t)} = \frac{\widehat{\pi}_k^{(t)} P(\Gamma_i|\widehat{\Theta}_k^{(t)})}{\sum_{l=1}^K \widehat{\pi}_l^{(t)} P(\Gamma_i|\widehat{\Theta}_l^{(t)})}\,, \quad 1 \leq i \leq N\,, \quad 1 \leq k \leq K$$

M-step : Estimation des $\Theta_k^{(t+1)} = (\widehat{\pi}_k^{(t+1)}, \widehat{\Sigma}_k^{(t+1)})$

$$\left\{ \begin{array}{lcl} \widehat{\pi}_k^{(t+1)} & = & \frac{\sum_{i=1}^N \widehat{z}_{ik}^{(t)}}{N} \\ \widehat{\Sigma}_k^{(t+1)} & = & \frac{\sum_{i=1}^N \widehat{z}_{ik}^{(t)} \Gamma_i}{\sum_{j=1}^N \widehat{z}_{jk}^{(t)} n_j} \end{array} \right. , \quad 1 \leq k \leq K$$

Jusqu'à $|E \log L(\widehat{\Theta}^{(t+1)}) - E \log L(\widehat{\Theta}^{(t)})| < \eta$

TAB. 4.2 – Algorithme EM pour matrices de covariance.

– Données aberrantes.

L'initialisation des paramètres est importante puisqu'elle influe sur l'optimisation de la fonctionnelle et peut amener à la convergence vers des minima locaux souvent très éloignés du minimum global. La plupart du temps, l'exploration exhaustive de toutes les configurations possibles n'est pas envisageable car cela nécessite des calculs extrêmement coûteux. L'étape d'initialisation peut se faire en choisissant les probabilités d'appartenance par tirage uniforme. Les clusters sont ainsi constitués de manière aléatoire. C'est de cette manière que nous initialisons notre procédure. Nous imposons également une contrainte selon laquelle aucun cluster n'est vide à la première itération. Pour rendre la procédure moins sensible aux minima locaux, nous procédons à une variante de l'algorithme EM, appelée *Stochastic EM* [CD85], dans laquelle on rajoute une étape intermédiaire (entre la ré-estimation et la maximisation) qui consiste à faire une classification par tirage aléatoire des z_{ik} selon une loi multinomiale. On peut également envisager une méthode subsidiaire de classification, soit de type hiérarchique en sélectionnant une configuration suivant le nombre de clusters désiré, soit de type centres mobiles, ou encore de type *fuzzy C-means*, analogue aux centres mobiles en rajoutant un degré d'appartenance aux clusters.

La seconde contrainte est la dimensionnalité des données, formalisée par le phénomène de Hugues [Bel61] : la taille minimale de l'échantillon d'apprentissage augmente exponentiellement avec le nombre de paramètres libres à estimer. Dans le cas d'une classification de matrices de covariance, le nombre de paramètres libres est de $\frac{p(p+1)}{2}$ et le problème suscité par la dimension des données entre pleinement en jeu.

La présence de valeurs manquantes est un problème qui a notamment été déjà évoqué dans la section consacrée à la géostatistique. Ce problème peut être corrigé à l'aide d'astuces plus ou moins élémentaires :

– Reprise de l'acquisition des données.
– Estimation de la valeur par tirage aléatoire.
– Estimation par EM, régression, interpolation (cf krigeage en géostatistique) ...

Les données aberrantes apparaissent la plupart du temps suite à l'acquisition, au transfert ou à la compression des données. La difficulté est alors de pouvoir les détecter parmi l'ensemble des données. Le cas échéant, il est tout à fait possible de traiter le problème de la même façon qu'au cas précédent en les considérant comme des valeurs manquantes.

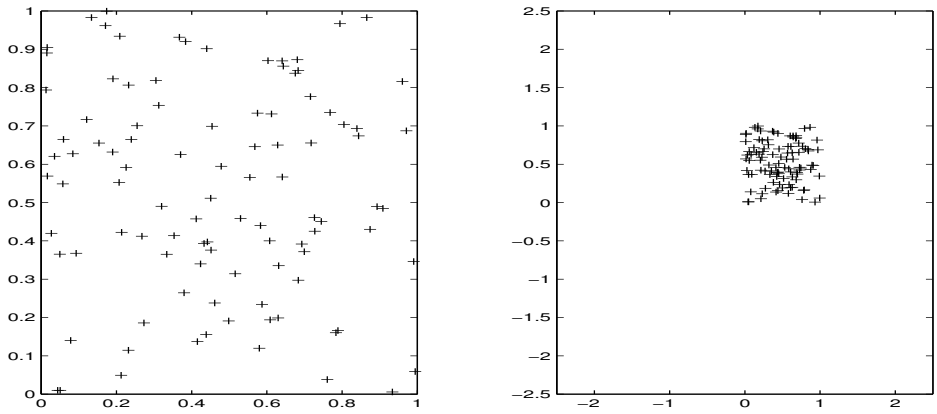

FIG. 4.2 – Tirage uniforme de 100 points à deux résolutions différentes. Y'a-t-il un ou plusieurs clusters ?

4.2.5 Sélection du modèle

Le choix du nombre de paramètres, et plus particulièrement le nombre de clusters à estimer, est un problème difficile dans les algorithmes de classification. Habituellement, le nombre de clusters est choisi de manière arbitraire par l'utilisateur. Cette tâche est néanmoins délicate comme le montre l'exemple de la figure 4.2. Un ensemble de 100 points ont été obtenus par tirage uniforme sur $[0,1]^2$. La figure présente le nuage de points à deux niveaux d'échelle différents. Sur le graphique de gauche, les points sont placés de manière disparate alors que sur le graphique de droite, ces mêmes points paraissent former un seul cluster. Simplement en visualisant les points à deux résolutions différentes, il semble déjà difficile de pouvoir décider d'un nombre optimal de clusters pour l'initialisation.

C'est pour répondre à ce type de problème que les critères d'information ont été introduits. Il en existe plusieurs approches mais le principe reste le même : minimiser une fonction objectif dépendant de la vraisemblance et pénalisée par un terme caractérisant la complexité du modèle à estimer. La fonction objectif est minimisée par rapport au nombre K de clusters. Dans ce travail, nous nous intéressons plus particulièrement au critère BIC (Bayesian Information Criterion) [Sch78]. La fonction objectif correspondante se formule de la manière suivante :

$$BIC(K) = -2\log L(\Theta) + \log(N)\check{n}(\Theta) \tag{4.39}$$

où $\check{n}(\Theta)$ est le nombre de ddl des paramètres (π_k et Σ_k) à estimer. Dans notre méthode,

ce nombre est donné par :

$$\begin{aligned} \check{n}(\Theta) &= ddl(\Theta) = ddl(\pi_k) + ddl(\Sigma_k) \\ &= (K-1) + K \times \frac{p(p+1)}{2} \end{aligned} \tag{4.40}$$

Le nombre $K-1$ est égal au ddl des poids des clusters, noté $ddl(\pi_k)$, privé de un du fait de la contrainte de somme unitaire. La quantité $K \times \frac{p(p+1)}{2}$ est le ddl des matrices d'échelle de chaque cluster, noté $ddl(\Sigma_k)$, égale au nombre de clusters multiplié par le nombre de ddl de chaque matrice. Pour K fixé, nous appliquons l'algorithme EM et déterminons la valeur de la vraisemblance finale, après convergence. On calcule ensuite la valeur de la fonction BIC et ainsi de suite pour toutes les valeurs de K considérées. La valeur minimale K_0 de la fonction BIC sera le nombre optimal de clusters pour appliquer la procédure.

4.2.6 Validation des méthodes par simulation

Pour conclure la première partie de ce chapitre, nous proposons l'étude d'un exemple simple. Nous considérons un ensemble de $C = 3$ clusters de matrices de dimension $p = 2$. Les matrices d'échelle sont les suivantes :

$$\Sigma_1 = \begin{pmatrix} 0.9950 & -0.610 \\ -0.610 & 0.523 \end{pmatrix} ; \Sigma_2 = \begin{pmatrix} 0.2 & 0.3 \\ 0.3 & 2.0 \end{pmatrix} ; \Sigma_3 = \begin{pmatrix} 0.9659 & 0.068 \\ 0.068 & 1.1138 \end{pmatrix}$$

Un ensemble de $N = 150$ matrices de Wishart (50 par cluster) est simulé. Les ddl correspondants sont choisis uniformément et vont de 7 à 32. La classification supervisée par la procédure leave-one out fournit un taux de bonne reconnaissance de l'ordre de 99.33%, soit une matrice mal classée parmi un total de 150. Pour l'algorithme EM, le nombre de cluster est fixé à $K = 3$ et le seuil de convergence est choisi à $\eta = 10^{-5}$. Après 279 itérations, l'algorithme converge et obtient les matrices d'échelle estimées suivantes :

$$\widehat{\Sigma}_1 = \begin{pmatrix} 0.8826 & -0.5330 \\ -0.5330 & 0.4657 \end{pmatrix} ; \widehat{\Sigma}_2 = \begin{pmatrix} 0.2095 & 0.2722 \\ 0.2722 & 1.6743 \end{pmatrix} ; \widehat{\Sigma}_3 = \begin{pmatrix} 0.9879 & 0.0678 \\ 0.0678 & 0.9365 \end{pmatrix}$$

Les ellipses correspondantes sont présentées en trait pointillé à la figure 4.3, en même temps que les matrices théoriques Σ_1, Σ_2 et Σ_3 en trait plein. On constate que les estimées ont leurs valeurs propres inférieures à celles des matrices théoriques. Cela s'explique par le fait que les probabilités d'appartenance $\widehat{z}_{ik}$ que l'on obtient à la fin de la procédure ne sont pas égales à 0 ou 1, contrairement au modèle théorique. En effet, la convergence de l'algorithme dépend d'un seuil strictement positif fixé a priori et implique

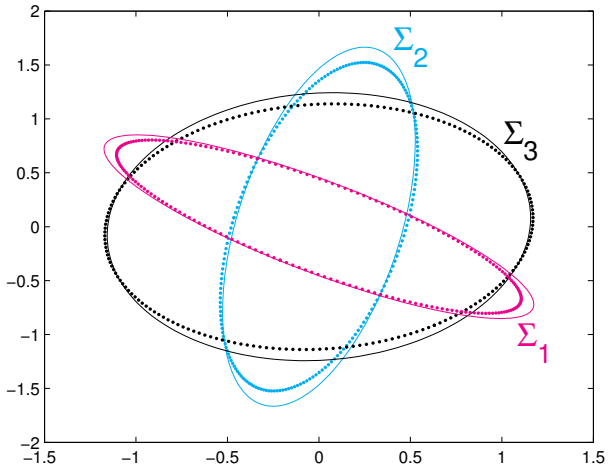

FIG. 4.3 – Matrices d'échelle estimées par EM pour l'exemple de simulation.

que la vraisemblance obtenue soit toujours strictement inférieure à la vraisemblance théorique. Les matrices d'échelle étant calculées par combinaison linéaire des éléments du cluster pondérée par les $\widehat{z}_{ik}$, leurs éléments sont donc inférieurs à ceux des matrices d'échelle théoriques. Puisque nous connaissons l'étiquetage des matrices a priori, nous comparons directement les résultats en repérant les éléments bien ou mal classés. Il y a trois matrices simulées avec Σ_2, de ddl respectifs égaux à 9, 13 et 14, qui ont été confondus avec le troisième cluster, toutes les matrices simulées avec Σ_1 ayant été bien classées. La fonction BIC a été tracée sur la figure 4.4 pour un nombre de clusters allant de 1 à 22. On vérifie que la courbe admet bien un minimum en $K = 3$, ce qui est conforme à notre choix de paramètres. Pour $K > 3$, la vraisemblance converge vers une valeur théorique correspondant au cas où $K = N$ avec une unique matrice par cluster. Le terme $-2 \log L(K)$ devient alors négligeable face au second terme qui n'est autre qu'une droite dépendante de K. En guise d'indication, la courbe $-2 \log L(K)$ a également été tracée. La vraisemblance étant croissante par rapport au nombre de clusters, on vérifie de même que $-2 \log L(K)$ est bien une fonction décroissante de K.

4.3 Analyse fractale

Nous avons déjà évoqué que l'analyse des résidus, obtenus après suppression d'une tendance, admet l'hypothèse de gaussianité. Dans la section précédente, nous avons proposé deux méthodes de classification de matrices de covariance d'échantillons de vecteurs aléatoires gaussiens issus d'un terme résiduel. La seconde partie de ce cha-

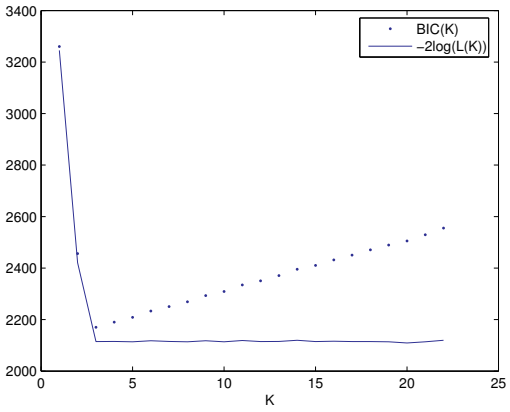

FIG. 4.4 – Tracés des courbes du critère BIC et de $-2\log(L(K))$ en fonction du nombre de clusters, pour l'exemple de simulation.

pitre s'intéresse à la nature fractale que peut présenter le résidu, après suppression de la tendance. L'hypothèse de gaussianité et la nature fractale sous-entend la notion de processus particulier appelé *mouvement brownien fractionnaire*. Nous en étudions les propriétés de base afin de proposer une méthode de segmentation de trajectoires planaires basée sur l'estimation de l'autosimilarité.

4.3.1 Mouvement brownien fractionnaire : définition, propriétés

On définit l'*autosimilarité* comme une préservation de la géométrie d'un objet, quelle que soit l'échelle à laquelle on l'observe. En réalité, il n'existe aucun objet présentant une telle similitude, au sens mathématique du terme mais on admet, par des procédés statistiques, que certains d'entre eux semblent présenter des caractéristiques se rapprochant de l'autosimilarité "parfaite". La notion d'autosimilarité est formalisée par un certain type de processus. Le mouvement brownien fractionnaire (MBF), issu des travaux de Mandelbrot et Van Ness [MN68], fait partie d'une classe de processus possédant la propriété d'autosimilarité, quantifiée par le paramètre de Hurst (H). Le MBF trouve de nombreuses applications en modélisation de signaux : biologie [Man83], analyse du mouvement [SGW03], séries économiques [CCS04], traitement d'image [LOKS86], réseau Ethernet [VV00]...

Un MBF de paramètre de Hurst $H \in]0,1[$, noté $(Z_t)_{t\in\mathbb{R}}$, est un processus gaussien centré réel vérifiant :

$$Z_0 = 0 \quad \text{et} \quad \forall (t,s) \in \mathbb{R}_+ \times \mathbb{R}_+, \quad E\left(|Z_t - Z_s|^2\right) = |t-s|^{2H} \tag{4.41}$$

On en déduit l'expression de la fonction de covariance :

$$\forall (t,s) \in \mathbb{R}_+ \times \mathbb{R}_+ , \quad E(Z_t Z_s) = \frac{c_H^2}{2} \left(|t|^{2H} + |s|^{2H} - |t-s|^{2H} \right) \tag{4.42}$$

où

$$c_H^2 = \frac{a^2 \cos(\pi H) \Gamma(1-2H)}{\pi H} \tag{4.43}$$

La constante a^2 définit une classe de processus à un facteur d'échelle près [JHJ01]. Dans la suite on supposera pour simplifier :

$$a = \sqrt{\frac{\pi H}{\Gamma(1-2H)\cos \pi H}} \tag{4.44}$$

pour avoir $c_H^2 = 1$ et donc la variance de Z_1 égale à 1. On dit que le processus $(Z_t)_{t \in \mathbb{R}}$ est *autosimilaire* d'ordre H car il vérifie la relation :

$$\forall (t,\alpha) \in \mathbb{R}_+ \times \mathbb{R}_+ , \quad Z_{\alpha t} \sim \alpha^H Z_t \tag{4.45}$$

où $\sim$ désigne l'égalité en loi. $(Z_t)_{t \geq 0}$ est l'unique processus gaussien centré continu autosimilaire à accroissements stationnaires [Her04]. Notons la dualité entre le paramètre de Hurst et la dimension fractale, qui est toujours égale presque sûrement à $2 - H$. En particulier, pour $H = 1/2$, (Z_t) est le mouvement brownien classique où la fonction de covariance se simplifie :

$$E(Z_t Z_s) = \min(t,s) \tag{4.46}$$

Les accroissement d'ordre τ de (Z_t) sont définis de la façon suivante :

$$Z_t^\tau = Z_{t+\tau} - Z_t \tag{4.47}$$

et sont appelés *bruits gaussiens fractionnaires*. Les accroissements d'un MBF sont des processus gaussiens autosimilaires, stationnaires et donc de même paramètre H :

$$\forall (t,\alpha) \in \mathbb{R}_+ \times \mathbb{R}_+ , \quad Z_{\alpha t}^\tau \sim \alpha^H Z_t^\tau \tag{4.48}$$

$$\forall t \in \mathbb{R}_+ , \quad Z_t^\tau \sim Z_0^\tau \tag{4.49}$$

Les accroissements du mouvement brownien classique sont indépendants car, pour $t > s + \tau$:

$$\begin{aligned} E(Z_t^\tau Z_s^\tau) &= E(Z_{t+\tau} Z_{s+\tau}) - E(Z_{t+\tau} Z_s) - E(Z_t Z_{s+\tau}) + E(Z_t Z_s) \\ &= (s+\tau) - s - (s+\tau) + s = 0 \end{aligned} \tag{4.50}$$

D'après (4.42), il est facile de déduire l'expression de la fonction d'autocovariance de l'accroissement :

$$\gamma(k) = E(Z_{t+k}^1 Z_t^1) = \frac{1}{2} \left(|k-1|^{2H} - 2|k|^{2H} + |k+1|^{2H} \right) \tag{4.51}$$

La fonction d'autocovariance se comporte comme $|k|^{2H-2}$ quand k tend vers l'infini [Ber94] :

$$\gamma(k) \sim H(2H-1)|k|^{2H-2} \quad (k \to +\infty) \tag{4.52}$$

Si $H > 1/2$, une propriété des séries entières affirme que le terme de droite n'est pas sommable. Dans ce cas, le processus est dit à *longue-mémoire* (ou $1/f$), et à *courte-mémoire* si $H \leq 1/2$.

Représentations du MBF

Soit $k(.,.)$ le noyau d'intégration fractionnaire de $\mathcal{L}^2(\mathbb{R})$ défini de la manière suivante :

$$\forall (t,s) \in \mathbb{R}_+ \times \mathbb{R}_+ , \quad k(t,s) = |t-s|^{H-1/2} - |t|^{H-1/2} \tag{4.53}$$

Le MBF a une représentation dite *moyenne mobile* [ST94]. Chacune des variables du processus Z_t pour t fixé s'écrit comme une somme de gaussiennes centrées réduites pondérée par les coefficients du noyau $k(.,.)$:

$$\forall t \in \mathbb{R}_+ , \quad Z_t = \int_0^{+\infty} k(t,s) W(ds) \sim \mathcal{N}\left(0, \|k(t,.)\|^2\right) \tag{4.54}$$

où W est une mesure brownienne, *i.e.* une isométrie de $\mathcal{L}^2(\mathbb{R})$ vers un espace gaussien. Le processus $(W(ds))_{s\in\mathbb{R}}$ est également un bruit blanc gaussien. À l'aide de la transformée de Fourier du noyau k :

$$\widehat{k}(t,s) = \frac{e^{its}-1}{|s|^{H+1/2}} \tag{4.55}$$

on obtient la représentation *harmonisable* du processus :

$$\forall t \in \mathbb{R}_+ , \quad Z_t = \int_0^{+\infty} \frac{e^{its}-1}{|s|^{H+1/2}} \widehat{W}(ds) \sim \mathcal{N}\left(0, \|\widehat{k}(t,.)\|^2\right) \tag{4.56}$$

où $(\widehat{W}(ds))_{s\in\mathbb{R}_+}$ est la transformée de Fourier du bruit blanc $(W(ds))$.

4.3.2 Simulation de trajectoires échantillonnées irrégulièrement

Il existe de nombreuses méthodes de simulation de trajectoires browniennes fractionnaires : approches stochastiques, multi-échelle, spectrales, par la matrice de covariance du processus [AS96, Coe00, DM03, JHJ01, Sel95, WC94] ... La plupart de ces méthodes supposent que la période d'échantillonnage soit constante. Nous avons choisi de simuler des trajectoires à partir de la matrice de covariance du processus en utilisant la méthode

de Cholesky, qui prend en compte l'irrégularité d'échantillonnage via l'équation (4.42). Les méthodes à base d'ondelettes, celles de type Paxson ou de simulations spectrales, fournissant de bonnes simulations, peuvent être utilisées. Une manière simple serait de simuler une trajectoire régulière puis de la ré-échantillonner avec un pas non constant.

L'algorithme de Cholesky permet de considérer l'échantillonnage de manière intrinsèque à l'aide de la covariance. De la même manière qu'avec la méthode de simulation de matrices de Wishart, une trajectoire de MBF échantillonnée irrégulièrement est obtenue en considérant la décomposition de Cholesky de la matrice de covariance du processus. Soient $(t_2, ..., t_N)$ les instants d'échantillonnage d'une trajectoire de longueur $N-1$ et H son paramètre de Hurst. On calcule la matrice de covariance associée via (4.42) :

$$V_H(i,j) = E(Z_{t_i} Z_{t_j}) = \frac{1}{2}\left(|t_i|^{2H} + |t_j|^{2H} - |t_i - t_j|^{2H}\right) \ , \quad 2 \leq i,j \leq N \tag{4.57}$$

La matrice V_H est symétrique définie positive et peut donc être décomposée par la méthode de Cholesky :

$$V_H = T\,{}^t\!T \tag{4.58}$$

où T est la matrice triangulaire inférieure dans la décomposition de Cholesky. Si Y est un échantillon de longueur $N-1$ d'une variable gaussienne centrée réduite, alors l'ensemble $(0, TY)$ est une trajectoire de longueur N d'un MBF aux instants $(t_1, ..., t_N)$ avec $t_1 = 0$ et d'autosimilarité H.

La méthode de Cholesky est dite *exacte* car la matrice de covariance calculée dans le cas discret coïncide avec le cas continu. En général, on l'applique à la matrice de covariance, notée V, du bruit gaussien fractionnaire car, pour gagner en temps de calcul, une alternative est d'opter pour l'algorithme de Levinson [Pel98] qui permet de reconstruire V dans la décomposition de Cholesky à partir des éléments de la première ligne. Dans le cas d'un échantillonnage irrégulier, cet algorithme n'est plus applicable puisque V n'a plus nécessairement la propriété Toeplitz. L'avantage de la simulation par Cholesky est la prise en compte de manière intrinsèque de n'importe quel type d'échantillonnage, par l'intermédiaire de la covariance. L'irrégularité des échantillons simulés est obtenue en contrôlant les valeurs du paramètre temporel t qui constitue la base du calcul de la matrice de covariance. La particularité des MBF est que :

La matrice de covariance du processus est entièrement déterminée par les t_i et H

L'inconvénient de la méthode de Cholesky est son coût élevé en temps de calcul puisque la complexité est en $\mathcal{O}(N^3)$ [Coe00], et impose donc de simuler des trajectoires de longueur limitée. La figure 4.5 présente une série de trajectoires browniennes fractionnaires

simulées par la méthode de Cholesky pour H=0.01, 0.25, 0.5, 0.7, 0.9 et 0.99. Parallèlement, la figure 4.6 présente les matrices de covariance à valeurs normalisées entre 0 et 1. Quand H se rapproche de 0, la matrice de covariance du processus tend continûment vers la matrice constante égale à 1/2. Par conséquent, le signal est de plus en plus erratique puisqu'il s'apparente à un échantillon de la variable gaussienne Y de la simulation. Quand H se rapproche de 1, V_H tend vers la matrice $V_1(t_i, t_j) = t_i t_j$. Par conséquent, la matrice T a sa première colonne contenant les dates t_i, et 0 partout ailleurs. Avec le produit TY, la trajectoire simulée tend vers une droite affine passant par l'origine et ayant pour équation tY_{t_1}. Enfin, la formule (4.42) montre que la variance d'un MBF croît avec le temps :

$$var(Z_t) = E(Z_t Z_t) = |t|^{2H} \tag{4.59}$$

Cette propriété est illustrée par la figure 4.7 avec une simulation de six trajectoires pour $H = 0.4$. La parabole en trait épais correspond au tracé de la courbe de l'écart-type $\pm|t|^H$. Du fait de la variance croissante du processus, les réalisations des variables aléatoires formant les points d'une trajectoire sont de moins en moins contraintes à être voisines de l'origine.

4.3.3 Extensions multidimensionnelles du MBF

Jusqu'à présent, nous nous sommes restreints au cas classique unidimensionnel où H était compris entre 0 et 1 et constant le long des trajectoires du processus. Perrin *et al.* [PHBJ+01] ont proposé une généralisation du MBF à l'ordre supérieur, dans laquelle le paramètre de Hurst peut être supérieur à 1. Également, les extensions multidimensionnelles du MBF ont fait l'objet de nombreux travaux théoriques. Parmi les différentes variantes, le *mouvement brownien multifractionnaire* (MBMF) [PLV95] a été introduit pour pallier à la propriété d'autosimilarité constante d'un MBF, ce qui constitue une forte limitation dans la modélisation de phénomènes aléatoires. Le MBMF est une généralisation du MBF où le paramètre H est remplacé par une fonction du temps $H(t)$ définie sur $\mathbb{R}_+$ et à valeurs dans $]0, 1[$. Sa fonction de covariance est la suivante [ACLV00] :

$$E(Z_t Z_s) = \left(\int_0^{+\infty} \frac{1 - e^{iu}}{|u|^{H(t)+H(s)+1}} du \right) \left(|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |t-s|^{H(t)+H(s)} \right) \tag{4.60}$$

Le paramètre H est dépendant du temps t, ce qui implique que le processus est donc localement autosimilaire, impliquant la non-stationnarité des accroissements. La figure 4.8 en donne un exemple avec une trajectoire de MBMF pour H variant continûment de 0.2 à 0.5. La fonction correspondante $H(t) = arctan(t)$ est présentée sur le graphique de droite. Les autres extensions multidimensionnelles font référence au cas où le paramètre

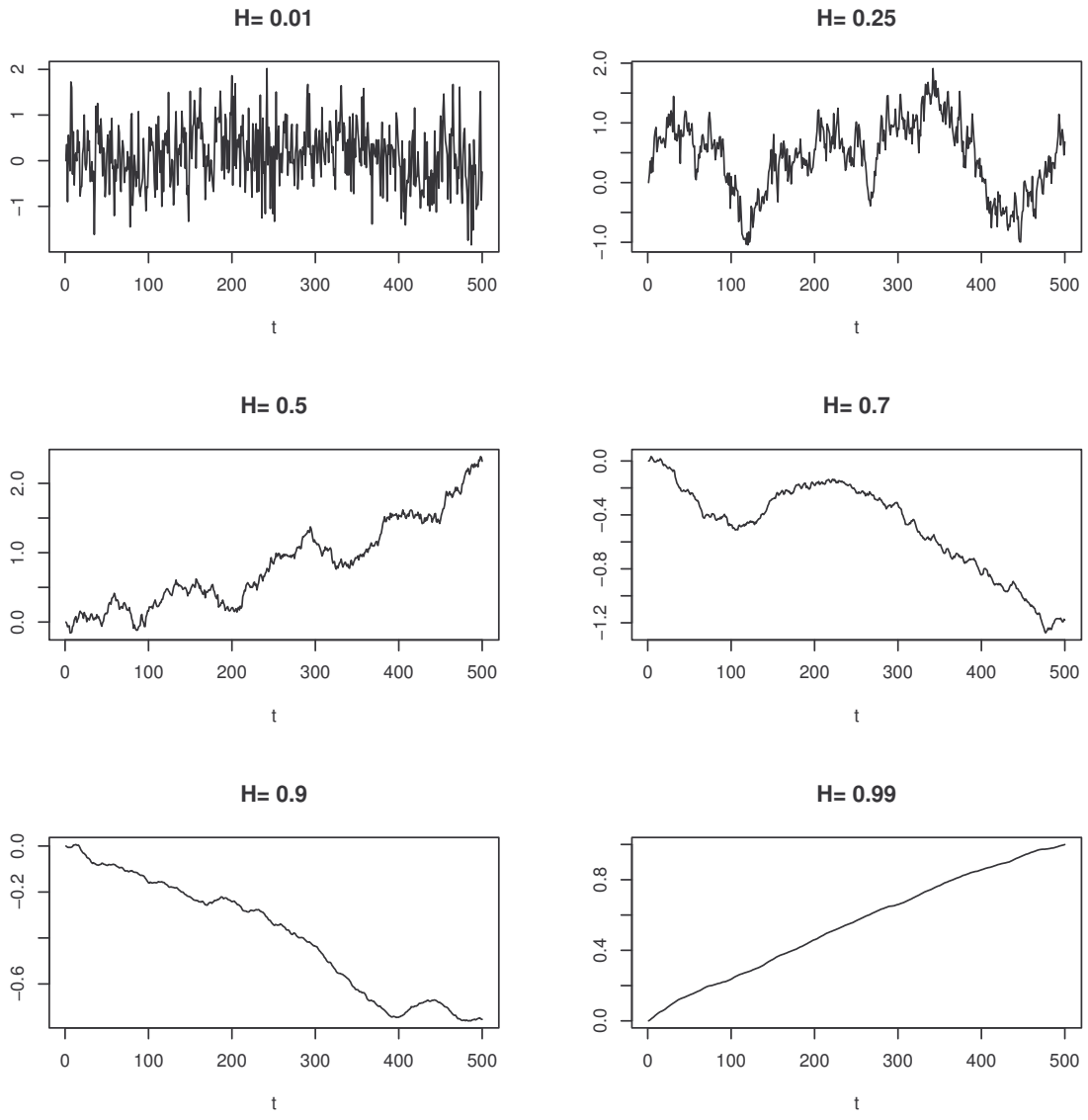

FIG. 4.5 – Trajectoires browniennes fractionnaires avec $N = 500$ simulées par la méthode de Cholesky pour H=0.01, 0.25, 0.5, 0.7, 0.9 et 0.99.

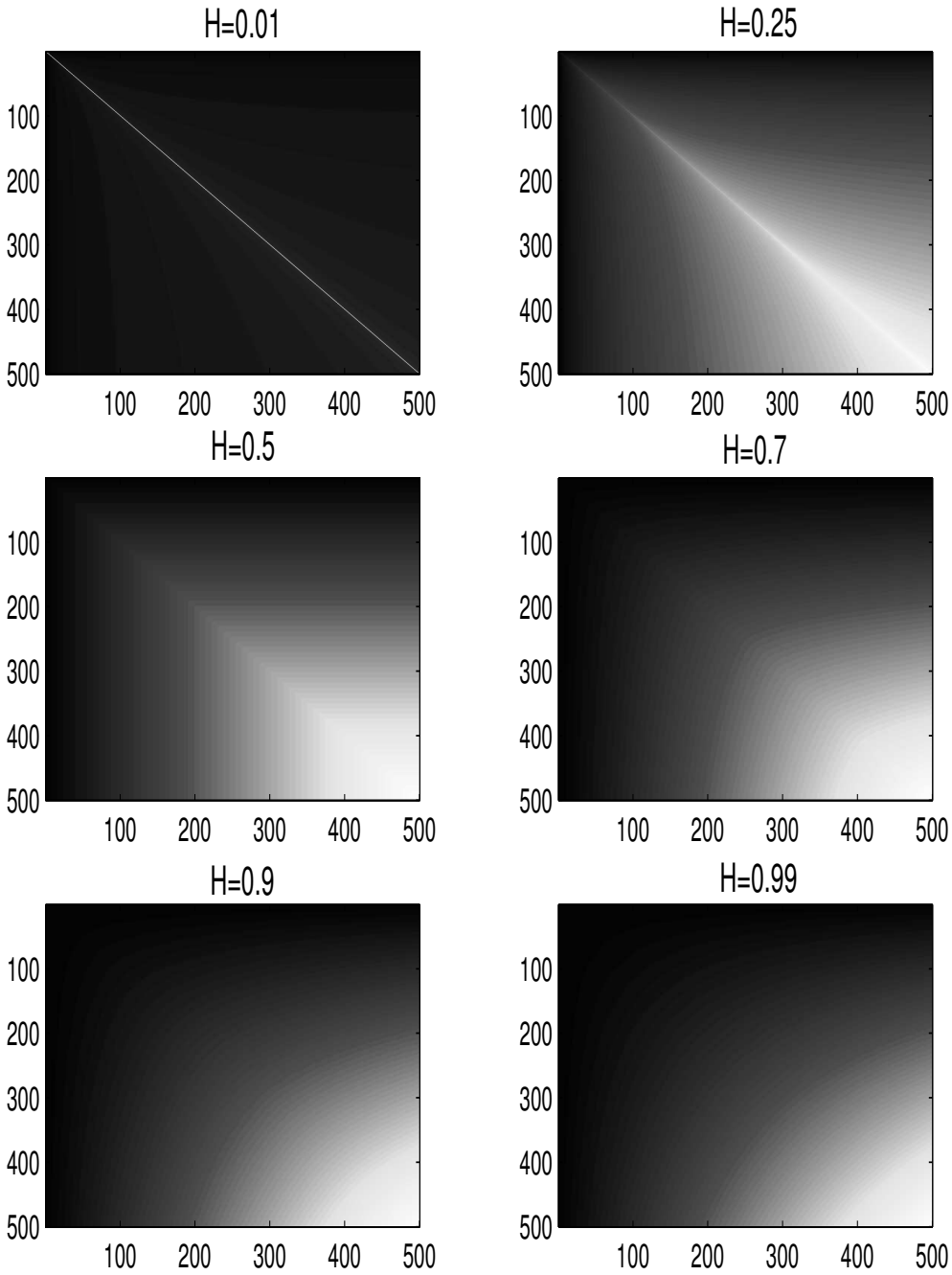

FIG. 4.6 – Matrice de covariance avec $N = 500$ pour H=0.01 ($\simeq$ bruit blanc), 0.25, 0.5 (mouvement brownien), 0.7, 0.9 et 0.99. Les valeurs sont comprises entre 0 (couleur noire) et 1 (couleur blanche).

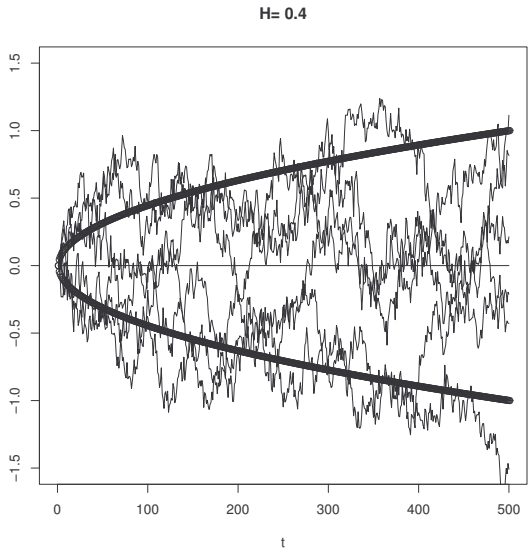

FIG. 4.7 – Simulation de 6 trajectoires avec H=0.4. La variance des variables aléatoires échantillonnées augmentent avec le temps.

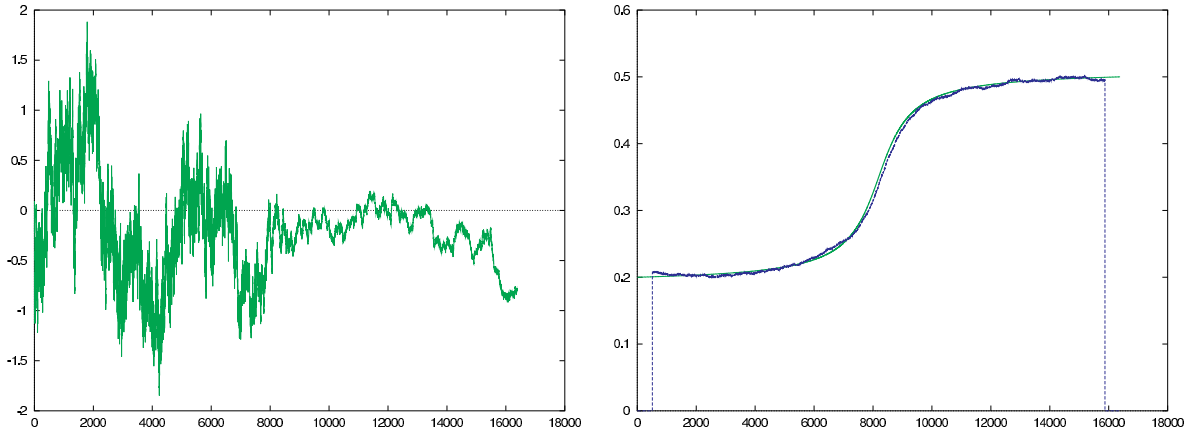

FIG. 4.8 – Trajectoire d'un processus brownien multifractionnaire avec sa fonction associée $H(t) = arctan(t)$ (source [PLV95]).

temporel est un multi-indice, *i.e.* $t \in \mathbb{R}^p$ avec $p > 1$. Le MBF en possède deux extensions classiques. Le *champ brownien fractionnaire* (ou mouvement brownien fractionnaire de Lévy) [Kah85] généralise de façon naturelle le MBF. C'est un processus gaussien centré de fonction de covariance :

$$E(Z_t Z_s) = \frac{1}{2}(\|t\|^{2H} + \|s\|^{2H} - \|t-s\|^{2H}) \tag{4.61}$$

où $0 < H < 1$ et $\|.\|$ désigne la norme euclidienne usuelle sur $\mathbb{R}^p$. Le champ brownien fractionnaire est autosimilaire de paramètre de Hurst H :

$$Z_{\alpha t} \sim \alpha^H Z_t \tag{4.62}$$

Le *drap brownien fractionnaire* [Kam96] est également un processus gaussien centré qui consiste à donner un paramètre de Hurst à chacune des dimensions, c'est à dire que l'on dispose d'un vecteur $H = (H_1, ..., H_p)$ pour chaque composante de l'indice temporel t de type vectoriel. La fonction de covariance se formule comme un produit de covariances en fonction des $t = (t_i)_{1 \leq i \leq p}$ et $s = (s_i)_{1 \leq i \leq p}$:

$$E(Z_t Z_s) = \prod_{i=1}^{p} (|t_i|^{2H_i} + |s_i|^{2H_i} - |t_i - s_i|^{2H_i}) \tag{4.63}$$

Le drap brownien fractionnaire est quant à lui autosimilaire d'indice $H = \sum_{i=1}^{p} H_i$.

$$Z_{\alpha t} \sim \alpha^{\left(\sum_{i=1}^{p} H_i\right)} Z_t \tag{4.64}$$

La notion d'accroissement de processus est plus délicate à définir pour un indice de temps multidimensionnel et cette notion a été étendue dans un cadre plus général, ce qui permet de donner un sens à la stationnarité. Pour plus de détails, le lecteur intéressé trouvera une étude approfondie dans la thèse de Herbin [Her04].

4.3.4 Grande dimension et échantillonnage

Contrainte de dimensionnalité

Les domaines d'application des extensions multidimensionnelles du MBF sont relativement récents et peu nombreux. On les emploie notamment dans l'analyse d'images texturées [LOKS86, PPLV02] car la représentation matricielle des images permet de considérer le paramètre temporel comme un multi-indice à deux coordonnées (en ligne et en colonne). Mais, d'une manière générale, les analyses fractales appliquées notamment à l'étude du mouvement ne traitent que du cas unidimensionnel. Le plus souvent, on dispose d'une courbe paramétrée par l'indice de temps (réel), qui fournit à t fixé une

observation identifiée à un vecteur de $\mathbb{R}^p$. Dans ce cas, l'extension multidimensionnelle ne fait plus référence à l'indice temporel mais à la dimension dans laquelle sont plongés les points de la trajectoire. Considérons par exemple le cas bidimensionnel ($p = 2$). La figure 4.9 donne un exemple de six trajectoires bidimensionnelles simulées par Cholesky pour les mêmes H que précédemment. Supposons maintenant que les coefficients d'autosimilarités en x et en y soient différents. Est-il légitime de supposer une autosimilarité conjointe donnée par les trajectoires en x et y d'autosimilarité différente ? Une idée serait de considérer l'écriture complexe $Z_t = X_t + iY_t$. Le processus (Z_t) est alors gaussien complexe [AHSE95], où X_t et Y_t sont des variables mutuellement indépendantes, qui impose le même paramètre de Hurst pour les processus (X_t) et (Y_t) :

$$Z_{\alpha t} \sim \alpha^H Z_t \Longleftrightarrow (X_{\alpha t} + iY_{\alpha t}) \sim \alpha^H (X_t + iY_t) \tag{4.65}$$

Comme on le verra dans la suite, il est possible d'estimer H par maximum de vraisemblance que l'on aura préalablement adapté aux trajectoires à réalisations complexes [Bri75]. Malgré un ensemble de travaux théoriques sur le sujet [BN06, PY86, Spi58], cette considération n'a malheureusement que peu d'intérêt pratique dans l'estimation de l'autosimilarité, car il est clair qu'il suffirait simplement d'étudier indépendamment le processus (X_t) ou (Y_t). Ainsi, afin d'analyser une trajectoire multidimensionnelle échantillonnée par l'indice réel t, nous n'avons d'autre choix que d'estimer H de manière marginale, coordonnée par coordonnée.

Contrainte liée à l'échantillonnage

En pratique, il est difficile d'affirmer qu'un signal possède un caractère autosimilaire et certains tests sont proposés pour vérifier la bonne adéquation modèle/donnée [JHJ96, Bar99] : par exemple, *run test* pour la stationnarité et test du χ^2 pour la similarité des accroissements. Cependant, ces tests ne s'appliquent pas dans le cas de signaux obtenus par échantillonnage irrégulier. La modélisation par MBF pour ce type de données peut alors devenir inadéquate. La plupart des méthodes d'identification et d'estimation [Coe00, JHJ01, IL94, PLV94, GCL03] de H d'un MBF supposent également un échantillonnage régulier. Ces méthodes nécessitent, dans le cas contraire, une modification en profondeur des algorithmes. L'échantillonnage irrégulier des séries temporelles a donné lieu à beaucoup de travaux, et notamment en analyse spectrale [PST00] et nous proposons une vérification expérimentale des propriétés de l'EMV en estimant le paramètre de Hurst sur des trajectoires échantillonnées irrégulièrement [HSJL]. Également, nous analysons le comportement de l'estimateur pour des signaux de taille faible.

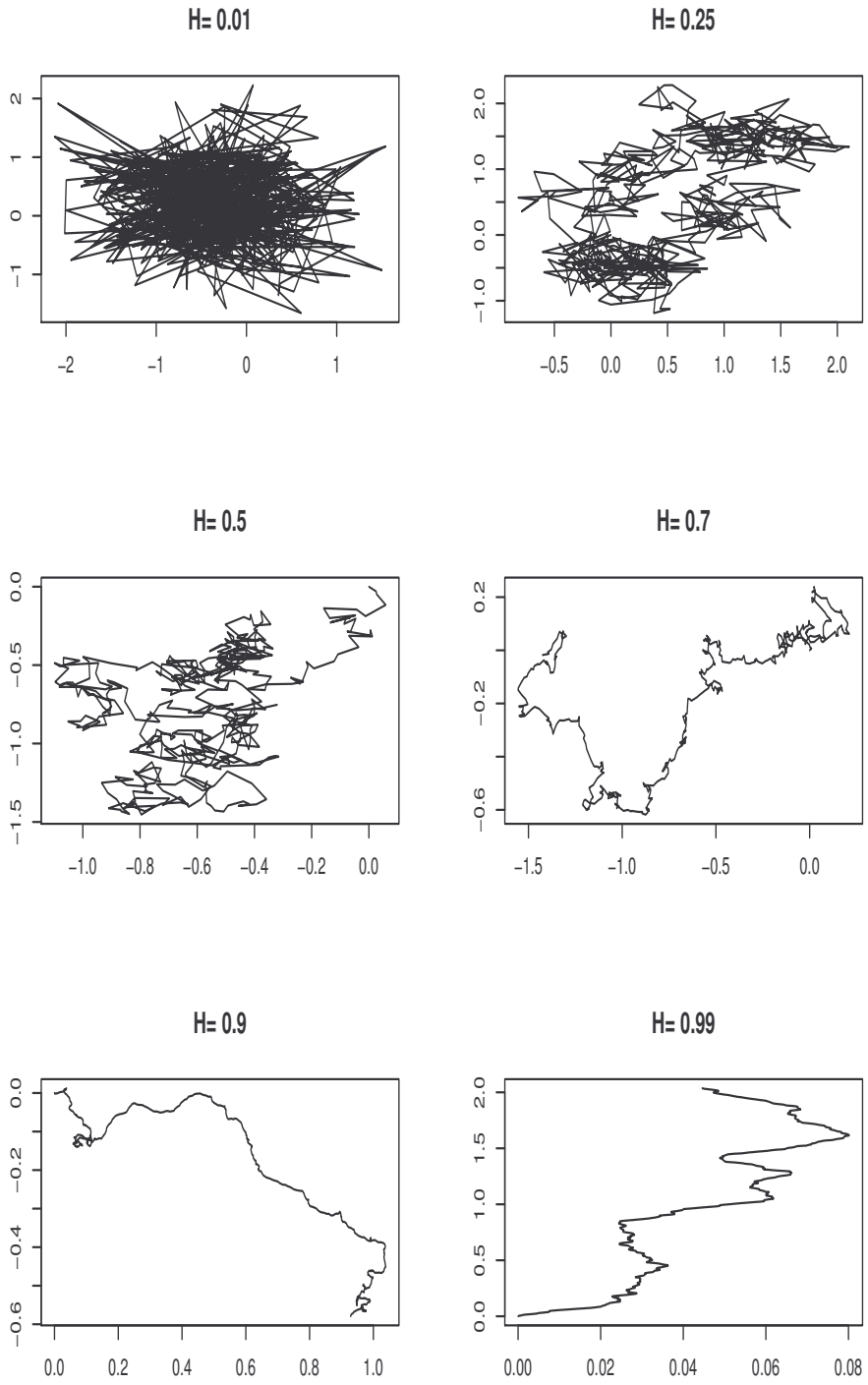

FIG. 4.9 – Trajectoires browniennes fractionnaires bidimensionnelles de longueur 500 simulées par la méthode de Cholesky pour H=0.01, 0.25, 0.5, 0.7, 0.9 et 0.99.

4.3.5 Estimation de l'autosimilarité avec un échantillonnage irrégulier

L'estimateur du maximum de vraisemblance (EMV) a été adapté par Dahlhaus [Dah89] pour le calcul du coefficient d'autosimilarité d'un MBF. Le choix de l'EMV est motivé d'une part par le fait qu'il bénéficie de bonnes propriétés asymptotiques [Ber94] : convergence et efficacité de l'estimateur, et d'autre part parce qu'il ne nécessite pas de connaître a priori la densité spectrale de puissance (DSP) du signal, contrairement à d'autres estimateurs plus rapides et asymptotiquement efficaces comme l'approximation de Whittle [FT86]. Le principal inconvénient de la méthode est le temps de calcul extrêmement élevé puisqu'il faut calculer l'inverse et le déterminant de la matrice de covariance à chaque itération.

Si l'on choisit d'estimer la DSP, la méthode la plus simple est celle du périodogramme. Pour un échantillonnage arbitraire, Scargle [Sca82] en propose une version généralisée, qui possède les mêmes propriétés statistiques (non biais asymptotique) que le périodogramme classique. Cette procédure nécessite néanmoins de recourir à des techniques de lissage local [Yue78] pour diminuer sa variance intrinsèquement élevée [BD91]. Enfin, l'approximation de Whittle impose de connaître le périodogramme d'une trajectoire de l'accroissement, qui n'est pas directement observable dans le cas d'un échantillonnage irrégulier.

Estimateur du maximum de vraisemblance pour le paramètre de Hurst

Soit $\mathbf{x} = {}^t(x_1, ..., x_N)$ une trajectoire du processus (X_t) aux instants d'échantillonnage $\mathbf{t} = {}^t(t_1, ..., t_N)$ et d'autosimilarité supposée constante. On cherche à maximiser par rapport à (m, a, H) une probabilité de vraisemblance donnée par la fonction de densité mutinormale [VV00] :

$$L(\mathbf{x}; m, a, H) = (2\pi)^{-\frac{N}{2}} (a^N |V_H|)^{-\frac{1}{2}} e^{-\frac{1}{2a}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t})} \tag{4.66}$$

où $|.|$ désigne le déterminant. Le logarithme de la vraisemblance est donné par :

$$\begin{aligned} \log L(\mathbf{x}; m, a, H) &= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\log\left(a^N |V_H|\right) - \frac{1}{2a}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t}) \quad (4.67) \\ &= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(a) - \frac{1}{2}\log(|V_H|) - \frac{1}{2a}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t}) \end{aligned}$$

Pour obtenir l'estimé de m, on dérive l'expression par rapport à m avec la condition de dérivation nulle :

$$\begin{aligned} \frac{\partial}{\partial m}\log L(\mathbf{x};m,a,H) &= -\frac{1}{2a}\frac{\partial}{\partial m}\left({}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t})\right) \\ &= \frac{1}{a}\left({}^t\mathbf{t}V_H^{-1}\mathbf{x} - m{}^t\mathbf{t}V_H^{-1}\mathbf{t}\right) = 0 \end{aligned} \tag{4.68}$$

Soit :

$$\widehat{m} = \frac{{}^t\mathbf{t}V_H^{-1}\mathbf{x}}{{}^t\mathbf{t}V_H^{-1}\mathbf{t}} \tag{4.69}$$

De même, l'estimé $\widehat{a}$ est obtenu en dérivant le logarithme de la vraisemblance par rapport à a. On obtient :

$$\frac{\partial}{\partial a}\log L(\mathbf{x};m,a,H) = -\frac{1}{2}\left(\frac{N}{a} - \frac{1}{a^2}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t})\right) = 0 \tag{4.70}$$

Donc :

$$\widehat{a} = \frac{1}{N}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t}) \tag{4.71}$$

Le problème consiste dorénavant à estimer H. Aux constantes près, cela se résume à maximiser relativement à H la fonction objectif suivante :

$$-N\log\left(\frac{1}{N}{}^t(\mathbf{x}-m\mathbf{t})V_H^{-1}(\mathbf{x}-m\mathbf{t})\right) - \log|V_H| \tag{4.72}$$

Pour maximiser la fonction objectif, nous appliquons une méthode quasi-Newton. En toute rigueur, il faudrait démontrer la concavité de (4.72) considérée comme une fonction de H. Il s'agit d'une question ardue et nous nous limiterons à une vérification expérimentale.

Dans le cas d'un échantillonnage quelconque, Lai [Lai04] montre que l'estimateur $\widehat{H}$ du paramètre de Hurst d'un MBF est asymptotiquement gaussien, sans biais et d'écart-type :

$$\sigma(\widehat{H}_{EMV}) = \frac{2}{trace\left([V_H^{-1}V_H']^2\right)} \tag{4.73}$$

où V_H' est la matrice dérivée de V_H :

$$V_H'(i,j) = \frac{\partial V_H(i,j)}{\partial H} \tag{4.74}$$

4.3.6 Étude expérimentale de la précision et du biais de l'EMV

Nous proposons d'étudier l'EMV du paramètre de Hurst pour des trajectoires browniennes fractionnaires échantillonnées irrégulièrement. Le but étant d'étudier le comportement de l'estimateur pour de signaux de longueur faible et de vérifier le maintien

	Signal 1 ($reg.$)	Signal 2(Irr_U)
Whittle	0.565	0.531
EMV	0.562	0.570

TAB. 4.3 – Exemple d'estimation de H par Whittle avec irrégularité d'échantillonnage.

de ses propriétés asymptotiques. La vérification expérimentale sera effectuée sans la tendance à estimer ($m = 0$), qui n'est pas à prendre en compte en raison de la méthode de simulation. Nous allons considérer deux types d'irrégularité temporelle. Dans les deux cas, nous procédons par sous-échantillonnage dans un échantillon régulier de référence : $\{t_i\}_{i=1}^N$ avec $t_i - t_{i-1} = 1$. Le premier sous-échantillonnage noté Irr_S est systématique, et considère alternativement des instants consécutifs et des instants séparés de deux unités de temps, à savoir : $\ldots, t_i, t_{i+1}, t_{i+3}, t_{i+4}, \ldots$. Le second sous-échantillonnage, noté Irr_U, est aléatoire et s'obtient par tirage uniforme dans l'échantillon de référence.

Exemple simple

Nous commençons par un exemple simple montrant les divergences des résultats suivant l'échantillonnage. Soient deux signaux simulés de taille $N = 1024$ avec $H = 0.6$ par Cholesky avec respectivement une régularité d'échantillonnage ($t_i = i$) et une irrégularité issue de la loi uniforme (Irr_U). Ces deux signaux ont été obtenus à partir du même échantillon gaussien. Nous estimons par l'approximation de Whittle (échantillonnage supposé régulier) et par EMV les coefficients d'autosimilarité des deux signaux. Les résultats sont présentés au tableau 4.3. On constate sur cet exemple que la non-robustesse de l'estimateur de Whittle sur le type d'échantillonnage puisque celui-ci sous-estime la valeur théorique de H. L'estimateur proposé donne une meilleure estimation (proche de 0.6) avec une différence moins significative. Par conséquent, l'irrégularité d'échantillonnage semble jouer un rôle important, ce qui nous amène à étudier son influence dans l'estimation de H. Cet exemple illustratif ne permet pas de mesurer la réelle efficacité de l'EMV et il convient donc, pour une certaine longueur N, de procéder à une étude systématique effectuée en répétant les procédures expérimentales de simulation et d'estimation, de manière à étudier la distribution d'échantillonnage de l'estimateur de H, son biais et sa précision.

N	32	64	128	256	512	1024
R	3053	1492	750	383	187	96

TAB. 4.4 – Calcul du nombre minimum de simulations en fonction de la taille du signal.

Choix des paramètres, stratégie de tests

Pour mener l'analyse, nous nous référons à l'article de Jennane *et al.* [JHJ01]. La démarche est la suivante. Ayant fixé un mode d'échantillonnage, nous étudions des signaux de différentes tailles (soit N) pour différentes valeurs de l'exposant de Hurst (H). Pour un mode d'échantillonnage et pour N et H fixés, nous répétons R fois les opérations de simulation d'un MBF (sous les contraintes précédentes) et d'estimation de H. On obtient ainsi un échantillon de taille suffisante pour approcher la distribution de l'estimateur $\widehat{H}$ du maximum de vraisemblance. Nous étudions alors expérimentalement les propriétés de biais et de convergence de l'estimateur en question. Le choix de R est fondé sur le contrôle des risques et de la puissance des tests de comparaison de moyennes et de variances que nous allons mener. Conformément à l'approche de Jennane *et al.*, nous adoptons pour le calcul de R l'hypothèse selon laquelle la variance de l'estimateur serait égale à la borne de Cramer-Rao correspondante, dont leur article donne la valeur. Cette borne dépend de H, et l'on adopte une stratégie "pessimiste" en choisissant la valeur maximum tout H confondus pour un N fixé. En supposant que la densité de l'estimateur soit gaussienne et avec un niveau de confiance de 1%, on trouve le tableau 4.4. Dans ce qui suit, nous comparons les résultats obtenus pour chaque irrégularité d'échantillonnage avec le cas classique d'estimation (régularité) et nous étudions l'influence de l'irrégularité temporelle sur le biais de l'estimateur. Toute la démarche est résumée en figure 4.10.

Afin de valider les observations, nous avons procédé à des tests de comparaison d'échantillons en nous intéressant à la précision et à la présence éventuelle d'un biais et ce, pour les deux irrégularités. Dans tous les cas, nous considérons un risque de première espèce α égal à 0.01. Nous appliquons successivement ("$Fish + Stud_1$", fig. 4.10), dans le cas gaussien, le test de Fisher de comparaison des variances puis le test de Student de comparaison des moyennes [Sap90]. L'hypothèse nulle H_0 du test de Fisher est l'homogénéité des variances des populations dont sont issus les deux échantillons, l'alternative étant l'hétérogénéité des variances. Puisque ces derniers ont la même taille, la statistique de Fisher à $R-1$ ddl est simplement le rapport entre s_1^2 et s_2^2, estimateurs empiriques des variances :

$$F(R-1, R-1) = \frac{s_1^2}{s_2^2} \quad \text{avec} \quad s_1^2 > s_2^2 \tag{4.75}$$

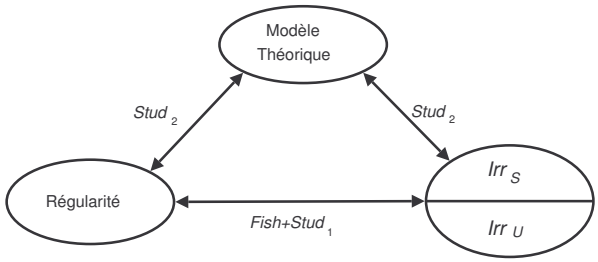

FIG. 4.10 – Stratégie de tests pour la comparaison et l'étude du biais de l'estimateur de H par EMV - $Stud_2$: Test de Student (Intervalle de confiance) ; $Fish + Stud_1$: Test de Fisher + Test de Student (Comparaison d'échantillons).

Pour une irrégularité donnée, ce test est appliqué pour chacune des cases homologues des deux tableaux. Pour les cas où il n'y a pas rejet de H_0, nous appliquons le test de Student où l'hypothèse nulle est l'égalité des moyennes $E(H_1)$ et $E(H_2)$. La statistique de test est la suivante :

$$T_{2R-2} = \sqrt{\frac{R-1}{s_1^2 + s_2^2}} \left(\overline{H}_1 - \overline{H}_2\right) \tag{4.76}$$

où $\overline{H}_1$ et $\overline{H}_2$ sont les moyennes empiriques de l'estimateur. Si l'hypothèse nulle des deux tests n'est pas rejetée, on conclut que l'irrégularité de l'échantillonnage n'a pas d'influence sur les propriétés de l'estimateur.

Pour juger du biais de l'estimateur, nous calculons l'intervalle de confiance (IC) au seuil α ("$Stud_2$" fig. 4.10) et vérifions qu'il contient la valeur théorique H du paramètre de Hurst. Une estimation de H étant indépendante des autres, on applique le théorème central limite sur $\overline{H}$ et on en déduit que $\overline{H}$ suit approximativement une loi gaussienne. Puisque l'écart-type est inconnu, la statistique $\frac{\overline{H}-H}{s/\sqrt{R-1}}$ de l'échantillon de moyenne $\overline{H}$ et d'écart-type empirique s suit approximativement une loi de Student T_{R-1} à $R-1$ ddl, ce qui donne l'encadrement :

$$-t_{\alpha/2} < \frac{\overline{H} - H}{s/\sqrt{R-1}} < t_{\alpha/2} \tag{4.77}$$

où $t_{\alpha/2}$ est $(1-\alpha)/2$ percentile de T_{R-1}. On en déduit l'IC :

$$\overline{H} - t_{\alpha/2}\frac{s}{\sqrt{R-1}} < H < \overline{H} + t_{\alpha/2}\frac{s}{\sqrt{R-1}} \tag{4.78}$$

De cette façon, on obtient l'IC qui permet de vérifier l'appartenance ou non de la valeur théorique choisie pour la simulation. Le calcul de l'IC est analogue à tester l'hypothèse nulle $E(\widehat{H}) = H$ avec la même statistique $\frac{\overline{H}-H}{s/\sqrt{R-1}}$.

Comparaison d'échantillons ($Fish + Stud_1$)

Le tableau 4.5 contient les moyennes et les écarts-types obtenus pour un échantillonnage régulier et croise les valeurs de H avec la taille des signaux. Les valeurs des tableaux sont arrondies mais les tests sont menés sur les valeurs exactes. On retrouve les propriétés usuelles de la méthode d'estimation : l'écart-type de l'estimateur diminue quand N grandit, illustrant ainsi le fait que l'EMV est convergent. Globalement, les moyennes obtenues sous-estiment les valeurs de H.

Le tableau 4.6 présente les mêmes estimations en considérant le cas Irr_S. Tous les écarts-types obtenus sont inférieurs ou égaux à ceux du tableau 4.5. L'irrégularité Irr_S réduit donc l'écart-type de l'estimateur et les valeurs théoriques de H sont sous-estimées. Nous avons comparé ces résultats avec ceux obtenus dans le tableau 4.5 à l'aide des tests statistiques expliqués plus haut. Les cellules en gris clair (resp. foncé) correspondent au non rejet de H_0 pour le test de Fisher (resp. Fisher et Student). Les non-rejets sont comparables pour les deux tableaux à partir de la taille $N = 512$. Pour la taille maximale 1024, les résultats obtenus dans les tableaux 4.5 et 4.6 sont similaires.

Le tableau 4.7 contient les estimés avec l'irrégularité Irr_U. Les résultats obtenus sont nettement meilleurs qu'avec les deux cas précédents en terme de variance. Tous les écarts-types sont inférieurs ou égaux à ceux des deux autres tableaux et la différence est plus forte globalement pour des valeurs de H autour de 0.5. Les non-rejets sont comparables pour les tableaux 4.5 et 4.7 à partir d'une taille comprise entre 512 et 1024. Seuls les moyennes et écarts-types des échantillons pour la taille 1024 se rapprochent des résultats obtenus avec un échantillonnage régulier. On constate que pour une irrégularité périodique (Irr_S), les valeurs des estimés sont plus proches du cas régulier, les tests de comparaison d'échantillons étant beaucoup plus rejetés dans le tableau 4.7.

En conclusion, plus l'échantillonnage est erratique, plus la convergence des estimateurs est lente et cette particularité tend à diminuer avec la taille des signaux. L'échantillonnage n'influence significativement les estimations que pour des signaux de longueur faible.

H\N	32	64	128	256	512	1024
0.1	0.113±0.068	0.100±0.047	0.098±0.034	0.099±0.024	0.098±0.016	0.098±0.011
0.2	0.202±0.093	0.199±0.064	0.200±0.043	0.201±0.030	0.199±0.021	0.198±0.014
0.3	0.291±0.107	0.295±0.073	0.298±0.050	0.299±0.034	0.297±0.025	0.298±0.018
0.4	0.386±0.115	0.391±0.080	0.394±0.056	0.397±0.038	0.401±0.025	0.401±0.017
0.5	0.485±0.117	0.494±0.081	0.497±0.058	0.499±0.039	0.498±0.029	0.498±0.021
0.6	0.581±0.119	0.592±0.082	0.597±0.055	0.599±0.041	0.599±0.030	0.601±0.020
0.7	0.676±0.115	0.685±0.081	0.691±0.056	0.694±0.040	0.699±0.031	0.701±0.019
0.8	0.769±0.110	0.788±0.078	0.795±0.058	0.798±0.043	0.801±0.030	0.803±0.021
0.9	0.858±0.092	0.875±0.068	0.887±0.050	0.892±0.038	0.899±0.028	0.900±0.019

TAB. 4.5 – Estimations par EMV avec régularité d'échantillonnage.

H\N	32	64	128	256	512	1024
0.1	0.112± 0.066	0.103 ± 0.045	0.102±0.032	0.102±0.022	0.102±0.015	0.102±0.011
0.2	0.197±0.086	0.195 ± 0.059	0.196 ± 0.040	0.197±0.027	0.197±0.019	0.198±0.014
0.3	0.293 ± 0.098	0.296±0.069	0.299±0.047	0.300±0.033	0.300±0.024	0.300±0.017
0.4	0.390 ± 0.105	0.394 ± 0.071	0.397±0.048	0.398±0.033	0.398±0.023	0.398±0.016
0.5	0.483 ± 0.110	0.490 ± 0.073	0.493 ± 0.054	0.495 ± 0.038	0.496±0.024	0.497±0.017
0.6	0.586 ± 0.111	0.592 ± 0.076	0.598±0.053	0.601±0.038	0.601±0.026	0.602±0.018
0.7	0.679 ± 0.107	0.691 ± 0.075	0.693 ± 0.054	0.694 ± 0.038	0.698±0.027	0.697±0.018
0.8	0.774 ± 0.099	0.787 ± 0.070	0.793 ± 0.052	0.797±0.039	0.799±0.027	0.797±0.019
0.9	0.863 ± 0.087	0.880 ± 0.062	0.888 ± 0.047	0.895 ± 0.035	0.898±0.025	0.897±0.019

TAB. 4.6 – Estimations par EMV avec Irr_S, comparaison avec le tableau 4.5. Les cellules en gris clair (resp. foncé) correspondent au non rejet de H_0 pour le test de Fisher (resp. Fisher et Student).

H\N	32	64	128	256	512	1024
0.1	0.109±0.060	0.101±0.040	0.101±0.029	0.099±0.020	0.101±0.013	0.101±0.010
0.2	0.204±0.077	0.202±0.053	0.202±0.036	0.202±0.025	0.201±0.017	0.201±0.013
0.3	0.300±0.087	0.300±0.060	0.300±0.038	0.302±0.027	0.300±0.020	0.302±0.015
0.4	0.396±0.090	0.398±0.062	0.397±0.042	0.399±0.028	0.397±0.020	0.402±0.016
0.5	0.495±0.092	0.496±0.064	0.502±0.045	0.500±0.032	0.501±0.022	0.502±0.017
0.6	0.589±0.091	0.594±0.064	0.596±0.044	0.599±0.029	0.598±0.021	0.602±0.018
0.7	0.692±0.092	0.698±0.064	0.695±0.045	0.699±0.032	0.697±0.022	0.701±0.018
0.8	0.784±0.086	0.791±0.062	0.797±0.044	0.801±0.032	0.798±0.021	0.801±0.018
0.9	0.877±0.071	0.887±0.054	0.897±0.041	0.898±0.031	0.897±0.021	0.901±0.018

TAB. 4.7 – Estimations par EMV avec Irr_U, comparaison avec le tableau 4.5.

Influence de l'irrégularité d'échantillonnage sur le biais de l'estimateur ($Stud_2$)

Nous étudions maintenant l'importance du biais de l'estimateur en fonction du choix de l'irrégularité d'échantillonnage. Pour cela, et en complémentarité avec le calcul de l'IC, nous appliquons un test de Student d'égalité de l'espérance de l'estimateur avec la valeur théorique ($H_0 : E(\widehat{H}) = H$). Les tableaux 4.8, 4.9 et 4.10 présentent l'ensemble des p-values du test pour chaque type d'échantillonnage. Puisque le test est bilatéral symétrique, il faut comparer les résultats obtenus avec la moitié du seuil de référence choisi initialement ($\frac{\alpha}{2} = 0.005$). Le test de Student est équivalent à déterminer si la valeur théorique appartient à l'IC mais a l'avantage de préciser l'analyse par l'intermédiaire des p-values.

Pour les trois tableaux, le non-rejet du test correspond globalement aux signaux de grande longueur. Plus la taille des signaux est importante et plus la valeur théorique appartient à l'IC, d'où le non-biais asymptotique de l'estimateur. Il y a 35 non-rejets pour la régularité d'échantillonnage, 30 pour Irr_S, 42 pour Irr_U et nous obtenons donc de ce point de vue les meilleurs résultats avec Irr_U. La moyenne des p-values des non-rejets est de 0.221 pour la régularité d'échantillonnage, 0.133 pour Irr_S et 0.168 pour Irr_U. La régularité d'échantillonnage accepte moins de tests que Irr_U, mais les p-values sont globalement plus élevées. Dans un intervalle fixé, plus le nombre d'observations est grand et plus l'échantillonnage Irr_U tend vers un échantillonnage régulier. Ce qui explique que les résultats de Irr_U se rapprochent de la régularité d'échantillonnage en termes de p-values quand N grandit.

En analysant le maximum des p-values par colonne, on observe une certaine structure pour chacun des tableaux. Pour la régularité d'échantillonnage, le maximum est atteint pour $H = 0.1$, pour des signaux de longueur faible, jusqu'à $H = 0.9$ pour des signaux longs. Pour Irr_S, le maximum est atteint pour $H = 0.3$ quelque soit N et pour Irr_U, la dynamique est plus aléatoire mais concerne aussi pour la plupart des cas $H = 0.3$. L'irrégularité d'échantillonnage a donc moins d'influence pour les H proches de 0.1 car les variations des signaux sont plus grandes. Par conséquent, l'irrégularité Irr_S rajoute globalement un biais. Pour Irr_U, les valeurs obtenues avec une irrégularité uniforme sont plus proches du modèle théorique : le test de Student pour l'IC accepte 42 valeurs contre 30 pour Irr_S et 35 pour la régularité d'échantillonnage.

Nous concluons donc un biais réduit pour la régularité d'échantillonnage mais aussi pour Irr_U. L'irrégularité "déterministe" Irr_S donne significativement de moins bons résultats, ce qui coïncide avec les analyses de l'IC, impliquant un biais plus prononcé.

H\N	32	64	128	256	512	1024
0.1	$< 10^{-16}$	0.426	0.107	0.229	0.076	0.081
0.2	0.173	0.241	0.409	0.253	0.268	0.103
0.3	2.428×10^{-6}	0.005	0.108	0.246	0.081	0.141
0.4	5.211×10^{-12}	3.333×10^{-6}	0.003	0.066	0.383	0.350
0.5	1.636×10^{-13}	0.001	0.088	0.387	0.183	0.234
0.6	$< 10^{-16}$	5.158×10^{-5}	0.047	0.397	0.360	0.248
0.7	$< 10^{-16}$	5.535×10^{13}	3.269×10^{-6}	0.002	0.316	0.258
0.8	$< 10^{-16}$	4.351×10^{-9}	0.006	0.216	0.406	0.124
0.9	$< 10^{-16}$	$< 10^{-16}$	2.526×10^{-13}	8.625×10^{-6}	0.277	0.426

TAB. 4.8 – Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas régulier. Les cellules grisées correspondent au non-rejet du test (p-values $> \frac{\alpha}{2}$).

H\N	32	64	128	256	512	1024
0.1	$< 10^{-16}$	0.004	0.056	0.048	0.021	0.046
0.2	0.024	4.757×10^{-4}	0.004	0.018	0.023	0.054
0.3	6.990×10^{-5}	0.011	0.202	0.440	0.460	0.437
0.4	2.341×10^{-7}	0.001	0.060	0.082	0.141	0.137
0.5	1.110×10^{-16}	2.763×10^{-7}	1.022×10^{-4}	0.003	0.013	0.052
0.6	9.403×10^{-13}	5.379×10^{-5}	0.103	0.360	0.315	0.133
0.7	$< 10^{-16}$	8.776×10^{-7}	3.920×10^{-4}	0.001	0.122	0.077
0.8	$< 10^{-16}$	1.098×10^{-13}	2.847×10^{-4}	0.073	0.243	0.078
0.9	$< 10^{-16}$	$< 10^{-16}$	1.171×10^{-11}	0.004	0.103	0.069

TAB. 4.9 – Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas Irr_S.

H\N	32	64	128	256	512	1024
0.1	4.440×10^{-16}	0.263	0.226	0.109	0.153	0.125
0.2	0.003	0.052	0.050	0.060	0.115	0.153
0.3	0.432	0.439	0.488	0.084	0.435	0.168
0.4	0.005	0.086	0.034	0.163	0.036	0.179
0.5	0.002	0.012	0.074	0.384	0.273	0.190
0.6	4.393×10^{-11}	6.702×10^{-5}	0.008	0.228	0.129	0.200
0.7	2.078×10^{-6}	0.097	0.003	0.317	0.059	0.211
0.8	$< 10^{-16}$	2.664×10^{-9}	0.023	0.271	0.079	0.223
0.9	$< 10^{-16}$	$< 10^{-16}$	0.018	0.107	0.030	0.253

TAB. 4.10 – Test de Student ($H_0 : E(\widehat{H}) = H$) : ensemble des p-values pour le cas Irr_U.

4.3.7 Application à la segmentation multidimensionnelle

Sous les contraintes évoquées plus haut, nous proposons une méthode de segmentation multidimensionnelle appliquée à une courbe paramétrée par le temps.

Modèle proposé

On considère le modèle proposé par Norros [Nor94], utilisé pour l'étude du traffic Ethernet, qui exprime une trajectoire d'un MBF avec une tendance linéaire à estimer a posteriori :

$$\begin{array}{rccl} X_{\cdot} & : & \mathbb{R}_+ & \rightarrow & \mathcal{L}^2(\Omega) \\ & & t & \mapsto & X_t = mt + \sqrt{a}Z_t \end{array} \tag{4.79}$$

Le réel m induit une tendance linéaire à (Z_t) et la quantité a est la variance de la variable $X_1 - m$, avec l'hypothèse de départ $E(Z_1 Z_1) = 1$. Le processus (Z_t) est un mouvement brownien fractionnaire d'autosimilarité H et de matrice de covariance V_H. Dans ce cas, $(X_t - mt)$ est un MBF d'autosimilarité H et de covariance égale à :

$$E((X_{t_i} - mt_i)(X_{t_j} - mt_j)) = aE(Z_{t_i} Z_{t_j}) = aV_H(i,j) \tag{4.80}$$

Évidemment, l'espérance de la variable aléatoire X_t est égale à mt puisque (Z_t) est de moyenne nulle. Ce modèle présente donc un cadre général car dans la pratique, il est souvent nécessaire de supprimer la tendance du signal pour se ramener ensuite à la modélisation par MBF. On considère l'ensemble $\mathbf{X}$:

$$\begin{array}{rccl} \mathbf{X} & : & \mathbb{R}_+ & \rightarrow & \mathbb{R}^p \\ & & t & \mapsto & \mathbf{X}(t) = (\mathbf{x}_1(t), \ldots, \mathbf{x}_p(t)) \end{array} \tag{4.81}$$

où $\mathbf{x}_i(t)$ est le $t$ème élément de la trajectoire $\mathbf{x}_i$ du MBF (X_{it}) avec tendance, comme il est défini à la formule (4.79). Sans la tendance, le MBF (Z_{it}) associé à pour autosimilarité H_i. L'ensemble $\mathbf{X}$ est donc une courbe discrétisée et paramétrée par le temps. À t fixé, $\mathbf{X}(t)$ est un vecteur de $\mathbb{R}^p$ dont chaque élément est une réalisation de la variable aléatoire X_{it} du processus (X_{it}). La figure 4.11 donne un exemple de courbes de longueur 500 pour $p = 2$ avec différentes valeurs de H en abscisse et en ordonnée. Dans chaque cas, les trajectoires marginales sans tendance ont été simulées avec l'algorithme de Cholesky. Chaque composante de $\mathbf{X}$ possède donc une signature du paramètre de Hurst, que l'on notera H_i et que nous estimons par l'approche du maximum de vraisemblance [Leh80, LOKS86]. Pour extraire une signature conjointe et unique de $\mathbf{X}$, on fait appel aux opérateurs de fusion des signatures marginales H_i. Quelque soit l'opérateur, la signature conjointe $\mathbf{H}$ est une somme pondérée des signatures marginales :

$$\mathbf{H} = \sum_{i=1}^{p} c_i H_i \tag{4.82}$$

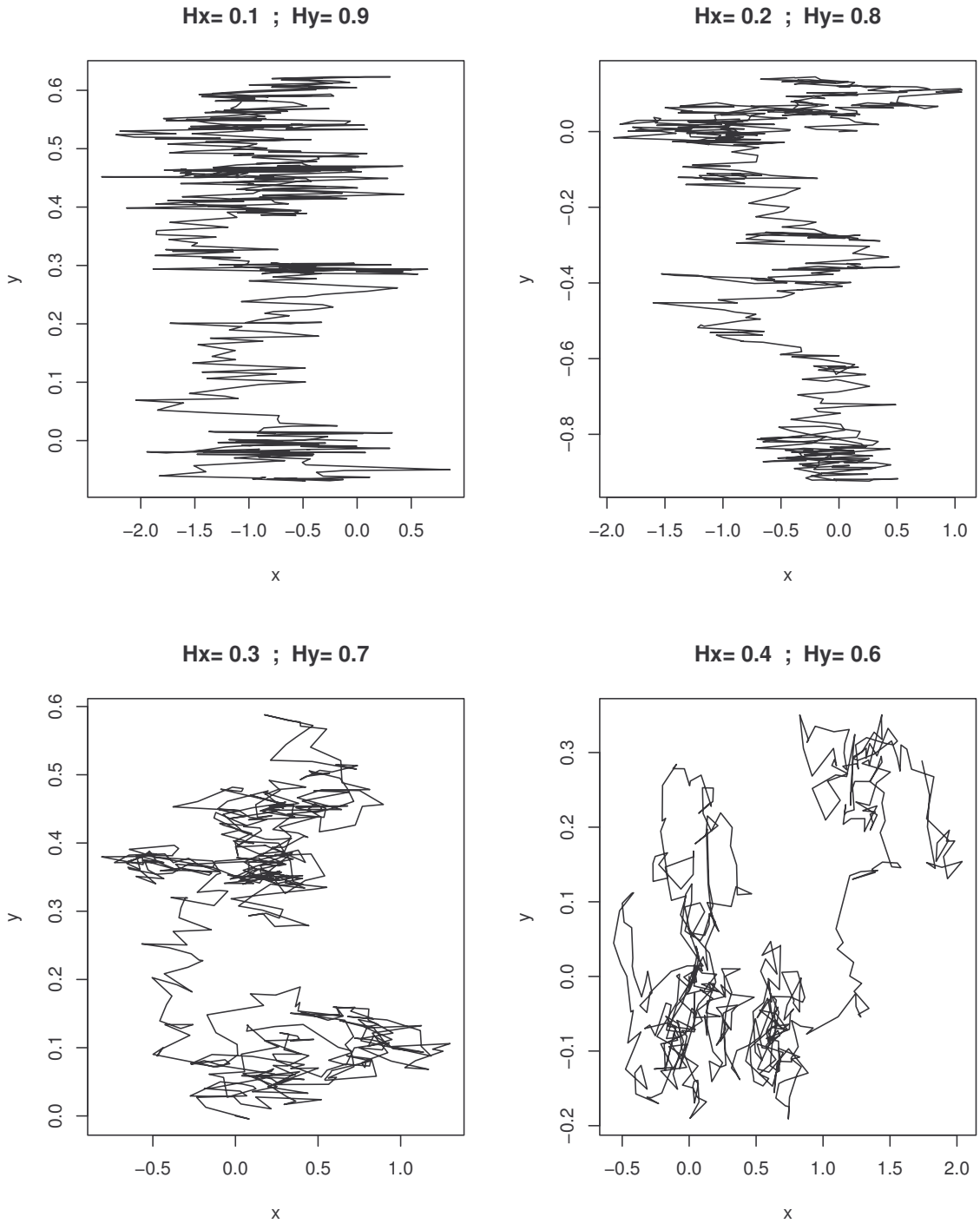

FIG. 4.11 – Trajectoires browniennes fractionnaires bidimensionnelles de longueur $N = 500$ simulées par la méthode de Cholesky pour différentes valeurs de H en abscisse et en ordonnée.

Par contre, cela ne veut absolument pas dire que **X** est autosimilaire d'indice **H**. Ce dernier étant simplement considéré comme une signature commune qui s'apparente à un opérateur de fusion en logique floue.

Algorithme de segmentation

Supposons dorénavant que nous disposions d'une trajectoire échantillonnée irrégulièrement, mais cette fois-ci d'autosimilarité constante par morceaux. L'objectif est de segmenter cette trajectoire en fonction des différentes valeurs estimées de H. Chaque valeur, noté h_i, est calculée sur un voisinage local centré en un point t_i. On obtient donc un ensemble de valeurs (soit N) et l'on note le signal à segmenter $\mathbf{h} = (h_1, ..., h_N)$. La procédure de segmentation choisie est l'approche proposée récemment par Kehagias [Keh04]. Cette méthode bénéficie à la fois de la simplicité des algorithmes de type "Sliding Windows" [KCHP93] ainsi que de la propriété de convergence vers une segmentation optimale [Keh04]. Une *segmentation* de **h** est la donnée d'un ensemble de dates de capture[2] $\overline{t}_0, \ldots, \overline{t}_K$ dont la valeur est constante sur chaque *segment* $]\overline{t}_{k-1}, \overline{t}_k]$ et égale à la moyenne des points de **h** sur ce segment. L'entier K est le nombre de segments, appelé aussi l'*ordre* de la segmentation. La segmentation est *optimale* car elle vérifie un critère de minimisation d'un certain coût [Hub00]. Précisément, l'hypothèse qui sera faite ici est que sur chaque segment $[\overline{t}_{k-1}, \overline{t}_k]$, les points de **h** sont distribués suivant une loi gaussienne de moyenne :

$$\widehat{\mu}_k = \frac{\sum_{t=\overline{t}_{k-1}+1}^{\overline{t}_k} h_t}{card(]\overline{t}_{k-1}, \overline{t}_k])} \tag{4.83}$$

et d'écart-type non biaisé :

$$\widehat{\sigma}_k = \sqrt{\frac{\sum_{t=\overline{t}_{k-1}+1}^{\overline{t}_k} (h_t - \widehat{\mu}_k)^2}{card(]\overline{t}_{k-1}, \overline{t}_k]) - 1}} \tag{4.84}$$

La technique de minimisation est fondée sur les modèles de Markov cachés. Le modèle est un gauche-droite de matrice de transition :

$$P = \begin{pmatrix} p & 1-p & 0 & \cdots & \cdots & 0 \\ 0 & \ddots & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 1-p & \ddots & \vdots \\ \vdots & & \ddots & p & 1-p & 0 \\ 0 & \cdots & \cdots & 0 & 0 & 1 \end{pmatrix} \tag{4.85}$$

[2]Les dates de capture $\overline{t}_0$ et $\overline{t}_K$ coïncident naturellement avec t_1 et t_N.

La probabilité de transition p, choisie a priori par l'utilisateur, mesure le potentiel de passer d'un segment à un autre. Plus cette probabilité est grande, plus le changement d'état sera favorisé. La séquence d'état $\mathbf{z} = (z_1, \ldots, z_N)$ est définie de la façon suivante :

$$z_i = k \quad \text{pour} \quad \overline{t}_{k-1} < i \leq \overline{t}_k \tag{4.86}$$

La vraisemblance à maximiser par rapport à P, μ_k et σ_k est la suivante :

$$L(\mathbf{z}, \mathbf{h}; P, \mu_k, \sigma_k) = \prod_{t=1}^{N} \left(P_{z_{t-1} z_t} . \exp\left(-\frac{(h_t - \mu_{z_t})^2}{2\sigma_t^2} \right) \right) \tag{4.87}$$

où $P_{z_0 z_1}$ est la probabilité initiale du MMC qui doit être fixée égale à 1 d'après le choix du modèle. Cette vraisemblance est égale au produit des probabilités de transition avec les images des densités gaussiennes dont les points de $\mathbf{h}$ sont supposées être des réalisations. Les différentes étapes d'estimation de la segmentation sont obtenues via la procédure de Viterbi [Vit67], et ce afin de déterminer une séquence d'état optimale à chaque itération (*i.e.* une segmentation étant donnée l'évolution des paramètres). La complexité de calcul d'une segmentation par l'algorithme de Viterbi est en $\mathcal{O}(N.K^2)$. Les procédures sont résumées aux tables 4.11 et 4.12.

Algorithme de Viterbi pour la segmentation de série temporelle

Entrées :
- Série temporelle $\mathbf{h} = (h_1, ..., h_N)$.
- Nombre d'états K.
- Matrice de transition P.
- Moyennes $(m_k)_{k=1}^K$ de chaque segment.
- Écarts-type $(\sigma_k)_{k=1}^K$ de chaque segment.

Sortie :
- Segmentation estimée $\widehat{\mathbf{z}} = (\widehat{z}_1, \ldots, \widehat{z}_N)$ suivant les paramètres d'entrée.

Procédure :

Récurrence : construction de la matrice $q = [q_1, \ldots, q_N]$ de taille $K \times N$:

Poser $q_0 = {}^t(1, 0, \ldots, 0)$.

Pour $t = 1, \ldots, N$

Pour $k = 1, \ldots, K$

$$q_{kt} = \max_{1 \leq j \leq K} \left(q_{j(t-1)} . P_{jk} . \exp\left(-\frac{(h_t - \mu_{z_t})^2}{2\sigma_t^2} \right) \right)$$

$$r_{kt} = \operatorname*{argmax}_{1 \leq j \leq K} \left(q_{j(t-1)} . P_{jk} . \exp\left(-\frac{(h_t - \mu_{z_t})^2}{2\sigma_t^2} \right) \right)$$

Fin Pour

Fin Pour

Terminaison :

$\widehat{z}_N = \operatorname{argmax}_{1 \leq k \leq K} q_{kN}$

Séquence d'état optimale :

Pour $t = N, N-1, \ldots, 2$

$\widehat{z}_{t-1} = r_{\widehat{z}_t t}$

Fin Pour

TAB. 4.11 – Algorithme de Viterbi pour la segmentation de série temporelle.

Procédure de segmentation pour une série temporelle

Entrées :
- Série temporelle $\mathbf{h} = (h_1, ..., h_N)$.
- Nombre d'états K (ordre de la segmentation).
- Probabilité de transition p.
- Seuil de convergence η.

Sortie :
- Segmentation optimale $\widehat{\mathbf{z}}$.

Procédure :
Choisir aléatoirement une segmentation $\widehat{\mathbf{z}}^{(0)}$.
$i \leftarrow 1$.
Calcul des $(\overline{t}_k)^{(i)}$ de $\widehat{\mathbf{z}}^{(i-1)}$.
Calcul des moyennes $\mu_k^{(i)}$ des segments à partir des $(\overline{t}_k)^{(i)}$ et 4.83.
Calcul des écarts-types $\sigma_k^{(i)}$ des segments à partir des $(\overline{t}_k)^{(i)}$ et 4.84.
Calcul de $\widehat{\mathbf{z}}^{(i)}$ à l'aide de l'algorithme de Viterbi.

Jusqu'à $|L(\widehat{\mathbf{z}}^{(i)}, \mathbf{h}; P, \mu_k^{(i)}, \sigma_k^{(i)}) - L(\widehat{\mathbf{z}}^{(i-1)}, \mathbf{h}; P, \mu_k^{(i-1)}, \sigma_k^{(i-1)})| < \eta$

TAB. 4.12 – Procédure de segmentation pour une série temporelle.

4.4 Conclusion

Dans ce chapitre, nous avons proposé des approches probabilistes pour analyser des échantillons de variables aléatoires gaussiennes résiduelles, obtenues en supprimant la tendance locale. La première partie fût dédiée à une nouvelle application de la distribution de Wishart à la classification supervisée ou non de matrices de covariance. La distribution de Wishart est typiquement utilisée pour caractériser les matrices de covariance d'échantillons d'un vecteur aléatoire gaussien. La classification supervisée permet de comparer les résultats en fonction du niveau de lissage choisi pour supprimer la tendance des variables. Pour la classification non supervisée, nous appliquons la procédure EM et validons la méthode en calculant la différence symétrique entre partitions. Concernant la sélection du modèle, nous utilisons le critère BIC permettant d'estimer un nombre pertinent de clusters à chercher.

La seconde partie visait à introduire une méthode de segmentation multidimensionnelle fondée sur la signature de l'autosimilarité, sous des contraintes de dimensionnalité et d'échantillonnage irrégulier. La méthode consiste à estimer l'autosimilarité de manière marginale et à consider ensuite une signature conjointe par le biais d'une application de fusion sur les indices estimés. Nous avons pu constater que la prise en compte d'une information exogène par la covariance a permis de pallier à la contrainte temporelle, en terme de simulation de trajectoires ou d'estimation du paramètre de Hurst. La simulation est établie à partir de la décomposition de Cholesky, méthode exacte mais coûteuse en temps de calcul. L'estimation se fait par maximum de vraisemblance dont nous avons proposé de vérifier expérimentalement le comportement de l'estimateur pour des signaux de longueur faible, ainsi que le maintien de ses propriétés asymptotiques. De manière générale, nous confirmons naturellement les résultats théoriques liés aux estimateurs du maximum de vraisemblance, et en particulier, le caractère asymptotiquement sans biais et la convergence. L'analyse de l'échantillonnage aléatoire confirme en particulier un biais réduit de l'estimateur, par rapport à un échantillonnage déterministe.

Chapitre 5

Applications

5.1 Introduction

Le présent chapitre consiste à donner un certain nombre d'exemples illustratifs des concepts présentés dans ce travail. Il sera consacré en grande partie à l'étude du mouvement humain à l'aide de figures chorégraphiques pour lesquelles nous appliquons les méthodes de statistique géométrique et le modèle de Wishart. Pour l'étude de l'autosimilarité, nous étudions un exemple de déplacement de bar évoluant dans un bassin aquacole. Les exemples seront étudiés sous différents aspects : classification, représentations graphiques, segmentation et reconstruction partielle des données. La section 2 concerne le choix et les interprétations de la connaissance exogène pour le contexte applicatif. Pour l'analyse, nous choisissons un graphe exprimant une contiguïté temporelle entre les observations. Nous donnons les interprétations de la covariance, de l'espérance et de l'ACP relationnelles via les propriétés formulées au second chapitre. La section 3 est consacrée aux applications de l'ACP relationnelle sur les données réelles : représentations graphiques et reconstruction partielle des données. La section 4 expose les exemples de l'analyse d'opérateurs avec les représentations graphiques et la classification supervisée. Nous considérons trois cas de figure : opérateurs de covariance normés ou non et opérateurs de covariance relationnelle normés. En section 5, le modèle de Wishart est appliqué à la classification de mouvements dansés. Les taux de reconnaissance pour la classification supervisée sont calculés à partir de la procédure leave-one out au plus proche centre de gravité de la classe. Pour la classification non supervisée, nous considérons la différence symétrique entre les partitions obtenues et celle de référence. Enfin, nous présentons à la section 6 un exemple de segmentation par l'autosimilarité appliquée au déplacement d'un bar.

5.2 Mise en place du formalisme

5.2.1 Expression de la connaissance exogène

Pour l'analyse, les variables aléatoires et les individus sont choisis équipondérés :

$$M = I_p \quad \text{et} \quad p_i = \frac{1}{n}\,, \quad 1 \leq i \leq n \tag{5.1}$$

On considère que l'espace Ω contient l'ensemble des dates de capture :

$$\Omega = \{t_1, \ldots, t_n\} \tag{5.2}$$

où n est égal au nombre de lignes du sous-tableau considéré, et doit donc être systématiquement remplacé par la taille du mouvement associé. Pour l'analyse du mouvement, nous reprenons le graphe défini au chapitre précédent :

$$g_\theta = \{(t_i, t_j) \in \Omega \times \Omega \ / \ |t_i - t_j| \leq \theta\} \tag{5.3}$$

Il s'agit d'une relation de contiguïté temporelle, symétrique et paramétré par θ qui va limiter les liaisons pour le graphe entre les individus. Le paramètre θ est une variable aléatoire exogène fixée. Cette variable induit une relation binaire agissant sur les observations qui constituent les dates de capture, et sa valeur est donc identique pour toutes les variables d'un sous-tableau. En d'autres termes, le graphe agit sur les observations associées aux individus, *i.e.* sur un groupe de réalisations des p variables et non sur une réalisation individuelle. Notons que des critères d'optimisation de θ, basées par exemple sur l'autocorrélation des variables ou encore sur le minimum d'entropie des valeurs propres d'une ACP relationnelle, peuvent être envisagés, mais nous considérerons que sa valeur est donnée a priori et n'est donc pas sujet à estimation. Si $\theta = n$, le graphe est complet et toutes les observations sont connectées entre elles par un chemin de longueur un. Dans ce cas, on notera dans la suite $\theta = max$ pour éviter toute confusion. La métrique du laplacien est une matrice-bande dont la largeur dépend de θ et de l'échantillonnage induit par les dates de capture.

5.2.2 Covariance et espérance relationnelles : exemples simples

Le graphe de la formule (5.3) est adapté à l'analyse de données spatio-temporelles en jouant un double rôle. D'une part, il permet une analyse de proximité temporelle entre les observations. En effet, deux observations espacées de plus de θ n'auront pas

d'influence puisqu'elles ne sont pas connectées pour le graphe : seuls les couples d'observations reliées par un chemin de longueur un auront un impact dans le calcul de la covariance. Le graphe tient compte également de l'ordre temporel des individus car, par construction, toute permutation d'observations va perturber la structure du graphe. Toutes ces considérations ne sont naturellement pas prises en compte par la covariance. La figure 5.1 présente en (a) une série ainsi qu'une permutation en (b). Les écarts-types globaux des deux séries sont tous deux égaux à 89.59. Avec une valeur du paramètre $\theta = 5$, l'écart-type relationnel de la série permutée est très supérieur à celle de la série initiale (82.75 au lieu 18.14). En revanche, le comportement erratique de la série permutée fournit une variance relationnelle importante. La propriété 2 sur la relation

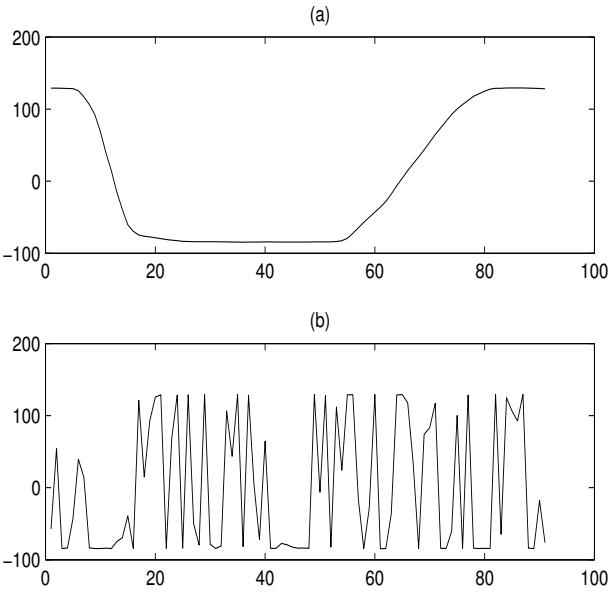

FIG. 5.1 – (a) Série temporelle X, $\sigma(X) = 89.59$ et $\sigma_g(X) = 18.14$; (b) Série $\tilde{X}$: permutation aléatoire de X, $\sigma(\tilde{X}) = 89.59$ et $\sigma_g(\tilde{X}) = 82.75$, avec $\theta = 5$.

d'équivalence de graphe nous indique qu'il n'y a aucune perte de généralité à choisir un graphe symétrique en fixant a priori la pondération des arêtes du graphe. Pour notre cas, ces poids sont choisis équiprobables car on suppose que toutes les arêtes ont la même influence sans avoir un rôle prépondérant.

L'opérateur espérance relationnelle $E(.|g_\theta)$ s'interprète comme un filtre par une fonction porte. Ainsi, le terme $X - E(X|g_\theta)$, que l'on suppose gaussien, ne conserve que les hautes fréquences de la variable. Pour un θ suffisamment petit, la propriété 4 mettant en jeu la covariance et l'espérance relationnelles implique que les parties linéaires d'une série n'influencent pas la variance relationnelle. La figure 5.2 est le graphique de deux

variables aléatoires. Ces variables sont orthogonales (décorrélées) pour la métrique du laplacien : $Corr_g(C_1, C_2) = 0$, avec $\theta = 5$. En revanche, la corrélation standard est non nulle ($Corr(C_1, C_2) = -0.19$), valeur significative pour le nombre d'observations considérées. Contrairement à la covariance classique, la covariance relationnelle ne tient pas compte des observations ayant eu lieu entre les dates 20 et 55. Concrètement, cela signifie que les périodes d'immobilité du danseur et plus généralement de celles où la trajectoire est rectiligne sont "gommées" par la covariance relationnelle, pour un θ suffisamment petit. L'espérance relationnelle est calculée suivant un graphe de connexité qui est modifié en fonction de la permutation de la série. On peut vérifier que la différence entre la variable aléatoire et son espérance relationnelle est nulle dès que la variable varie linéairement en fonction du temps. Donc, les observations pour lesquelles la variable aléatoire coïncide avec son espérance relationnelle n'influent pas sur la variance relationnelle puisque celle-ci est nulle d'après la propriété 4.

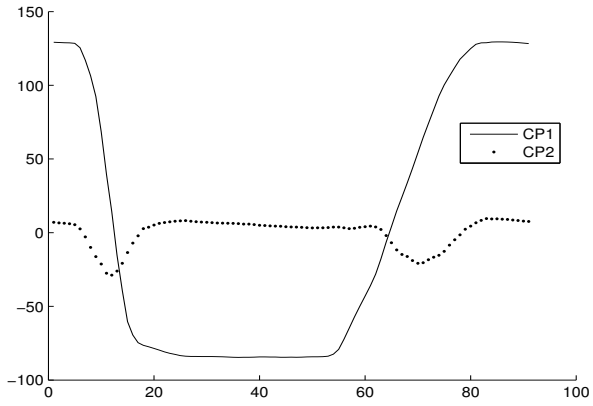

FIG. 5.2 – Représentation graphique de deux variables aléatoires décorrélées pour la covariance relationnelle ($\theta = 5$) et non décorrélées pour la covariance.

5.2.3 ACP relationnelle : exemples simples

Nous proposons d'analyser deux exemples simples montrant la justification de l'emploi de la méthode. On considère deux variables aléatoires telles que la disposition des observations sur le plan soit en "dents de scie" (fig. 5.3). On pratique deux ACP : une relationnelle avec $\theta = 5$ et une standard. Le premier vecteur propre de chaque ACP est montré figure 5.3. On remarque que la direction locale de la trajectoire analysée est bien mise en valeur par une ACP relationnelle contrairement à l'ACP standard où le vecteur (G) caractérise la dispersion globale du nuage.

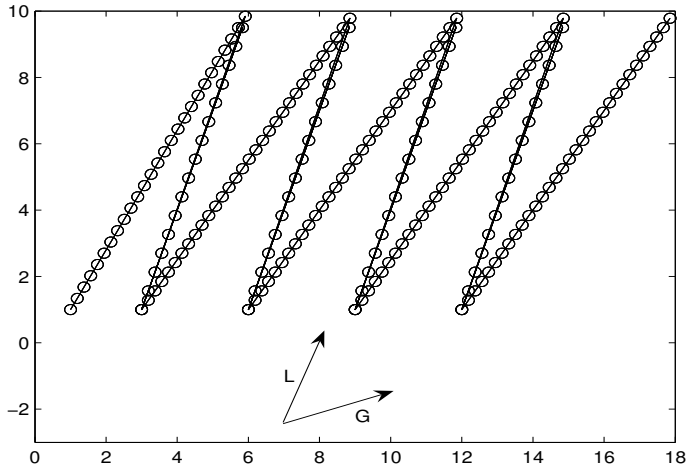

FIG. 5.3 – Disposition des observations (petits cercles) avec deux vecteurs propres : l'un issu d'une ACP standard (G) et l'autre d'une ACP relationnelle (L) avec $\theta = 5$.

Pour le second exemple, on substitue le graphe de la formule (5.3) par une relation de voisinage définie comme l'appartenance (=1) ou non (=0) de deux observations à une même classe. L'information exogène est alors donnée par une variable indicatrice de catégorie sur les observations. Cet exemple est illustré à la figure 5.4. Comme précédemment, le vecteur (G) donne la direction principale d'inertie de l'ensemble alors que le vecteur (L) s'attache d'autant plus à la relation imposée sans pour autant discriminer de manière optimale les classes. Notons pour cela que l'ACP relationnelle n'est en général pas équivalente à l'analyse discriminante. Le vecteur (D) sur la figure représente le premier axe d'une analyse discriminante simple appliquée aux observations. Le vecteur (L) effectue une synthèse des directions principales locales à chaque classe, et ne coïncide pas avec le vecteur (D), qui distingue plutôt les centres de gravité et donc les classes.

5.3 ACP relationnelle

5.3.1 Présentation des données

L'analyse du mouvement assisté par ordinateur est un sujet de Recherche récurrent [AC99, CS95, Per01] et l'on peut citer les trois axes applicatifs majeurs [MG01] : la surveillance, le contrôle et l'analyse. La *surveillance*, qui se situe principalement dans le

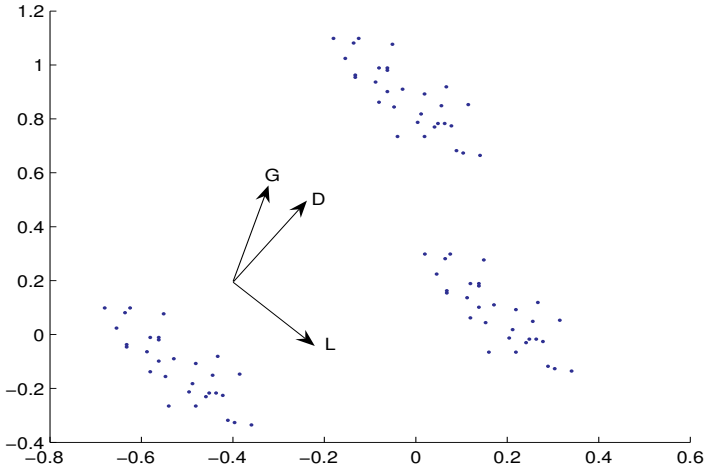

FIG. 5.4 – Exemple sur un ensemble de trois classes : ACP standard (G), ACP relationnelle (L) et analyse discriminante (D).

contexte du traitement d'images, désigne toutes les approches permettant de détecter et suivre un objet au cours du temps. Par exemple, l'un des objectifs peut être d'identifier des comportements anormaux ou suspects d'un ou plusieurs objets cible filmés dans un lieu sécurisé, comme un parking, un musée, une banque ... Le *contrôle*, que l'on retrouve presque systématiquement dans le développement des jeux vidéo, concerne les techniques de capture permettant de calquer les gestuelles et les comportements d'un objet. Le but étant de développer certaines fonctionnalités, d'animer un personnage, ou encore de créer une réalité virtuelle sur un ordinateur. Enfin, l'*analyse* fait référence aux méthodes qui ont pour finalités premières de "comprendre" le mouvement, c'est à dire d'en extraire l'information pertinente, que l'on modélise à l'aide de techniques formelles par des procédés statistiques ou géométriques.

Le cadre applicatif de ce travail s'inscrit dans ce troisième axe. Le Ballet Atlantique Régine Chopinot (BARC) a été fondé en 1986 sous la première appellation du Centre Chorégraphique National de la Rochelle. Les danseurs membres de ce ballet se sont prêtés au jeu de l'enregistrement de leur performance chorégraphique à l'aide de tout un arsenal de capture (figure 5.5).

Les danseurs sont équipés de 15 marqueurs positionnés sur diverses parties de leur corps : bras, tête, poignets, mains, torse, pelvis, hanches, jambes et pieds. Les coordonnées spatiales sont enregistrées pendant l'exécution d'un mouvement typique de danse contemporaine avec une fréquence d'échantillonnage de 25 Hz. Un mouvement constitué de n captures est donc identifié par un ensemble de $p = 15 \times 3 = 45$ variables

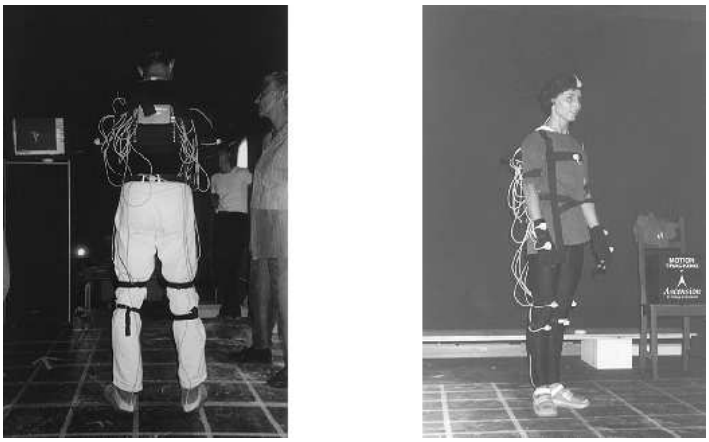

FIG. 5.5 – Danseurs équipés du dispositif de capture.

aléatoires réelles discrétisées, rangées dans un tableau de taille $n \times 45$. Les mouvements ont été capturés au cyberdôme du Centre National de la Bande Dessinée et de l'Image (CNBDI) à Angoulême. Une représentation schématique du Cyberdôme est présentée à la figure 5.6. Grâce à la collaboration de XD-Productions, le système de capture est doté de marqueurs électro-magnétiques qui fournit la position spatiale des capteurs au cours du temps. Nous disposons d'un corpus de 219 mouvements répartis dans les 14 groupes de mouvements typiques de danse contemporaine. Le tableau 5.1 nous donne l'appellation de chaque groupe ainsi que leur effectif. Nous disposons au total d'un

Mouvement-type	Effectif
Marche	17
Enveloppé	13
Chute pliée	20
Enroulé tête	15
Enroulé bassin	21
Grand plié	8
Marche glissée	8
Saut attitude	12
Chute repliée	18
Chute allongée	22
Jeté	25
Tour	22
Suspension	12
Saut temps levé	6

TAB. 5.1 – Mouvements-type de danse contemporaine et effectif par groupe.

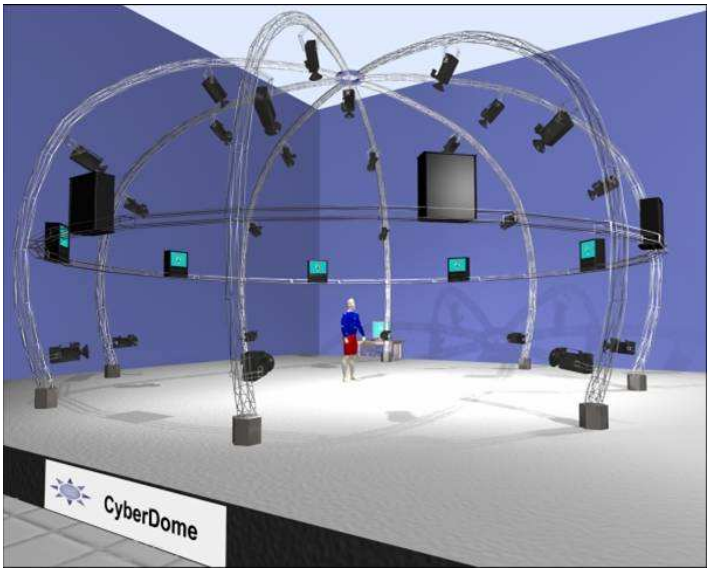

FIG. 5.6 – Cyberdôme du CNBDI.

corpus de $N = 219$ mouvements se répartissant dans l'un des 14 mouvements type[1] : Marche (ma), Enveloppé (en), Chute Pliée (cp), Enroulé Tête (et), Enroulé Bassin (eb), Grand Plié (gp), Marche Glissée (mg), Saut Attitude (sa), Chute Repliée (cr), Chute Allongée (ca), Jeté (jt), Tour (tr), Suspension (su) ou Saut Temps Levé (st). Chaque mouvement type a été réalisé à plusieurs reprises par des danseurs de différentes corpulences et possédant chacun un niveau de maîtrise propre de la danse. Les classes de mouvements ne sont pas homogènes et vont de 6 éléments pour la classe 'Saut Temps Levé' à 25 pour la classe 'Jeté' ($6 \leq N_k \leq 25$). De même, les tailles des mouvements vont de 52 à 338 captures ($52 \leq n_k \leq 338$). La figure 5.7 donne le schéma de la structure du graphe considéré avec $\theta = 3$ en ayant fixé une date de capture t_i. Comme il a été vu au chapitre 2, une arête reliant deux observations est pondérée par le produit des poids des individus correspondants[2].

La capture des mouvements respecte la norme standard internationale H-ANIM ISO/IEC 19774 de 2006 qui spécifie la forme et la structure d'humanoïdes représentés sur ordinateur. Formellement, pour un mouvement individuel, on se ramène à procéder à l'analyse simultanée d'un groupe de trajectoires représentant le déplacement tridimensionnel d'un capteur. L'ensemble nous donne la description du déplacement d'un squelette utilisé conventionnellement en capture de mouvement humain avec marqueurs [ACLS98].

[1]Les abréviations seront utilisées pour les représentations graphiques.
[2]On utilise la notation p_i au lieu de p_{t_i} pour ne pas encombrer les notations.

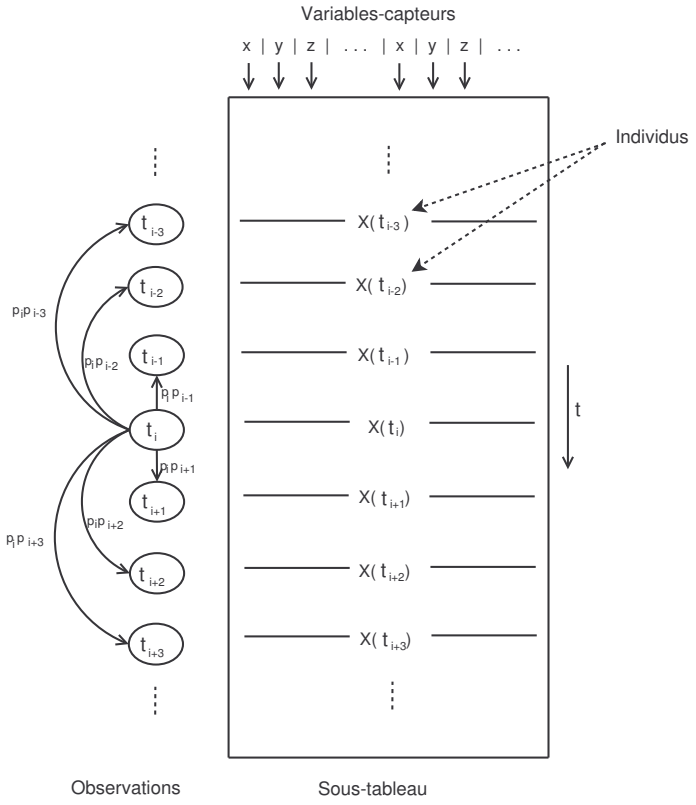

FIG. 5.7 – Structure du graphe ($\theta = 3$) pour une observation fixée t_i.

Malgré des liens évidents entre l'informatique et la danse, comme l'élaboration d'effets visuels pendant une représentation chorégraphique, la simulation d'une réalité virtuelle ou encore le développement d'un logiciel d'auto-évaluation, ce n'est que très récemment que des études scientifiques concernant le mouvement dansé ont été publiées [CBV04, TM04]. Dans la thèse [Che04], Chenevière applique les modèles de Markov cachés (MMC) à la reconnaissance et la classification de données spatio-temporelles de coordonnées de capteurs équipant des danseurs. La représentation proposée est de type variation angulaire en considérant les coordonnées sphériques de la trajectoire des capteurs. Cette étape de pré-traitement est au détriment de la perte d'information relative au positionnement et à l'orientation. Une approche locale est envisagée où chaque capteur est associé à un MMC auquel on attribue un poids suivant son pouvoir discriminant relatif au mouvement considéré. Les difficultés et motivations qui nous amènent à étudier ce type de données sont les suivantes :

- Information commune.
- Immobilité et sensibilité des capteurs.
- Typologie des mouvements.

Pour le premier point, il apparaît par exemple des distances inter-capteurs morphologiquement constante : distances coude/épaule, pied/genou, poignet/coude ... qui constitue une part de l'information commune à tous les mouvements. Il faudrait donc localiser et supprimer cette redondance afin de présenter des méthodes de discriminations plus fiables. Le second point désigne les imprécisions des capteurs qui imposent d'en étudier l'impact sur l'analyse et la description des mouvements. Enfin, pour étudier la typologie des mouvements présentant des caractéristiques hétérogènes (trajectoire rectiligne, nombre de captures ...) même pour deux mouvements d'un même groupe, il faut considérer des signatures spécifiques afin de pouvoir les identifier et les comparer. Toutes ces considérations seront étudiées dans le présent travail. Une fois le mouvement capturé, nous disposons d'un logiciel développé dans le cadre de la thèse de Chenevière [Che04]. Il s'agit d'un logiciel de rendu graphique permettant l'affichage de l'évolution des 15 capteurs, dont la figure 5.8 en donne une illustration. À chaque instant il est possible de repérer la position d'un capteur ou de visualiser la courbe décrivant sa trajectoire.

5.3.2 Représentations graphiques

Notre analyse est basée sur les coordonnées des capteurs par rapport à un repère cartésien fixe dans l'espace. Ainsi, un même mouvement exécuté avec une rotation préalable des capteurs produira des projections des variables différentes. Cependant, les vecteurs propres subissent la même rotation tandis que les valeurs propres et les CP

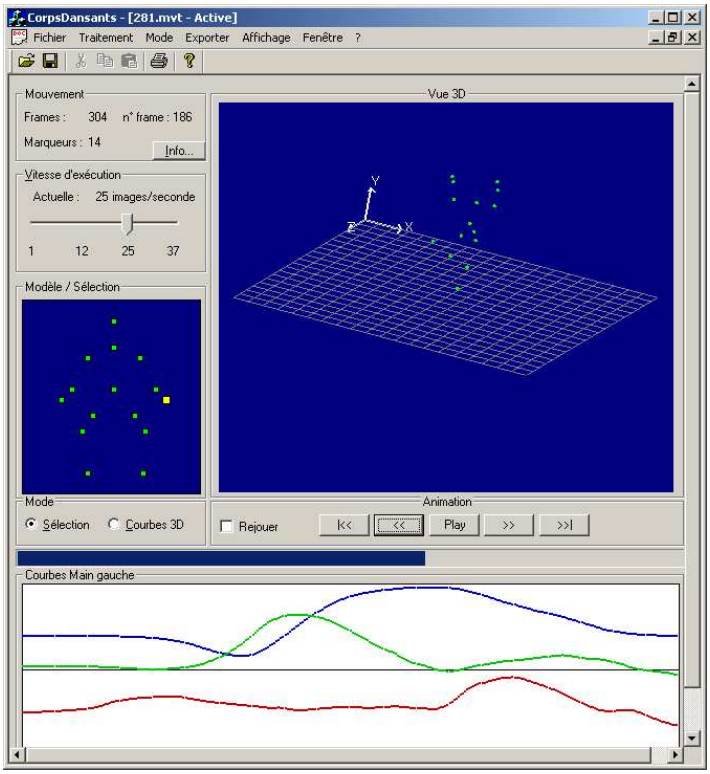

FIG. 5.8 – Logiciel de reconstitution 3D de mouvements dansés.

sont indépendantes de l'orientation initiale du danseur. La position initiale du danseur quant à elle n'a aucune influence sur l'analyse à cause du centrage automatique des variables aléatoires par le laplacien. Puisque la métrique M est l'identité, chaque CP est un vecteur constitué des produits scalaires usuels entre le vecteur propre associé et chacun des individus. L'invariance des CP implique donc l'invariance des projections du nuage des instants sur les plans factoriels. Comme la position initiale du danseur par rapport au repère est arbitraire, une des alternatives serait soit de procéder à un prétraitement des données, soit de travailler avec le tableau des distances inter-capteurs évoluant au cours du temps, ce qui produirait la mise en valeur de caractéristiques différentes du mouvement.

Dans le contexte applicatif de l'analyse du mouvement, la propriété 4 signifie que les périodes de temps où la trajectoire du mouvement est rectiligne n'a aucune influence sur le calcul de la covariance relationnelle. Ces dates de captures n'ont en effet aucun intérêt pour l'analyse locale puisque l'exécutant est immobile. Dans cette posture, le danseur exprime une attitude mais la covariance relationnelle ne prend en compte que les liens éventuels entre les déplacements des capteurs.

Nous présentons maintenant des exemples de l'ACP relationnelle appliquée à l'étude

des mouvements Marche et Tour [HLSJ06a]. Pour les représentations graphiques, les variables aléatoires sont étiquetées par leur abréviation. Par exemple, "Bgy" désigne le Bras Gauche en y, Pdx le Pied Droit en x. Pour éviter la confusion, "W" désigne le poignet ("Wrist" en anglais). Du fait de sa périodicité et de sa linéarité, la Marche est un mouvement relativement simple pour notre étude. En effet, des caractéristiques propres au mouvement se dégagent clairement sur les espaces des variables et des individus. La Marche est constituée d'un nombre $n = 200$ individus. Pour ce mouvement, le danseur se déplace suivant la bissectrice du plan xOz[3]. Pour une ACP standard (fig. 5.9), on distingue sur le premier plan factoriel deux groupes de variables. À gauche, l'ensemble des coordonnées en x et z des capteurs contribue fortement au premier axe. À droite, très proche de l'origine, il y a toutes les coordonnées en y. Celles-ci évoluent peu puisque, globalement, la Marche est une translation horizontale des capteurs. C'est pourquoi ces variables sont proches de l'origine sur ce plan, qui du reste possède la quasi-totalité de l'inertie avec 99.85%.

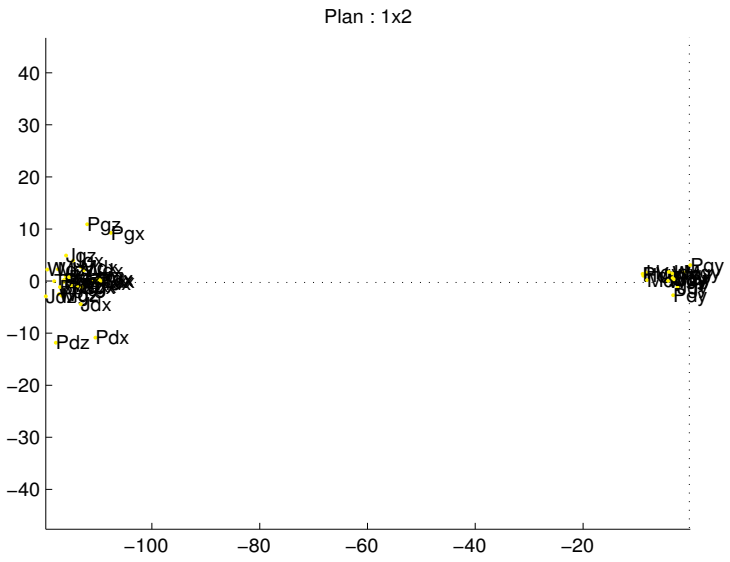

FIG. 5.9 – ACP standard : projection des variables aléatoires sur le plan 1×2 (mouvement Marche), inertie : 99.85%.

Pour une ACP relationnelle ($\theta = 5$, fig. 5.10), on constate la mise en valeur de mouvements locaux caractérisés par les pas du danseur. Sur le groupe de gauche, les couples (Pgx,Pgz) et (Pdx,Pdz) sont décorrélés : si le pied gauche est fixe, le pied droit est en mouvement et vice versa. On peut également distinguer une moindre décorrélation entre (Jgx,Jgz) et (Jdx,Jdz). Alors que le pied est immobile, la jambe bouge, ce qui implique

[3]Le trièdre (x, y, z) est direct mais c'est l'axe y qui détermine la hauteur.

une corrélation néanmoins faible du fait de l'ampleur. D'autre part, les variables en x et z nous renseignent clairement quant aux rôles de la première et seconde CP. Le premier axe caractérise la translation du mouvement et le second axe est porté par le déplacement des pieds.

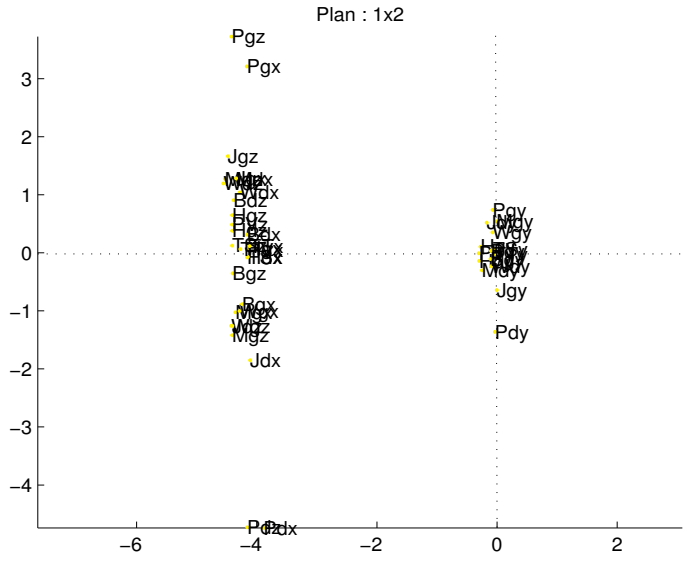

FIG. 5.10 – ACP relationnelle avec $\theta = 5$: projection des variables aléatoires sur le plan 1×2 (mouvement Marche), inertie : 88.96%.

À titre de vérification, on analyse la projection des individus sur le plan engendré par les deux premiers vecteurs propres (fig. 5.11). Puisque le premier axe est porté par la translation des capteurs, la contribution à celui-ci est d'autant plus élevée que la durée du mouvement est grande. Pour le mouvement étudié, le danseur effectue sept pas, visibles sur la figure 5.11. En fonction de la disposition des variables (fig. 5.10), on reconnaît le premier pas décrit par le pied gauche, le second par le pied droit et ainsi de suite. On remarque également que le danseur passe par l'origine du repère spatiale entre son troisième et quatrième pas.

Le tableau 5.2 mentionne les parts d'inertie contenues dans les quatre premiers axes factoriels ainsi que leur cumul. Avec une analyse locale, on constate une entropie plus forte des parts d'inertie sur les axes, où des mouvements locaux sont mis en valeur. Ces derniers étant non perçus avec une ACP standard.

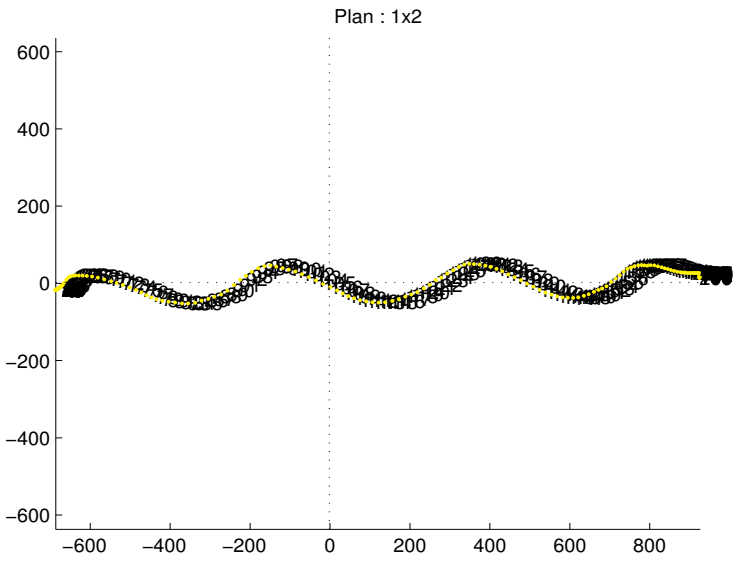

FIG. 5.11 – ACP relationnelle avec $\theta = 5$: projection des individus sur le plan 1×2 (mouvement Marche).

Valeurs propres	λ_1	λ_2	λ_3	λ_4
ACP standard	0.9970	0.0015	0.0007	0.0003
Cumul	0.9970	0.9985	0.9992	0.9995
ACP relationnelle ($\theta = 5$)	0.7584	0.1312	0.0363	0.0232
Cumul	0.7584	0.8896	0.9259	0.9491

TAB. 5.2 – Part d'inertie des 4 premiers axes pour le mouvement Marche.

5.3.3 Reconstruction partielle

Afin de donner une interprétation "visuelle" des axes, il est possible de reconstruire approximativement le mouvement en se restreignant à un nombre limité de CP. Le nouveau sous-tableau partiel $\widehat{\mathbb{X}}$ est donné par la relation :

$$\widehat{\mathbb{X}} = \widetilde{C}{}^t\widetilde{U} + \mathbf{1}.\overline{X} \tag{5.4}$$

où $\widetilde{C}$ est la matrice des CP choisies (colonnes de C), $\widetilde{U}$ la matrice des vecteurs propres correspondants (colonnes de U) et $\overline{X}$ le vecteur ligne constitué des espérances des variables aléatoires de $\mathbb{X}$. Ainsi, il est facile de vérifier de manière visible le rôle et la part d'information restituée par le ou les axes sélectionnés pour la reconstruction partielle. Nous présentons cet exemple avec le mouvement Tour (56 individus). Le premier plan factoriel est donné par la figure 5.12. Les coordonnées dont nous disposons étant carté-

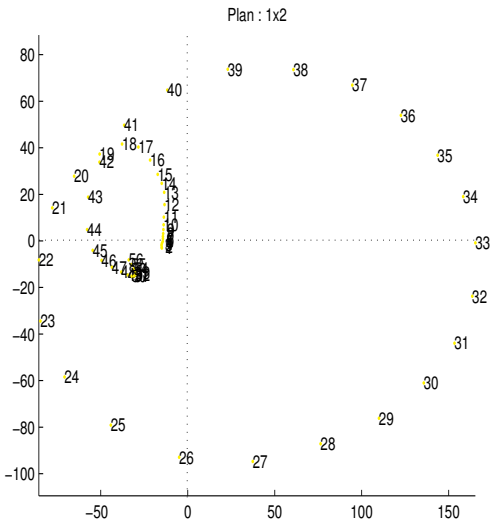

FIG. 5.12 – ACP relationnelle avec $\theta = 5$: projection des individus sur le plan 1×2 (mouvement Tour).

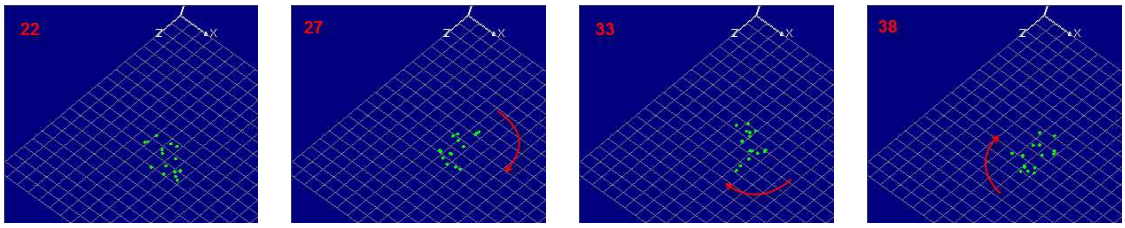

FIG. 5.13 – Mouvement Tour (vue de dessus).

siennes et non polaires, la première CP ne peut à elle seule restituer l'aspect circulaire contenu dans la gestuelle du danseur. La figure 5.13 montrent les captures correspondant aux individus se situant aux extrémités du plan factoriel (capture 22, 27, 33 et 38). Les résultats de la reconstruction partielle sont présentés à la figure 5.14. Sans le second axe nécessaire pour retrouver la circularité du mouvement, le tour est en quelque sorte "écrasé" sur un plan parallèle à xOy. La reconstruction de la Marche avec le premier axe présenterait une simple translation horizontale des points associés aux capteurs. Reconstruit partiellement en omettant cet axe, le mouvement obtenu est privé de l'information de translation, ce qui aboutit à l'exécution d'une Marche sur place.

La figure 5.15 présente l'erreur quadratique en fonction de θ de la reconstruction d'un Tour en (a) et d'une Marche en (b) avec les deux premiers axes principaux. La quantité calculée est la norme au carrée de Hilbert-Schmidt de la différence des tableaux,

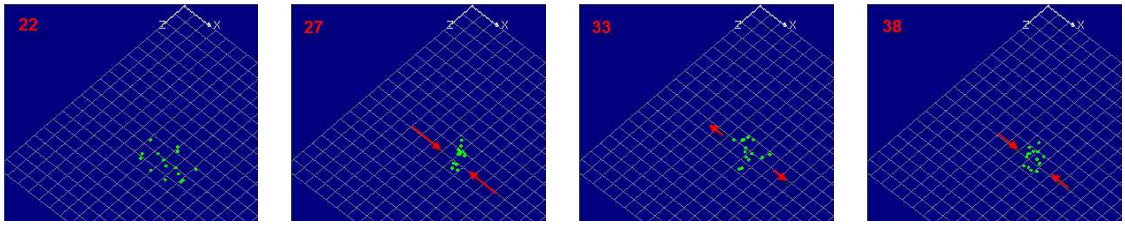

FIG. 5.14 – Reconstruction partielle d'un Tour avec le premier axe (vue de dessus) : perte de la circularité du mouvement.

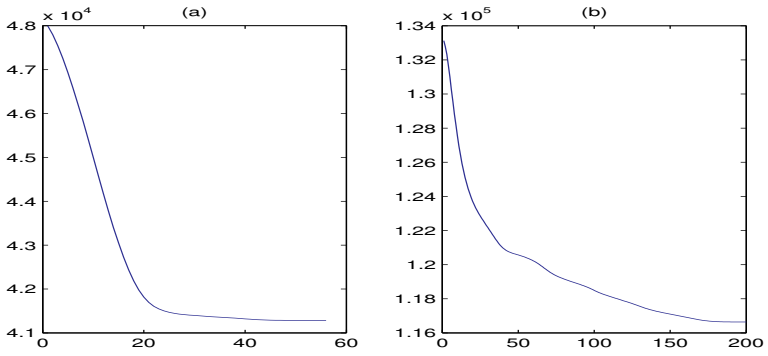

FIG. 5.15 – Norme au carré de la différence entre le tableau initial et la reconstruction partielle des 2 premiers axes en fonction des valeurs de θ, mouvement Tour (a) et Marche (b).

égale à $trace({}^t(\mathbb{X} - \widehat{\mathbb{X}})(\mathbb{X} - \widehat{\mathbb{X}}))$. On remarque que plus θ est grand (proche de l'ACP standard), plus l'erreur de reconstruction diminue. En accord avec le tableau 5.2, l'ACP relationnelle cherche à subdiviser l'information sur les axes de manière à extraire des comportements locaux du mouvement.

5.4 Analyse factorielle d'opérateurs

5.4.1 Choix des paramètres et interprétations

Il est important de rappeler que la signature choisie est la matrice de covariance et non de corrélation entre les variables-capteurs. Le système d'acquisition étant susceptible de fournir des données bruitées, normer préalablement les variables-capteurs a pour effet d'accroître ce bruit pour le calcul de la signature. Par ailleurs, l'utilisation de

la corrélation implique que toutes les variables-capteurs ont la même importance quel que soit le mouvement et cela accentue donc le rôle de certaines d'entres elles moins significatives. La normalisation d'un opérateur signifie que les opérateurs sont unitaires pour la norme de Hilbert-Schmidt :

$$\|\Gamma_k\|_{HS} = \sqrt{trace(\Gamma_k^2)} \tag{5.5}$$

Plus la norme d'un opérateur est grande et plus la covariance entre les variables-capteurs ainsi que leur variance sont globalement fortes. En ce sens, une normalisation homogène des variables par le coefficient RV plutôt que l'usage de la corrélation est plus pertinente et permet un plongement des opérateurs dans la sphère unité, ce qui facilite les interprétations. Si les opérateurs sont normés, on rappelle que le produit scalaire de Hilbert-Schmidt correspond au coefficient RV d'Escoufier. Concrètement, ce coefficient coïncide avec le cosinus entre les opérateurs et rend compte des proportionnalités entre les signatures, par exemple pour deux mouvements de même type exécutés par deux danseurs de corpulences différentes. Par ailleurs, on associe naturellement à la norme de Hilbert-Schmidt la distance suivante :

$$d(\Gamma_i, \Gamma_j) = \|\Gamma_i - \Gamma_j\|_{HS} \tag{5.6}$$

Cette valeur est faible lorsque les covariances entre les variables-capteurs sont mutuellement similaires pour deux mouvements.

5.4.2 Représentations graphiques

Pour les représentations graphiques de l'analyse d'opérateurs, nous prélevons un échantillon de 60 mouvements dans la population des 219 mouvements de référence et ce, afin de rendre plus claires les interprétations. Ce sous-ensemble contient huit instances de mouvement pour la classe Jeté, et quatre instances pour toutes les autres classes. Pour un ensemble de N mouvements constitués chacun de p variables-capteurs, l'espace potentiellement utilisé est de dimension $d = min\left(N, \frac{p(p+1)}{2}\right)$. Dans le cadre applicatif de ce travail, on a $N \ll \frac{p(p+1)}{2}$ et du fait des contraintes physiques et morphologiques, la possibilité d'exécution d'un mouvement par un être humain est naturellement limitée. On peut donc raisonnablement penser que la dimension des sous-espaces contenant l'information pertinente est très inférieure à d.

Pour l'analyse (et la classification) des opérateurs, nous considérons trois cas de figure : covariance normée ou non et covariance relationnelle normée. Pour le dernier cas, on fixe arbitrairement une valeur de θ égale à 10. La figure 5.16 donne les graphiques des sommes cumulées des valeurs propres d'une analyse relationnelle d'opérateurs pour

chacun des cas. Les unités en ordonnée sont comprises entre 0 et 1 mais représentent en fait le pourcentage d'information contenue dans les premiers axes. Le graphique en (a)

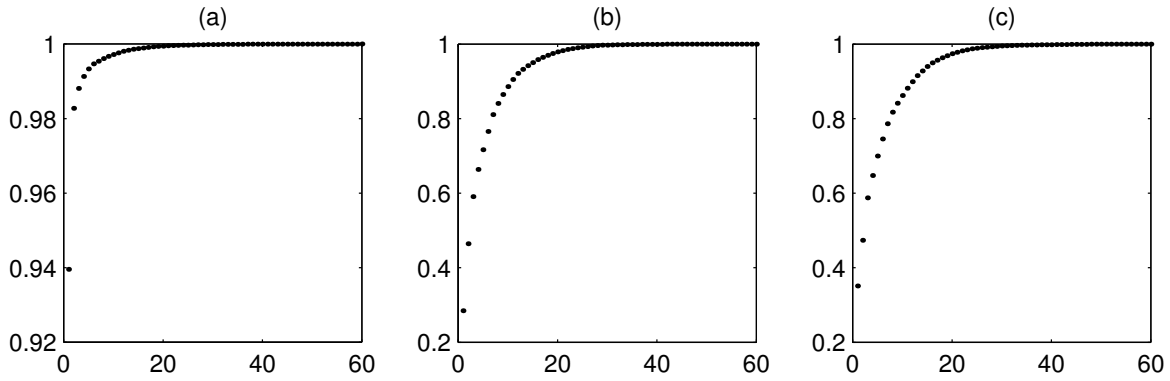

FIG. 5.16 – Somme cumulée des parts d'inertie des opérateurs principaux d'une analyse d'opérateurs de covariance relationnelle ; (a) Opérateurs non normés et $\theta = max$; (b) Opérateurs normés et $\theta = max$; (c) Opérateurs normés et $\theta = 10$.

montre que pour une non-normalisation des opérateurs, la part d'inertie sur les deux premiers axes est de plus de 99% de la totalité de l'information. Les graphiques en (b) et (c) correspondent respectivement au cas où les opérateurs normés présentent une répartition plus hétérogène qui est due à la disposition sphérique des opérateurs et à la prise en compte en plus de la proximité temporelle (pour le graphique en (c)). Les graphiques des projections sont présentés dans les trois sections qui suivent.

Analyse d'opérateurs de covariance non normés

La figure 5.17 contient les deux plans engendrés par les quatre premiers opérateurs principaux. En parallèle avec l'analyse des graphiques, le tableau 5.3 contient les moyennes des normes de Hilbert-Schmidt des opérateurs de chaque classe de mouvement. Les trois plus grandes normes calculées sont obtenues pour la Marche, mise en valeur par le premier axe sur le graphique, puis les mouvements Chute Allongée et la Chute Repliée caractérisés par le second axe. Comme il a été vu dans la section consacrée à l'ACP, les variables-capteurs du mouvement Marche ont une variance globalement forte et sont également très corrélées entre elles. Son opérateur a donc nécessairement une norme élevée, d'après la remarque qui a été faite plus haut. Par ailleurs, il en est de même pour la Chute Allongée et Repliée, mais de manière moins significative comme on peut le constater sur le tableau. Le second plan discrimine la moitié des mouvements de la classe Jeté divisée en quatre variantes différentes : Jeté devant, Jeté derrière et Jeté de côté. Les instances discriminées sont les Jetés de côté ayant une norme plus élevée.

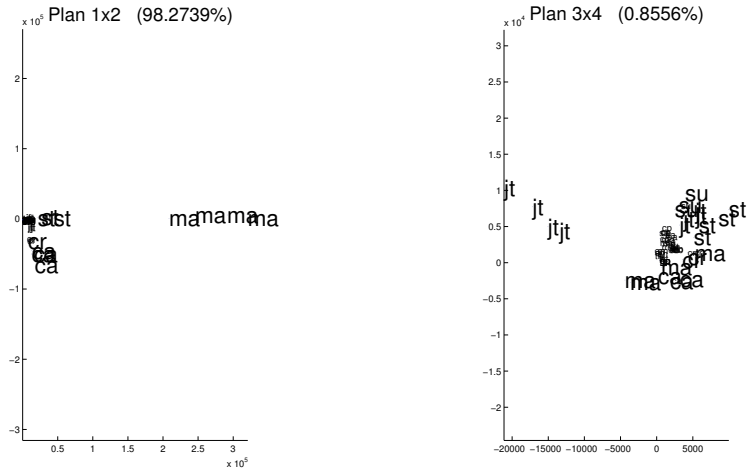

FIG. 5.17 – Représentations graphiques des deux premiers plans principaux d'une analyse d'opérateurs non normés pour $\theta = max$.

Mouvement	ma	en	cp	et	eb	gp	mg
Norme moyenne	2.6688	0.0558	0.0833	0.0623	0.0729	0.0154	0.0876
Mouvement	sa	cr	ca	jt	tr	su	st
Norme moyenne	0.0737	0.3307	0.5032	0.1890	0.0445	0.1117	0.3328

TAB. 5.3 – Moyenne des normes de Hilbert-Schmidt des instances de chaque classe de mouvement pour les 60 mouvements et $\theta = max$ (échelle : 10^5).

Analyse d'opérateurs de covariance normés

Nous passons maintenant à l'étude du second cas : analyse d'opérateurs normés de covariance. En figure 5.18 sont présentés les quatre premiers plans principaux de l'analyse d'opérateurs. Les huit premiers opérateurs principaux apportent plus de 86% de l'inertie totale dont le premier en représente approximativement 32% (fig. 5.16 (b)). Ainsi, au vue des pourcentages d'information, ces premiers axes nous permettent d'extraire les principales caractéristiques des mouvements et donc de pouvoir les différencier. Le compromis O_1 est l'opérateur "moyen" qui caractérise les informations communes à tous les mouvements. Il ne permet pas de discriminer les mouvements autrement que par leur norme et ne constitue donc pas a priori un bon critère de différenciation. Par exemple, le compromis peut être relié à des caractéristiques d'ordre morphologique communes à tous les danseurs, comme par exemple le "squelette" décrit par la disposition des capteurs ou les distances mutuelles entre certains d'entre eux qui restent fixent au

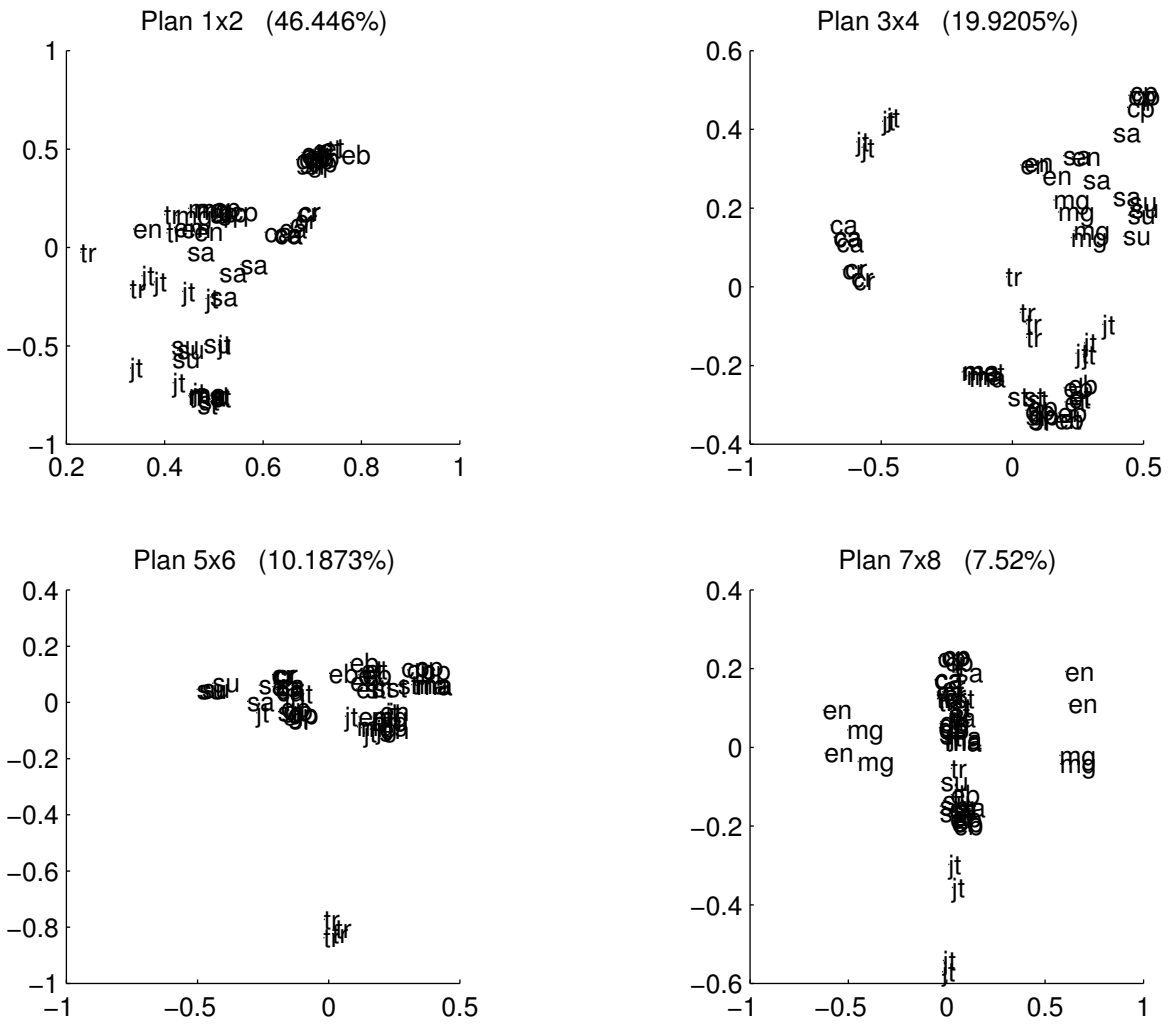

FIG. 5.18 – Représentations graphiques des quatre premiers plans principaux d'une analyse d'opérateurs normés pour $\theta = max$.

cours du temps (pied/jambe, poignet/main ...).

En ce qui concerne O_2, nous pourrions l'interpréter comme l'axe qui caractérise le "sens" du mouvement. Les éléments se situant en haut de l'axe (Enroulé bassin, Enroulé tête ou encore Grand plié) correspondent aux mouvements verticaux et ceux se situant en bas (Marche ou Saut temps levé) sont les mouvements à dominante horizontale. Considérons par exemple les trois mouvements types suivants : la Marche, le Saut attitude et l'Enroulé bassin. La Marche, qui est un mouvement plutôt horizontal, se situe en bas du graphique. L'Enroulé Bassin, qui lui est vertical se trouve en haut de l'axe. Enfin, le Saut Attitude est un mouvement intermédiaire aux deux autres et se situe au milieu, proche de l'origine. Ce mouvement consiste à exécuter un saut (sens

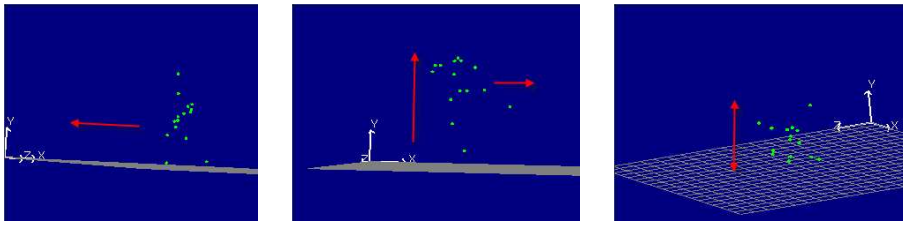

FIG. 5.19 – De gauche à droite, capture d'une Marche (mouvement à dominante horizontale) ; capture d'un Saut attitude (mouvement intermédiaire) ; capture d'un Enroulé bassin (mouvement à dominante verticale).

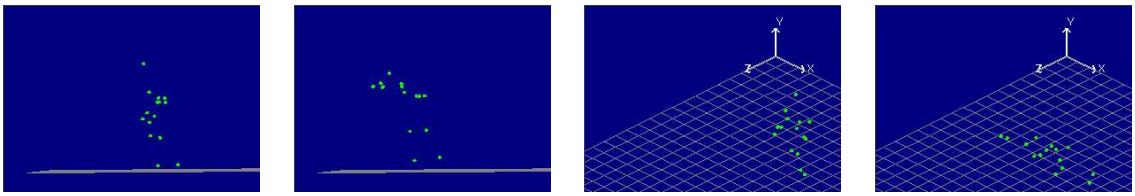

FIG. 5.20 – De gauche à droite, Suspension : position initiale et finale (déplacement faible du centre de gravité) ; Chute allongée : position initiale et finale (déplacement fort du centre de gravité).

vertical) en levant une jambe à l'horizontal (sens horizontal). Conscient que le lecteur n'est pas nécessairement expert en chorégraphie, la figure 5.19 illustre le propos en donnant une capture pour chacune des instances.

Pour O_3, les mouvements qui se situent à droite par rapport à l'origine de cet axe sont les mouvements où la position relative au sol du centre de gravité du danseur (qui se situe près du pelvis) varie peu. À l'opposé, les mouvements se situant plutôt à gauche sont ceux où il y a une grande mobilité du centre de gravité par rapport au sol. La figure 5.20 donne les captures initiales et finales de deux instances se trouvant à l'extrémité de cet axes : Suspension et Chute allongée. En visionnant les deux captures, on vérifie bien que le déplacement du centre de gravité de la première instance est relativement faible, par opposition au mouvement Chute allongée. Le plan principal 5×6, et plus précisément l'axe 5, discrimine clairement la classe Tour, *i.e.* les mouvements présentant une dominante d'exécution circulaire. À ce titre, la plupart des variables-capteurs pour ce type de mouvement ne présentent pas de liaison linéaire, donnant ainsi une forme particulière de matrice de covariance. Cela explique le fait que ce mouvement est isolé par au moins un des axes. Enfin, le plan principal 7×8 divise en deux parties les classes de mouvement Enveloppé et Marche glissée. À l'instar de la classe Jeté, on discerne en effet deux formes d'exécution : à droite et à gauche.

Analyse d'opérateurs de covariance relationnelle normés

Avec une analyse relationnelle ($\theta = 10$), les projections des opérateurs sont présentées à la figure 5.21. Les parts d'inertie sont distribuées de manière plus homogène sur les plans principaux et l'on remarque que les interprétations des axes données précédemment sont globalement les mêmes. En revanche, et contrairement à l'analyse globale des opérateurs, l'usage du laplacien implique une plus grande disparité des projections sur les plans principaux. En effet, la signature de la covariance relationnelle ne prend pas en compte les périodes d'immobilité du danseur, et ce phénomène peut être similaire pour deux mouvements n'appartenant pas à la même classe. De plus, il est clair que la covariance relationnelle discrimine deux mêmes mouvements exécutés avec des vitesses différentes, contrairement à la covariance. Ce phénomène est dû à l'emploi d'une fenêtre temporelle de taille réduite. Avec une fréquence d'échantillonnage identique pour deux mêmes mouvements exécutés à deux vitesses différentes, on peut remarquer que deux individus du mouvement plus lent seront séparés de plus de dates de capture. Ainsi, les degrés des observations du mouvement lent seront donc significativement plus faibles avec un nombre total d'individus plus important. La mise en valeur de ces caractéristiques supplémentaires par la covariance relationnelle explique que les classes sont plus étendues et ont donc tendance à se chevaucher davantage.

Analyse discriminante d'opérateurs

Les calculs ont montré que les quinze premiers axes d'une analyse factorielle d'opérateurs apportent approximativement 95% de l'inertie que cela soit pour l'échantillon de 60 instances (représentations graphiques) ou pour la totalité des mouvements. Nous fixons donc $m = 15$ le nombre de premières projections pour appliquer l'analyse discriminante.

Les graphiques des valeurs propres pour les trois cas déjà mentionnés sont montrés à la figure 5.22. Les valeurs propres sont à chaque fois au nombre de 13 puisqu'il y a 14 classes de mouvements a priori. On constate une homogénéité des pouvoirs discriminants des premiers axes pour chacun des cas, impliquant une disposition sphérique du nuage des centres de gravité des classes. À la figure 5.23, le graphique de gauche est la projection d'une analyse d'opérateurs normés ($\theta = 10$). Tous les mouvements ont été regroupés suivant leur classe (courbes sur le graphique). Le graphique de droite montre le plan engendré par les deux premiers opérateurs discriminants partiels. Tous les mouvements sont maintenant clairement distingués et les distances qui contribuent à séparer les groupes sont d'autant plus renforcées. L'interprétation de chacun des axes

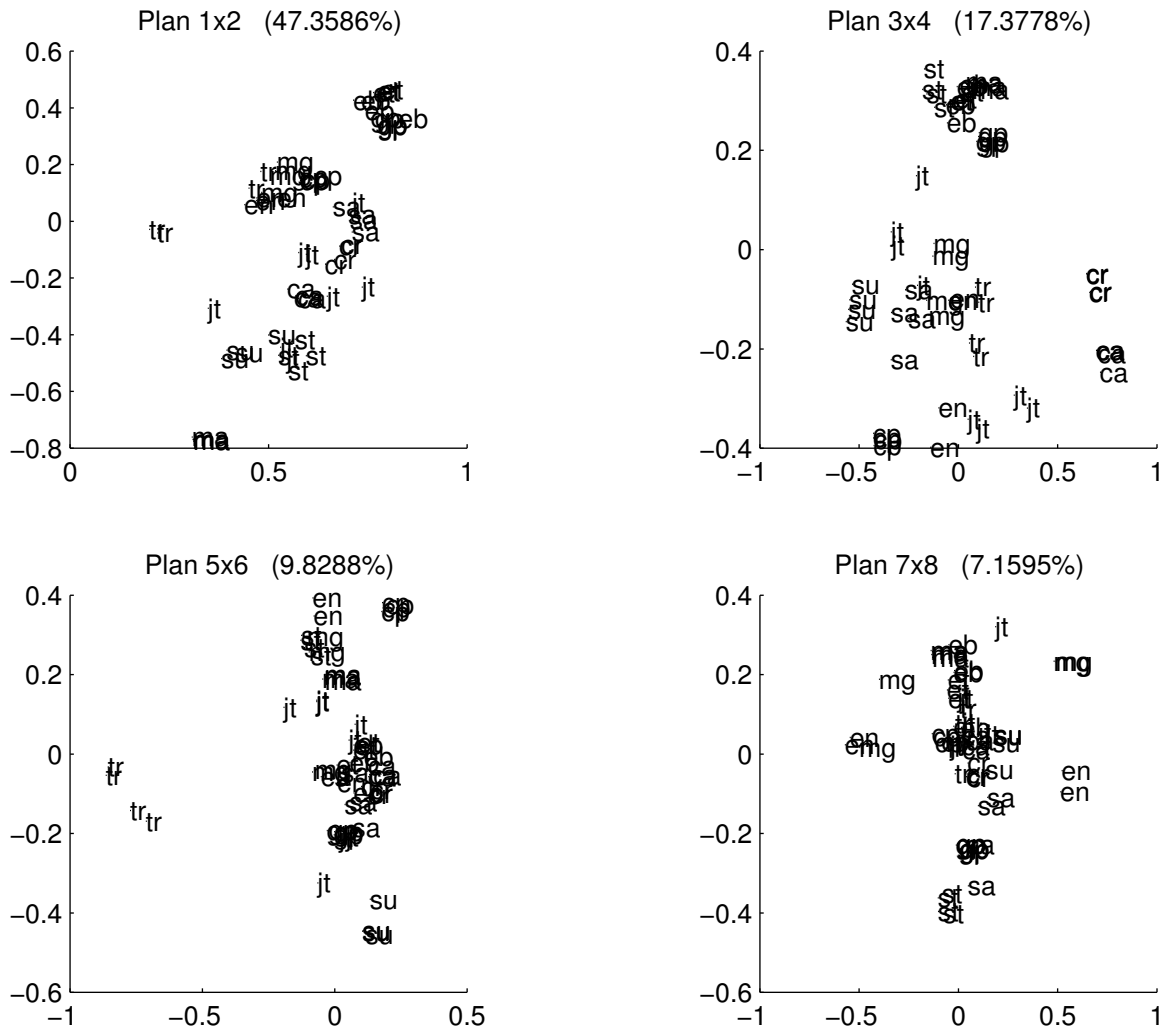

FIG. 5.21 – Représentations graphiques des quatre premiers plans principaux d'une analyse d'opérateurs normés pour $\theta = 10$.

est plus délicate et l'on peut remarquer que des classes homogènes telles que Chute repliée/allongée et Enroulé tête/bassin sont mutuellement proches sur le plan discriminant.

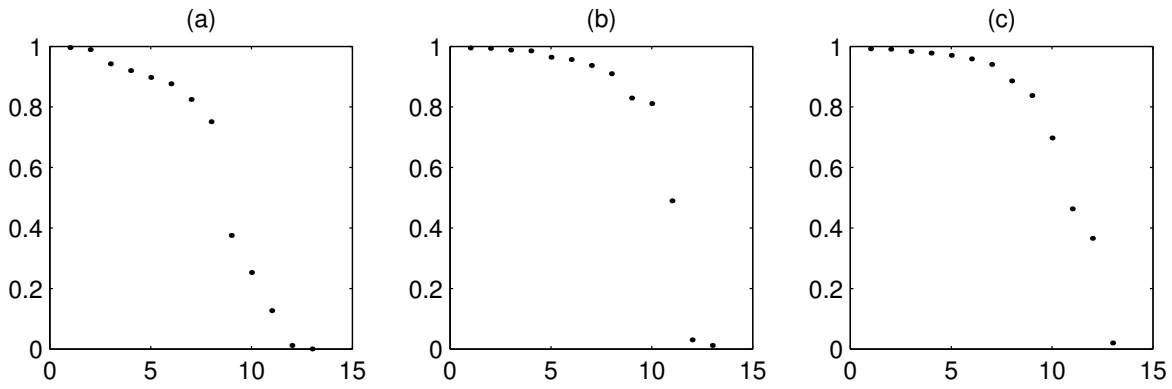

FIG. 5.22 – Valeurs propres d'une analyse discriminante d'opérateurs ; (a) Opérateurs non normés et $\theta = max$; (b) Opérateurs normés et $\theta = max$; (c) Opérateurs normés et $\theta = 10$.

5.4.3 Classification supervisée

Procédure de validation

Pour la classification supervisée, nous proposons de valider les méthodes présentées en calculant les pourcentages de bonne reconnaissance par la distance au plus proche centre de la classe (leave-one-out), qui est en fait un cas particulier de l'approche *validation croisée* [JDM00]. Parmi les méthodes d'estimation du taux d'erreur, telles que resubstitution, holdout ou bootstrap, la procédure leave-one out fournit une estimation du taux d'erreur à la fois non optimiste et non pessimiste, c'est à dire qu'il n'y a ni sous-estimation, ni sur-estimation. Notre choix est donc motivé par le fait que l'estimation du taux d'erreur est asymptotiquement non biaisée [JDM00]. En contre partie, les inconvénients sont d'une part la variance intrinsèquement élevée de l'estimateur et d'autre part les calculs coûteux de la procédure.

Quelque soit la méthode utilisée, on est amené à choisir initialement deux sous-ensembles de la population afin d'appliquer la procédure : un échantillon d'apprentissage et un échantillon de test. L'échantillon d'apprentissage permet de créer les classes d'appartenance par l'intermédiaire de l'algorithme testé. Avec l'échantillon de test, on calcule une estimation du taux d'erreur en procédant au classement via l'échantillon d'apprentissage.

La première étape de la procédure leave-one out est de prélever un élément dans la population de départ qui va constituer l'échantillon de test. La population privée

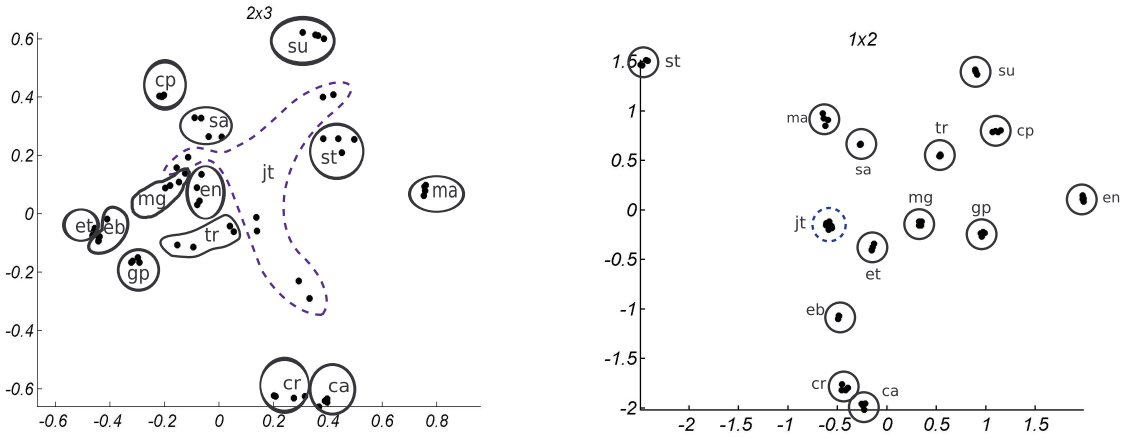

FIG. 5.23 – À gauche : plan 2×3 d'une analyse d'opérateurs normés ($\theta = 10$, inertie : 24%) ; à droite : plan 1×2 d'une analyse discriminante d'opérateurs normés ($\theta = 10$, inertie : 16%).

de cet élément correspondant donc quant à elle à l'échantillon d'apprentissage. L'approche est supervisée car on dispose a priori des étiquettes de chaque élément. Après pré-traitements éventuels de l'élément test, celui-ci est comparé à l'échantillon d'apprentissage. Le critère de classement est de déterminer si l'élément test et le centre de gravité d'une classe le plus proche (en terme de distance euclidienne dans l'espace de projection) sont issus du même label. La procédure est ensuite réitérée en prélevant un autre élément test. On calcule ensuite le taux de classement en sommant le nombre d'élément reconnu que l'on divise par la taille de la population. Avec les mêmes notations qu'au troisième chapitre, les encadrés 5.4 et 5.5 donnent les différentes étapes de la procédure leave-one out pour l'analyse d'opérateurs.

Résultats expérimentaux

Les résultats détaillés des taux de classification par validation croisée, pour les cas analyse d'opérateurs non normés, normés et analyse discriminante d'opérateurs normés, sont présentés aux tableaux 5.6, 5.7 et 5.8. Pour chacun d'eux est mentionné l'effectif des instances reconnues ou non ainsi que les taux de reconnaissance globaux, pour $\theta \in \{1, 2, 5, 10, max\}$. Nous conservons le choix de $m = 15$ premiers axes pour les projections.

Le tableau 5.6 (opérateurs non normés) donnent des taux de bonne reconnaissance plus faibles que les autres cas. Les taux augmentent les liaisons pour le graphe et les

Leave-one out pour l'analyse d'opérateurs de covariance relationnelle

Entrées :
- Tableau $\mathfrak{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_N)$.
- Graphes de relation binaire G_k de chaque sous-tableau $\mathbb{X}_k$.
- Métriques M_k sur les individus de $\mathbb{X}_k$.
- Nombre m de premières projections.

Sortie :
- Taux de bonne reconnaissance *taux*.

Procédure :
$\tau \leftarrow 0$
Pour k allant de 1 à N,
ARO sur $\mathfrak{X}$ *privé de* $\mathbb{X}_k$:
$\mathfrak{X}^k = (\mathbb{X}_1, \ldots, \mathbb{X}_{k-1}, \mathbb{X}_{k+1}, \ldots, \mathbb{X}_N)$
$[\mathcal{U}^k, \mathcal{L}^k, \Gamma^k, O^k] = ARO(\mathfrak{X}^k)$

Projection v *(vecteur colonne) de l'opérateur* Γ_k *sur les* m *premiers axes* :
Pour i allant de 1 à m, $v(i) = trace(\Gamma_k O_i^k)$; Fin Pour

Calcul des centres $\mathcal{G}^k$ *en déterminant les moyennes des lignes de* $\mathcal{U}^k(\mathcal{L}^k)^{1/2}$.

Calcul de l'indice du minimum de la distance euclidienne usuelle :
$ind = \text{argmax}_{j=1,\ldots,K-1} \|{}^t v - \mathcal{G}_m^k(j)\|$
où $\mathcal{G}_m^k(j)$ *est le* j*ème centre projeté sur les* m *premiers axes.*

Mise à jour de la variable τ :
Si $Class(\mathbb{X}_k) = Class(\mathbb{X}_{ind})$, alors $\tau \leftarrow \tau + 1$; Fin Si
Fin Pour

$taux = \tau / N$

TAB. 5.4 – Leave-one out pour l'analyse d'opérateurs de covariance relationnelle.

Leave-one out pour l'analyse discriminante d'opérateurs

Entrées :
- Tableau $\mathfrak{X} = (\mathbb{X}_1, \ldots, \mathbb{X}_N)$.
- Graphes de relation binaire G_k de chaque sous-tableau $\mathbb{X}_k$.
- Métriques M_k sur les individus de $\mathbb{X}_k$.
- Nombre m de premières projections.
- Métrique $\mathcal{P}$ de la fréquence des classes.

Sortie :
- Taux de bonne reconnaissance *taux*.

Procédure :
$\tau \leftarrow 0$
Pour k allant de 1 à N,
ADO sur $\mathfrak{X}$ privé de $\mathbb{X}_k$:
$\mathfrak{X}^k = (\mathbb{X}_1, \ldots, \mathbb{X}_{k-1}, \mathbb{X}_{k+1}, \ldots, \mathbb{X}_N)$
$[\mathfrak{U}^k, \mathfrak{L}^k, \mathcal{D}^k, \mathcal{V}^k, \Gamma^k, O^k] = ARDO(\mathfrak{X}^k, m)$

Projection v (vecteur colonne) de l'opérateur Γ_k sur les m premiers axes :
Pour i allant de 1 à m, $v(i) = trace(\Gamma_k O_i^k)$; Fin Pour

Mise à jour de $\mathcal{P}$ et centrage de v pour la métrique $\mathcal{P}$:
$\mathcal{P}^* = \mathcal{P}$; $\mathcal{P}^*(k,k) = \mathcal{P}(k,k) - \frac{1}{|class_k|}$
$v = v - {}^t\mathbf{1}\mathcal{P}^* v$

Projection projv (vecteur colonne) de v sur les axes discriminants :
$projv = {}^t v(\mathcal{V}^k)^{-1}\mathfrak{U}^k$

Calcul des centres $\mathcal{G}^k$ en déterminant les moyennes des lignes de $\mathfrak{U}^k(\mathfrak{L}^k)^{1/2}$.

Calcul de l'indice du minimum de la distance euclidienne usuelle :
$ind = \text{argmax}_{j=1,\ldots,K-1} \|projv - \mathcal{G}_m^k(j)\|$
où $\mathcal{G}_m^k(j)$ est le $j$ème centre projeté sur les m premiers axes.

Mise à jour de la variable τ :
Si $Class(\mathbb{X}_k) = Class(\mathbb{X}_{ind})$, alors $\tau \leftarrow \tau + 1$; Fin Si
Fin Pour
$taux = \tau / N$

TAB. 5.5 – Leave-one out pour l'analyse discriminante d'opérateurs.

meilleurs sont donc obtenus pour la covariance. Les mouvements ne sont discriminés que par la valeur de leur norme. Ainsi, les mouvements de la classe Marche, Chute allongée et repliée, qui fournissent la plus grande norme sont presque tous bien classés. Les mouvements bien classés et présentant une norme faible, comme les mouvements Enroulé bassin, font partie d'une classe de mouvements dont les éléments sont géométriquement proches, *i.e.* que la moyenne des normes de la classe est représentative de toutes les normes de ses éléments. Ainsi, un mouvement omis de la classe (principe de la validation croisée au plus proche centre de gravité) ne perturbe que très peu sa structure.

En accord avec les représentations graphiques de la section précédente, les résultats du tableau 5.7 (opérateurs normés) sont meilleurs que le cas des opérateurs non normés. Les mouvements suivants sont quasiment tous bien classés pour toutes les valeurs de θ : Marche, Chute, Grand plié, Saut attitude et Suspension. Les instances de la classe réduite Marche glissée (seulement huit éléments) possédant deux variantes d'exécution sont à moitié bien classés. Les taux globaux pour les différentes valeurs de θ sont presque stables. La covariance relationnelle ne prend pas en compte les périodes statistiques mais discrimine par contre les différentes vitesses d'exécution (niveau de maîtrise du danseur). Par conséquent, elle n'apporte pas de meilleurs résultats en terme de classification puisqu'elle agit en "compromis" sur les opérateurs en mettant en évidence de nouvelles caractéristiques (vitesse) tout en en supprimant d'autres (immobilité). Il en est de même pour le tableau 5.8 (analyse discriminante, opérateurs normés) qui donne les meilleurs taux de bonne reconnaissance.

5.5 Modèle de Wishart

En complémentarité avec les approches liées à la statistique géométrique, nous présentons les résultats d'une classification fondée sur les approches probabilistes de type loi de Wishart. La section qui suit décrit comment se ramener au modèle théorique et utiliser la variable exogène pour supprimer la tendance locale des variables aléatoires. Nous présentons ensuite le choix des paramètres et des niveaux de lissage et analysons les résultats de la classification supervisée ou non de mouvements par la loi de Wishart.

θ	1		2		5		10		max	
Reconnu ?	oui	non	oui	non	oui	non	oui	non	oui	non
Marche	17	0	17	0	17	0	17	0	16	1
Enveloppé	10	3	10	3	11	2	11	2	11	2
Chute pliée	17	3	17	3	17	3	16	4	20	0
Enroulé tête	5	10	5	10	5	10	5	10	4	11
Enroulé bassin	15	6	15	6	16	5	17	4	21	0
Grand plié	7	1	6	2	6	2	7	1	8	0
Marche glissée	3	5	3	5	3	5	3	5	0	8
Saut attitude	10	2	10	2	11	1	12	0	12	0
Chute repliée	15	3	15	3	15	3	15	3	18	0
Chute allongée	11	11	11	11	10	12	10	12	22	0
Jeté	15	10	15	10	15	10	17	8	18	7
Tour	20	2	20	2	21	1	22	0	15	7
Suspension	12	0	12	0	12	0	12	0	12	0
Saut tps levé	6	0	6	0	6	0	6	0	4	2
Effectifs	163	56	162	57	165	54	170	49	181	38
Taux (%)	74.43	25.57	73.97	26.03	75.34	24.66	77.63	22.37	82.65	17.35

TAB. 5.6 – Analyse d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs non normés ($m = 15$ premières projections et $C = 14$ classes).

5.5.1 Suppression de la tendance locale : interprétations

Afin d'admettre l'hypothèse de gaussianité, le graphe de la formule 5.3 permet accessoirement de supprimer la tendance locale de chaque variable aléatoire. En effet, si l'on suppose que les individus sont équipondérés, l'espérance relationnelle correspondante joue le rôle de moyenne mobile et la tendance est supprimée en retranchant localement de la variable les moyennes locales obtenues. La variable θ correspond au nombre de réalisations pour lesquelles on calcule une moyenne locale. Par conséquent, plus θ est faible, plus la variable perd de sa structure et le résidu obtenu $X - E(X|g_\theta)$ est erratique. Si θ est maximum, cela revient à centrer la variable puisqu'on lui retire sa moyenne globale. La figure 5.24 en donne une illustration où la variable considérée possède 300 réalisations, pour $\theta = 10$ et $\theta = 300$. Afin d'utiliser le modèle dans le cadre d'une classification, on admettra l'hypothèse selon laquelle l'échantillon obtenu en supprimant la tendance locale d'une variable est issu d'une variable aléatoire gaussienne. Notons que l'hypothèse de gaussianité des variables devrait être admise de manière plus rigoureuse à l'aide de tests d'hypothèses adéquates tels que les tests de normalité fondés sur les extensions multidimensionnelles des coefficients skewness et kurtosis [Mar80] ou plus récemment du test du χ^2 [EB04, EG04]. Enfin, nous considérons dans le modèle

θ	1		2		5		10		max	
Reconnu ?	oui	non	oui	non	oui	non	oui	non	oui	non
Marche	17	0	17	0	17	0	17	0	17	0
Enveloppé	11	2	11	2	11	2	11	2	11	2
Chute pliée	20	0	20	0	20	0	20	0	20	0
Enroulé tête	14	1	14	1	14	1	14	1	14	1
Enroulé bassin	16	5	16	5	16	5	16	5	16	5
Grand plié	8	0	8	0	8	0	8	0	8	0
Marche glissée	5	3	4	4	4	4	4	4	4	4
Saut attitude	12	0	12	0	12	0	12	0	12	0
Chute repliée	17	1	18	0	18	0	18	0	18	0
Chute allongée	22	0	22	0	22	0	22	0	22	0
Jeté	20	5	20	5	20	5	17	8	24	1
Tour	21	1	21	1	21	1	21	1	18	4
Suspension	12	0	12	0	12	0	12	0	12	0
Saut tps levé	6	0	6	0	6	0	6	0	4	2
Effectifs	201	18	201	18	201	18	198	21	200	19
Taux (%)	91.78	8.22	91.78	8.22	91.78	8.22	90.41	9.59	91.32	8.68

TAB. 5.7 – Analyse d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs normés ($m = 15$ premières projections et $C = 14$ classes).

des lois de Wishart centrées puisque l'espérance du résidu obtenu est très proche de l'origine.

La figure 5.25 montre une série temporelle avec son espérance relationnelle tracée en pointillé ($\Omega = [1, 100]$ et $\theta = 10$). On constate que la variable aléatoire coïncide avec son espérance relationnelle sur les intervalles [0,20], [40,60] et [80,100], c'est à dire sur les segments où la variable varie linéairement.

5.5.2 Choix des paramètres

Bien entendu, nous conservons la même base de données afin de pouvoir comparer les pourcentages de reconnaissance. Le niveau de lissage des variables, qui par abus de notation s'écrira aussi θ, est fixé aux mêmes valeurs : $\theta \in \{1, 2, 5, 10, max\}$. On rappelle que pour $\theta = 1$, le résidu obtenu correspond à la plus grande perte d'information, par opposition à $\theta = max$ où les variables sont simplement centrées. Également, la métrique du laplacien n'intervient pas pour le calcul des signatures afin de rester conforme au modèle.

θ	1		2		5		10		max	
Reconnu ?	oui	non	oui	non	oui	non	oui	non	oui	non
Marche	17	0	17	0	17	0	17	0	17	0
Enveloppé	11	2	11	2	11	2	11	2	11	2
Chute pliée	20	0	20	0	20	0	20	0	20	0
Enroulé tête	14	1	14	1	14	1	14	1	14	1
Enroulé bassin	16	5	16	5	16	5	16	5	16	5
Grand plié	8	0	8	0	8	0	8	0	8	0
Marche glissée	7	1	7	1	6	2	4	4	4	4
Saut attitude	12	0	12	0	12	0	12	0	12	0
Chute repliée	16	2	17	1	18	0	18	0	18	0
Chute allongée	22	0	22	0	22	0	22	0	22	0
Jeté	22	3	22	3	22	3	24	1	24	1
Tour	22	0	22	0	22	0	22	0	22	0
Suspension	12	0	12	0	12	0	12	0	12	0
Saut tps levé	6	0	6	0	6	0	6	0	6	0
Effectifs	205	14	206	13	206	13	206	13	206	13
Taux (%)	93.61	6.39	94.06	5.94	94.06	5.94	94.06	9.59	94.06	5.94

TAB. 5.8 – Analyse discriminante d'opérateurs de covariance : taux de reconnaissance par leave-one out sur les opérateurs normés ($m = 15$ premières projections et $C = 14$ classes).

Implémentée tel quel, la formule (4.38) donnant les estimés de z_{ik} peut amener à des formes indéterminées, en particulier à cause des déterminants mis à la puissance n_i, ou l'exponentielle d'une trace de matrice à grande dimension. Pour y remédier, la ré-estimation des z_{ik} peut se simplifier après avoir éliminé les constantes communes et mis tous les termes dans l'exponentielle. La nouvelle expression donnant les z_{ik} est la suivante :

$$z_{ik} = \frac{\exp\left\{\log(\pi_k) - \frac{n_i}{2}\left(\log|\Sigma_k| + trace\left(\Sigma_k^{-1}\left(\frac{1}{n_{ik}}{}^t\varepsilon_{ik}\varepsilon_{ik}\right)\right)\right)\right\}}{\sum_{l=1}^{K}\exp\left\{\log(\pi_l) - \frac{n_i}{2}\left(\log|\Sigma_l| + trace\left(\Sigma_l^{-1}\left(\frac{1}{n_{ik}}{}^t\varepsilon_{ik}\varepsilon_{ik}\right)\right)\right)\right\}} \tag{5.7}$$

Cela permet de compenser les valeurs "extrêmes" et d'obtenir des résultats numériques utilisables dans la suite de la procédure. Le passage au logarithme des poids des clusters implique des valeurs non nulles, autrement dit des clusters non vides qui peut donc amener à des erreurs numériques potentielles. Pour y palier, nous vérifions à chaque itération la présence de clusters vides et nous les supprimons de la procédure le cas échéant.

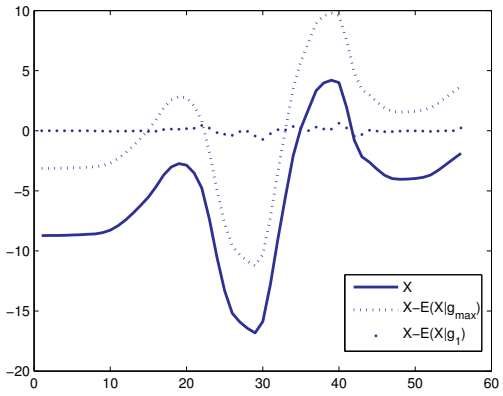

FIG. 5.24 – Suppression de la tendance locale d'une variable à 56 réalisations, pour $\theta = 1$ et $\theta = max$.

5.5.3 Classification supervisée

Pour l'ensemble des résultats présentés, nous avons considéré deux cas de figures. Le premier consiste à sélectionner un nombre de classes *a priori* égal au nombre de mouvements type ($C = 14$). Le second prend en compte les différentes variantes possibles pour certains types de mouvements. Entres autres, nous disposons de tours sautés ou non sautés, de marches orientées différemment ... Nous avons déjà noté que la matrice de covariance des variables-capteurs n'est pas robuste sur les différentes variantes d'exécution des mouvements-type. Nous comptabilisons cinq variantes possibles, qui s'additionnent aux 14 classes initiales, soit $C = 19$ classes.

Le but de la classification supervisée est de calculer le taux de reconnaissance en fonction du niveau de lissage et de déterminer dans quelle mesure se manifeste la perte d'information due à la suppression de la tendance. L'encadré 5.9 donne une synthèse de la procédure leave-one out pour la méthode de classification supervisée par la loi de Wishart. Le tableau 5.10 présente les taux de reconnaissance en pourcentage selon différentes tailles de la fenêtre de lissage. La procédure choisie est la validation croisée avec le plus proche voisin.

Le pourcentage de bonne reconnaissance augmente globalement en fonction de la taille de la fenêtre. Les taux vont de 76.26% ($C = 19$, 167 mouvements bien classés) jusqu'à 97.71% ($C = 14$, 214 mouvements bien classés). Le tableau 5.11 présente en plus les résultats détaillés de reconnaissance pour les 14 classes de référence. Les taux sont nettement meilleurs qu'avec les approches précédentes. On vérifie ainsi une certaine robustesse de la méthode, car les résultats sont satisfaisants dans des conditions

Leave-one out pour la classification supervisée par la loi de Wishart

Entrées :
- Nombre C de classes.
- Opérateurs $(\Gamma_i)_{i=1}^N$ répartis a priori dans les C classes.
- Taille $(n_i)_{i=1}^N$ des sous-tableaux sans tendance correspondants.

Sortie :
- Taux de bonne reconnaissance *taux*.

Procédure :
$\tau \leftarrow 0$
Pour k allant de 1 à N,
Classification supervisée par la loi de Wishart sur les opérateurs Γ_i privés de Γ_k.

Détermination de la classe c_k de Γ_k :

$$c_k = \underset{i\in\{1,\dots,n\}\setminus\{k\}}{\operatorname{argmax}} \left(\log f({}^t\varepsilon_k\varepsilon_k; \widehat{\Sigma}_i, n_k)\right)$$

Mise à jour de la variable τ :
Si $Class(\varepsilon_k) = c_k$, alors $\tau \leftarrow \tau + 1$; Fin Si
Fin Pour

$taux = \tau/N$

TAB. 5.9 – Leave-one out pour la classification supervisée par la loi de Wishart.

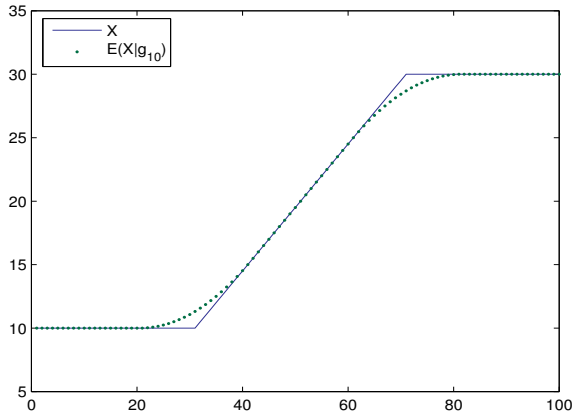

FIG. 5.25 – Tracés d'une variable et de son espérance relationnelle ($\theta = 10$).

θ	1	2	5	10	max
$C = 14$	77.17	84.93	93.60	94.06	97.71
$C = 19$	76.26	80.82	90.41	89.95	92.69

TAB. 5.10 – Classification supervisée en supprimant la tendance en fonction du niveau de lissage.

s'éloignant des hypothèses du modèle théorique pour les fortes valeurs du paramètre de lissage θ (stationnarité et modèle gaussien). Pour $\theta = 1$, les variables résiduelles obtenues correspondent à la plus grande perte d'information propre au mouvement. À l'inverse, les variables centrées ($\theta = max$) fournissent toute l'information apportée par le mouvement, jouant ainsi un rôle d'autant plus discriminatoire pour les classes. Lorsque l'on retire la tendance, on accède à d'autres critères caractéristiques de typologies moins évidentes que celles liées au côté spatial. Cela peut être lié soit à des aspects proches du bruit comme les erreurs de mesure ou l'imprécision des capteurs, soit à la morphologie du danseur, mise en valeur une fois que l'on a supprimé l'information propre à l'exécution. Globalement, les pourcentages sont plus faibles pour $C = 19$. Ce phénomène s'explique par un inconvénient de la procédure leave-one out, qui agit sur des classes dont certaines ont été subdivisées en deux sous-classes. Le fait de retirer un élément à une classe d'effectif faible risque de modifier fortement la nature de la classe en question.

Lissage θ	1		2		5		10		max	
Reconnu ?	oui	non	oui	non	oui	non	oui	non	oui	non
Marche	16	1	16	1	16	1	16	1	16	1
Enveloppé	10	3	10	3	11	2	11	2	11	2
Chute pliée	16	4	17	3	18	2	11	2	11	2
Enroulé tête	4	11	3	12	13	2	12	3	14	1
Enroulé bassin	8	13	17	4	18	3	18	3	21	0
Grand plié	8	0	8	0	8	0	8	0	8	0
Marche glissée	8	0	8	0	7	1	7	1	8	0
Saut attitude	10	2	12	0	12	0	12	0	12	0
Chute repliée	17	1	17	1	17	1	18	0	18	0
Chute allongée	20	2	21	1	22	0	22	0	22	0
Jeté	19	6	22	3	25	0	25	0	25	0
Tour	20	2	22	0	22	0	22	0	22	0
Suspension	9	3	9	3	11	1	11	1	11	1
Saut tps levé	4	2	4	2	5	1	6	0	6	0
Effectifs	169	50	186	33	205	14	206	13	214	5
Taux (%)	77.17	22.83	84.93	15.07	93.61	6.39	94.06	5.94	97.72	2.28

TAB. 5.11 – Classification supervisée avec modèle de Wishart : taux de reconnaissance par leave-one out en fonction di niveau de lissage θ ($C = 14$ classes).

5.5.4 Classification non supervisée

Procédure de validation

Pour la classification non supervisée (Algorithme EM), nous n'estimons pas de taux de reconnaissance puisque les clusters n'ont pas d'étiquetage a priori. Nous proposons de comparer les classements obtenus de la manière suivante. Supposons P et Q deux graphes de partition, *i.e.* $P(i,j) = 1$ si les opérateurs Γ_i et Γ_j sont dans le même cluster, et 0 sinon (idem pour Q). On adopte l'indice de cohérence suivant :

$$\mathcal{I}(P,Q) = \frac{1}{N^2} \sum_{i,j=1}^{N} (P(i,j)Q(i,j) + (1 - P(i,j))(1 - Q(i,j))) \tag{5.8}$$

La quantité $D(P,Q) = 1 - \mathcal{I}(P,Q) \in [0,1]$ correspond à une différence symétrique et s'interprète comme une distance entre P et Q. Plus D est proche de 0, plus les partitions sont homogènes. Ainsi, pour chaque estimation, nous calculons la distance entre la partition obtenue et la partition de référence des opérateurs connue a priori.

Résultats expérimentaux

La classification non supervisée extrait des sous-mouvements, issus des éventuelles variantes possibles pour un même mouvement type. Le tableau 5.12 contient les moyennes et écarts-types des distances entre partitions après 50 itérations pour chacun des cas, en fonction des valeurs de θ, de C et de K. Les moyennes diminuent globalement suivant

θ	$C = K = 14$	$C = 14, K = 19$	$C = K = 19$
max	0.101 ± 0.014	0.087 ± 0.013	0.079 ± 0.013
10	0.093 ± 0.010	0.085 ± 0.008	0.075 ± 0.007
5	0.097 ± 0.010	0.088 ± 0.009	0.081 ± 0.009
2	0.105 ± 0.013	0.098 ± 0.008	0.092 ± 0.009
1	0.110 ± 0.011	0.101 ± 0.009	0.094 ± 0.009

TAB. 5.12 – Classification non supervisée : moyennes et écarts-types des différences symétriques en fonction de K et de C (50 itérations).

que le nombre de classes ou de clusters augmente. Les écarts-types montrent que l'on se heurte quasi-systématiquement à un minimum local. Dans les 3 cas considérés, on constate une distance et un écart-type faibles pour $\theta = 10$, les moyennes et écarts-types

étant plus élevés pour $\theta = max$. En effet, les covariances analysées avec la tendance tiennent compte des périodes d'immobilité du danseur, ce qui n'est pas le cas si l'on supprime la tendance. Il y a donc des confusions possibles pour deux mouvements différents ayant des périodes statiques. Pour $\theta = 10$, il s'agit d'un bon compromis où l'immobilité est suffisamment négligée tout en conservant la structure du mouvement.

Pour le cas $\theta = 10$ qui correspond au meilleur résultat obtenu, la courbe du critère BIC a été tracée à la figure 5.26. Contrairement à l'exemple étudié au chapitre précédent, il n'y a pas de minimum global de la courbe : la courbe $-2\log L(K)$ est toujours en-dessous de la fonction BIC. De plus, la courbe présente des irrégularités suite à l'instabilité des résultats d'estimation (écarts-types non nuls du tableau 5.12). Ces phénomènes mettent clairement en évidence les problèmes d'initialisation et de dimensionnalité (phénomène de Hugues) déjà évoqués. En effet, par symétrie, nous avons $\frac{45\times46}{2} = 1035$ ddl pour chaque signature ce qui, comparé à seulement une base de 219 mouvements, perturbe l'algorithme d'estimation et empêche d'établir un nombre optimal de clusters.

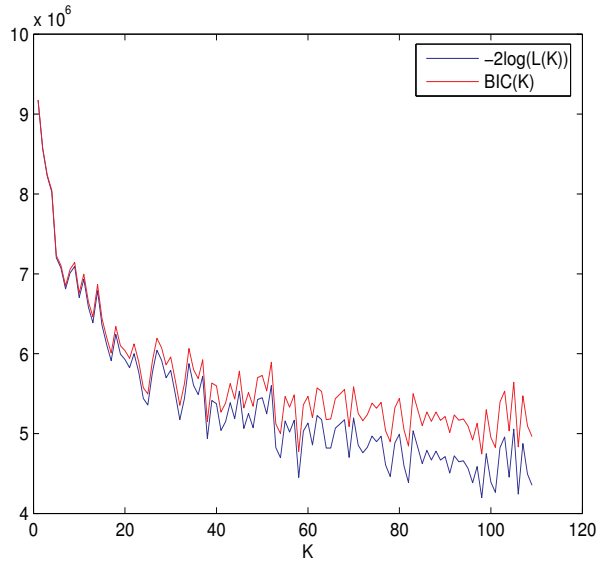

FIG. 5.26 – Graphique de la fonction BIC en fonction du nombre de clusters pour le lissage $\theta = 10$.

5.6 Analyse fractale

5.6.1 Présentation des données

Pour l'application relative à l'étude de l'autosimilarité avec irrégularité d'échantillonnage, nous travaillons avec des données de biologie marine fournie par l'Institut Français de Recherche pour l'Exploitation de la mer (IFREMER). Nous disposons d'un déplacement bidimensionnel de poisson bar nageant dans un aquarium de forme carrée. Pour capturer le mouvement, un marqueur est placé sur le dos du poisson. Une caméra dont l'axe de vision est perpendiculaire à la surface de l'eau du bassin est placée en haut de l'aquarium, comme illustré à la figure 5.27. Le bassin est filmé pendant une

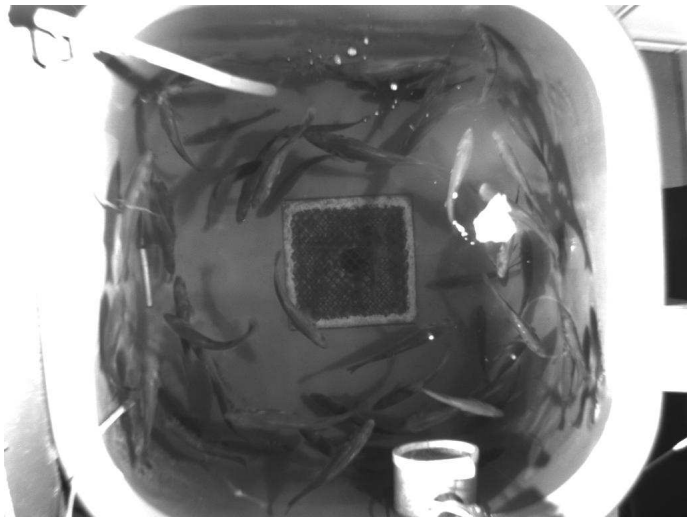

FIG. 5.27 – Bassin aquacole.

période de temps et les coordonnées bidimensionnelles du marqueur sont extraites par traitement d'image. Le déplacement est présenté à la figure 5.28. Le poisson nage en compagnie d'autres bars dans le bassin et cherche sa nourriture fournie par une trappe se situant en bas sur le graphique. La courbe paramétrée possède $N = 852$ captures. L'origine du repère correspond au coin en bas à gauche du bassin et le point $(6, 6)$ au coin supérieur droit. Le comportement du poisson (accélération, orientation, fuite ...) est perturbé par des stimulus tels que les changements de température de l'eau ou encore l'apport de nourriture. Cet exemple constituera une illustration d'une méthode de segmentation multidimensionnelle basée sur la signature de l'autosimilarité. Nous verrons que l'ajout d'une connaissance exogène telle que l'irrégularité d'échantillonnage perturbe l'estimation qui, si elle est fondée sur la matrice de covariance, permet de lever cette contrainte.

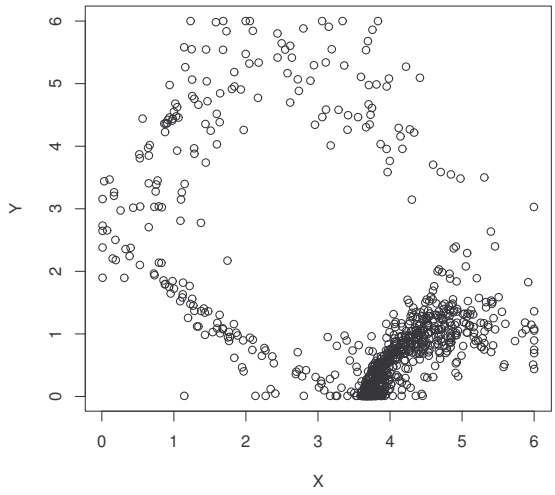

FIG. 5.28 – Déplacement de bar analysé ($N = 852$ captures).

5.6.2 Expression de la connaissance exogène

Au chapitre précédent, nous avions remarqué que l'autosimilarité issue d'une étude conjointe n'a de sens que si les estimations marginales fournissent les mêmes valeurs de H. Malheureusement, pour les données que nous utilisons, il est clair que ce n'est absolument pas le cas. Ainsi, nous analysons séparément les déplacements en X et Y en estimant successivement l'autosimilarité sur un voisinage temporel centré sur chaque date de capture. Pour cela, nous adoptons le voisinage symétrique centré en t :

$$\mathfrak{V}_{\theta,N_{\mathfrak{V}}}(t) = \{t' \,, \quad |t' - t| \leq \theta\} \tag{5.9}$$

où $N_{\mathfrak{V}}$ est la taille de la portion de trajectoire sur laquelle on détermine les points du voisinage.

5.6.3 Choix des paramètres

En moyenne, les dates de capture du déplacement sont espacées de deux dixièmes de seconde et nous fixons la largeur θ à 30, ce qui correspond à une durée moyenne d'environ six secondes par voisinage. De cette manière, les tailles de voisinage vont de 150 à 303 captures. Pour un voisinage fixé, nous avons en X et en Y une portion de trajectoire d'un mouvement brownien fractionnaire dont on estime par maximum de vraisemblance la tendance et la variance, afin d'aboutir à une valeur de H. La valeur

initiale est choisie égale à 0.5 et l'intervalle de recherche pour l'estimation à $[0.1, 0.9]$. Finalement, nous avons deux ensembles de N estimations locales de H (en X et en Y). Les différentes interprétations de l'irrégularité pour le poisson-bar seront évoquées dans la section qui suit.

Pour la méthode de segmentation, nous initialisons l'algorithme à l'aide de la méthode k-means (ou méthode des centres mobiles) [For65, Sap90]. La méthode des centres mobiles consiste à subdiviser les individus en k clusters distincts tels que chaque élément soient géométriquement proches. Au départ, on choisit les k-centres auxquels on affecte les points les plus proches. On ré-estime les centres de ces nouvelles classes, et ainsi de suite jusqu'à obtention d'une partition optimale. Le seuil de convergence η est fixé à 10^{-3} et le nombre d'états souhaités à $K = 3$.

5.6.4 Segmentation multidimensionnelle

Application au déplacement de bar

La figure 5.29 donne l'ensemble des estimations de H en X et Y. Les valeurs de $\widehat{H}$ sur les voisinages vont de 0.1 à 0.47 pour l'axe des abscisses, et de 0.1 à 0.49 pour l'axe des ordonnées. Les voisinages où l'EMV donne l'estimé égal à la borne inférieure de l'intervalle de recherche ($\widehat{H} = 0.1$) sont soit les zones présentant des valeurs aberrantes qui perturbent l'estimation (non-robustesse de l'EMV), ou encore celles où l'autosimilarité est trop hétérogène, ce qui ne permet pas d'obtenir une valeur raisonnable commune de H. Il s'agirait donc de réduire la taille des voisinages sur de telles zones. Néanmoins, et conformément aux résultats des tests d'appartenance à l'intervalle de confiance (tableau 4.10, irrégularité uniforme), le calcul de H sur un voisinage d'étendue au moins égale à 150 ($\theta = 30$) semble raisonnable en fournissant une estimation proche de la valeur théorique. Les voisinages présentant une mauvaise estimation de H ne concernent que 8 dates de capture pour X et 23 pour Y, pour un total de 852 voisinages. Pour le reste, les estimations donnent des valeurs homogènes sur des zones relativement étendues comparées à la taille du signal.

Une fois obtenues les estimations marginales, nous fusionnons les résultats à l'aide de l'opérateur d'agrégation du minimum et de la somme des estimations marginales et appliquons de fait la segmentation sur les signaux obtenus. L'opérateur du minimum donne d'autant plus d'importance aux trajectoires d'autosimilarité faible, *i.e.* très irrégulières. Le choix de l'opérateur "somme" est inspiré de l'expression du caractère autosimilaire du drap brownien fractionnaire. Le but de l'emploi de ces deux opéra-

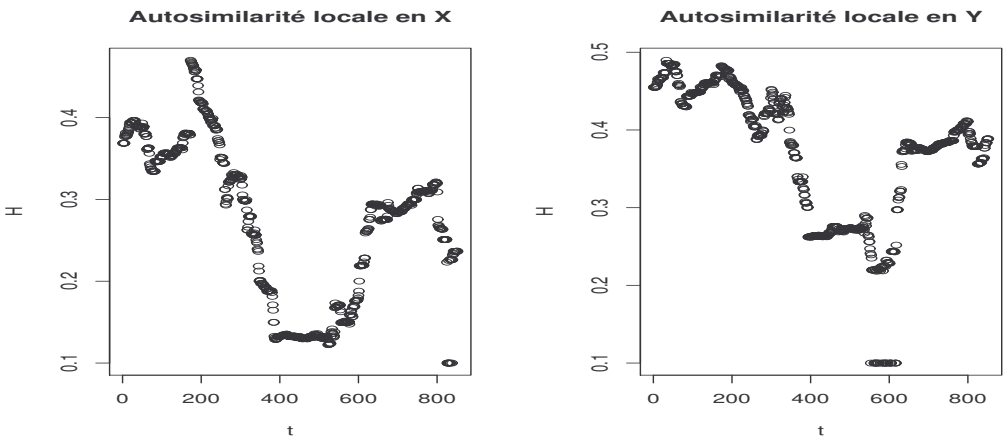

FIG. 5.29 – Résultat de l'estimation des autosimilarités locales en abscisse et en ordonnée.

teurs est de vérifier la cohérence des résultats. Les résultats de la segmentation avec les deux opérateurs d'agrégation sont donnés à la figure 5.30. Pour les deux cas, la segmentation à trois classes extrait les groupes de captures 1 à 347, 348 à 617 et 618 à 852, de moyennes respectives $\overline{H}_{+}$=0.807, 0.414 puis 0.656 pour la somme des estimations marginales, et $\overline{H}_{\min}$=0.359, 0.146 puis 0.279 pour le minimum. Les tracés du déplacement pour chacune des classes sont illustrés à la figure 5.31. La segmentation permet de dégager deux comportements hétérogènes du poisson. On constate que pour les captures 1 à 347, ($\overline{H}_{+} = 0.414$), le poisson se déplace de manière circulaire car soumis à la géométrie du bassin. Pour les captures allant de 348 à 617 ($\overline{H}_{+} = 0.414$), le comportement du poisson est stationnaire car il est à la recherche de nourriture fournie par une trappe se situant à la coordonnée (4,0) du bassin. La troisième portion de déplacement met en valeur les deux comportements (fuite et stationnarité) à cause des dernières estimations en ordonnée qui d'une part possède des valeurs qui ont tendance à diminuer, et d'autre part présentent également des zones d'autosimilarité hétérogènes.

Changements d'échelle

Nous venons de montrer l'homogénéité des estimations avec deux opérateurs d'agrégation. Dans la suite, on étudiera uniquement le cas de l'opérateur "somme". Certains articles [Ben04, Tur96, HHK+04] ont apporté des critiques sur la modélisation fractale, liées entres autres aux problèmes d'échelle et à la réelle nature fractale des données. Afin de valider en pratique la propriété autosimilaire des signaux, l'homogénéité des estimations est vérifiée en appliquant la même procédure à trois échelles temporelles

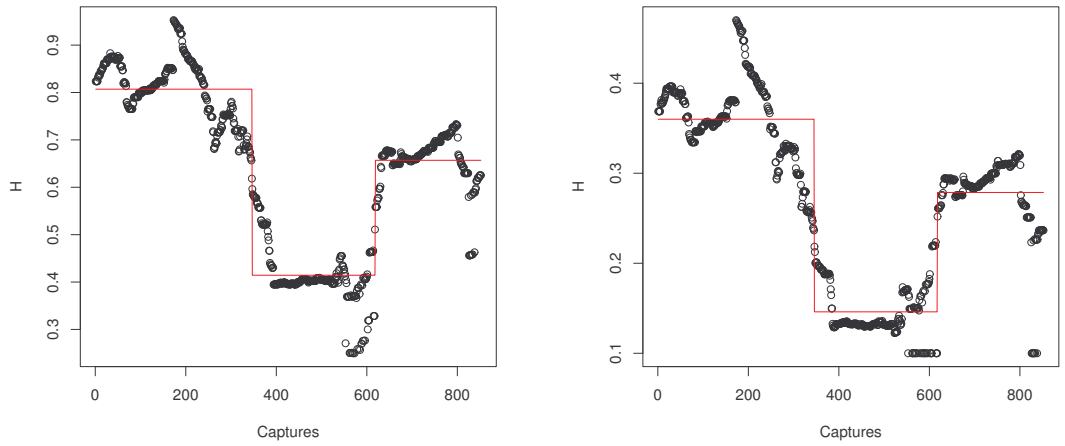

FIG. 5.30 – De gauche à droite, résultat de la segmentation pour la somme puis le minimum des estimations marginales de H.

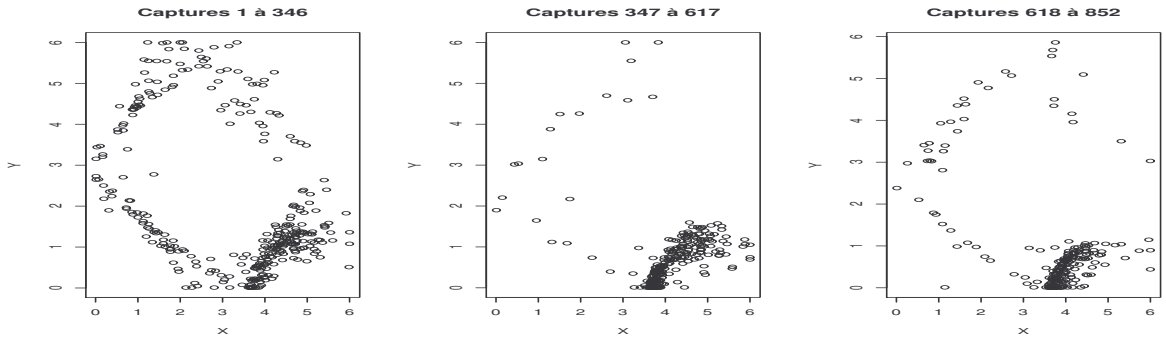

FIG. 5.31 – Tracés des portions de déplacement en fonction des résultats de la segmentation.

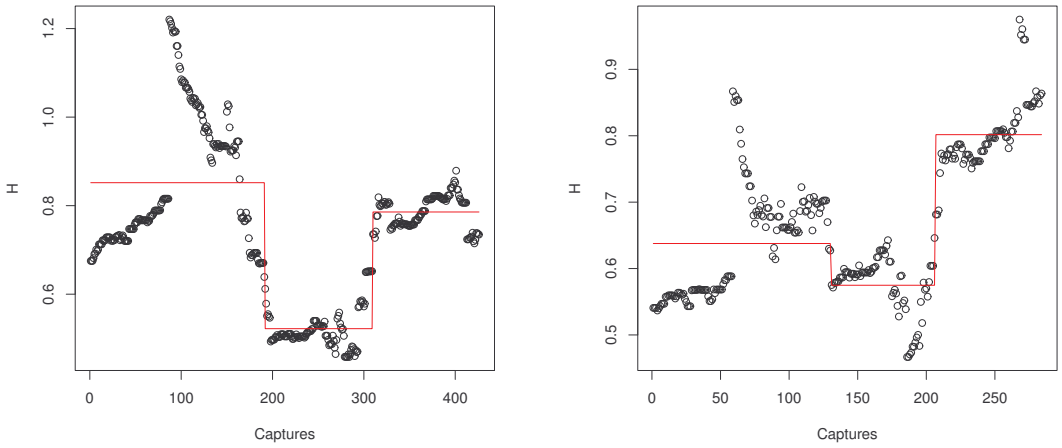

FIG. 5.32 – De gauche à droite, résultat de la segmentation pour la somme des estimations marginales de H avec une capture sur deux, puis une sur trois.

différentes. La première a été étudiée précédemment et nous considérons maintenant le signal avec deux échelles de temps successives : une capture sur deux puis une capture sur trois. Nous obtenons donc deux nouveaux signaux avec respectivement 426 et 284 captures. Les estimations et segmentations sont présentées à la figure 5.32. Les moyennes obtenues pour la subdivision à un sur deux sont égales à 0.863, 0.535 et 0.786. Celles calculées avec une capture sur trois donnent 0.638, 0.577 et 0.803. Les captures correspondant aux bornes de la segmentation sont respectivement de 180, 309 et 129, 207. Afin de pouvoir comparer avec le déplacement initial, ces captures correspondent en réalité aux captures 360, 618 (multiplié par deux) et 387, 621 (multiplié par trois), ce qui coïncide avec les captures 348, 618 de la segmentation du signal brut issu de l'opérateur "somme".

Analyse des composantes décorrélées

Puisque le poisson évolue dans un milieu clos, il existe par conséquent une certaine dépendance entre les coordonnées X et Y de son déplacement. Il convient donc de tester l'efficacité de la procédure de manière marginale, c'est à dire en l'appliquant aux composantes décorrélées du signal, obtenues par ACP. On représente de cette façon le signal dans une base orthogonale et on estime H sur chaque projection. Pour cela, nous appliquons une ACP non centrée aux déplacements en X et Y et réitérons la procédure sur les deux CP du signal. Le choix de la taille de la fenêtre restant inchangée, nous croisons les résultats obtenus sur chacune des deux CP C_1 et C_2. Avec l'opérateur

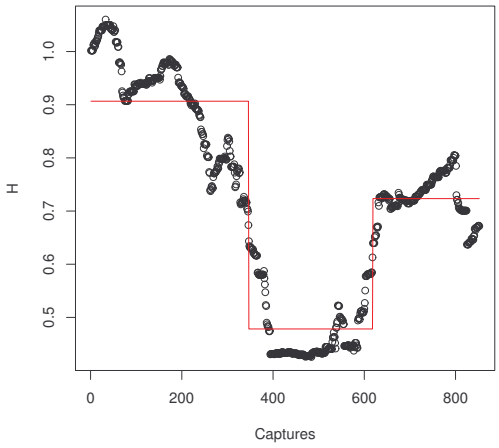

FIG. 5.33 – Résultat de la segmentation pour la somme des estimations marginales de H sur les composantes principales.

de la somme, nous obtenons des résultats similaires (figure 5.33). L'appartenance des estimations aux classes est identique par rapport à la trajectoire traitée initialement. Les segmentations calculées sur les CP ou sur la trajectoire brute donnent par conséquent les mêmes résultats.

5.7 Conclusion

Dans ce chapitre, nous avons appliqué les concepts présentés dans ce mémoire à l'analyse locale du mouvement. Dans ce cadre applicatif, nous avons choisi de définir la connaissance exogène par un graphe de proximité temporelle entre les observations, dont l'étendue est contrôlée par une variable fixée a priori par l'utilisateur. Dans ces cas, la covariance relationnelle tient compte de l'ordre temporel des captures et gomme les périodes d'immobilité du mouvement. L'espérance relationnelle joue en particulier le rôle de moyenne mobile et s'utilise pour supprimer la tendance locale des variables-capteurs.

L'ACP relationnelle autorise une reconnaissance et une analyse fine des liens inter-capteurs ainsi qu'un suivi de l'évolution dynamique du danseur au cours de l'exécution de son mouvement. Une simulation artificielle des données par reconstruction partielle permet de dégager les parties essentielles du mouvement. La qualité de cette reconstruction diffère suivant le nombre de liaisons pour le graphe.

Pour l'analyse d'opérateurs, nous avons étudié le cas normé, non normé et relationnel normé. La normalisation des opérateurs facilite les interprétations des axes graphiques, en imposant une disposition sphérique des opérateurs. La covariance relationnelle n'apporte pas significativement de meilleurs taux de reconnaissance pour la classification supervisée puisqu'elle met en valeur des caractéristiques plus discriminatoires que la covariance.

Dans le modèle de Wishart, nous avons abordé deux types de classification : supervisée ou non supervisée. La classification supervisée, fournissant de meilleurs taux de reconnaissance que l'analyse d'opérateurs même en dehors du cadre classique, permet de comparer les résultats en fonction du niveau de lissage choisi pour supprimer la tendance locale des variables-capteurs. La classification non supervisée permet d'extraire des sous-mouvements présents dans chaque classe de mouvements-type a priori. Grâce au critère de sélection BIC et à l'écart-type des différences symétriques entre partitions, nous avons noté un problème d'initialisation et de dimensionnalité pour la procédure EM.

La covariance permet également d'estimer l'autosimilarité du terme résiduel avec n'importe quel type d'échantillonnage. Nous en avons proposé un exemple avec un déplacement 2D de bar où les coordonnées sont supposées être des portions de trajectoires de mouvement brownien fractionnaire. Le paramètre H mesure l'autosimilarité spatiale et permet d'extraire des comportements locaux du poisson. Sur l'exemple présenté, les résultats sont cohérents par rapport aux déplacements et nous avons vérifié l'homogénéité en termes de changements d'échelles. De même, la segmentation obtenue sur les composantes décorrélées du signal est quasiment identique à celle calculée sur la trajectoire brute du poisson.

Chapitre 6

Conclusion générale

6.1 Synthèse

En analyse exploratoire, la statistique multidimensionnelle est la généralisation en dimension supérieure des méthodes d'estimation et de représentation des données. Dans cette thèse, nous nous sommes plus particulièrement intéressés aux approches fondées sur la covariance sous l'influence de l'ajout d'une connaissance exogène. La covariance est l'outil statistique de référence qui constitue une mesure de l'existence et du sens de la liaison linéaire entre les variables aléatoires. Le modèle choisi sépare le signal en une composante $\varphi(t)$, fonction déterministe du temps caractérisant la tendance, et une composante ϵ_t, terme résiduel supposé gaussien.

Les chapitres 2 et 3 concernent l'analyse du terme tendanciel $\varphi(t)$. Dans le chapitre 2, nous avons présenté un formalisme général de la covariance relationnelle qui permet de prendre en compte une information exogène de type relationnelle. Un certain nombre de propriétés qui éclairent le sens et les interprétations de cette notion ont été formulées. La covariance relationnelle est induite par la métrique du laplacien associé à la notion de graphe exprimant les liaisons par la relation binaire entre les individus. Elle agit non pas sur le graphe de la relation binaire mais plutôt sur la classe de graphes à laquelle il est associé. En particulier, cela signifie que la covariance relationnelle d'un graphe quelconque et de son symétrique coïncident. Implicitement, la covariance relationnelle est liée à la notion d'espérance relationnelle, opérateur auto-adjoint pour un certain produit scalaire lié aux propriétés topologiques du graphe, et se définit comme l'espérance d'une variable aléatoire conditionnellement au graphe de la relation binaire.

Au chapitre 3, nous avons considéré la covariance relationnelle comme une signature

dans les méthodes factorielles et plus particulièrement dans l'ACP, l'analyse factorielle d'opérateurs et l'analyse discriminante d'opérateurs. L'ACP relationnelle est un cas général de l'ACP standard, munie du laplacien comme métrique sur l'espace des variables aléatoires. Les extensions dans le cadre fonctionnel (ACP fonctionnelle) et dans le cas où les données présentent une liaison non linéaire (ACP à noyaux) ont également été présentées. Pour chacune des méthodes, nous avons construit le schéma de dualité permettant d'établir les liens avec l'ACP standard. L'analyse factorielle d'opérateurs de covariance relationnelle fait partie de l'analyse des tableaux à trois entrées. Elle est à la fois un cas général de l'analyse factorielle d'opérateurs et une variante de la méthode STATIS. Comme l'ACP, le principe est de construire une base orthonormée de l'espace des opérateurs de covariance relationnelle de manière à en obtenir une représentation synthétique. La technique s'étend à l'analyse factorielle discriminante d'opérateurs, dans laquelle on détermine les axes qui discriminent au mieux les classes formées par les projections des opérateurs de covariance relationnelle sur les premiers éléments de la base orthonormée.

Le terme résiduel a été étudié au chapitre 4 à l'aide d'approches statistiques, où ϵ_t a été supposé gaussien, hypothèse justifiée dans la mesure où l'on traite préalablement le signal X_t en lui retirant sa tendance (représentée par l'espérance relationnelle). Le cadre applicatif concerne la classification (distribution de Wishart) et la segmentation multidimensionnelle (analyse fractale). La matrice de covariance empirique de l'échantillon ϵ_t induit par définition une distribution de Wishart. Nous avons envisagé deux types de classification : supervisée et non supervisée. En classification supervisée, on associe à chaque classe une loi de Wishart de matrice d'échelle à estimer. Pour la classification non supervisée, nous avons défini le cadre formel de la procédure EM pour matrice de covariance. Nous nous sommes intéressés au caractère autosimilaire de ϵ_t à l'aide de l'analyse fractale. Nos résultats s'étendent au cas où l'échantillonnage est irrégulier, ce qui relâche la contrainte habituellement imposée par les méthodes d'estimation liées aux approches spectrales (Fourier, ondelettes). La méthode consiste à estimer le paramètre de Hurst H sur une trajectoire de processus brownien fractionnaire, dont la matrice de covariance permet la prise en compte de n'importe quel type d'échantillonnage (régulier ou non). L'échantillonnage irrégulier limitant le choix des techniques de simulation et d'estimation, nous nous sommes orientés vers les approches liées à la covariance : la méthode de Cholesky pour la simulation et le maximum de vraisemblance pour l'estimation. Les trajectoires de mouvement brownien fractionnaire ont été simulées en décomposant par l'algorithme de Cholesky la matrice de covariance du processus. Celle-ci étant théorique, la méthode est dite exacte et fournit d'excellentes simulations. L'estimation de H est établie par maximum de vraisemblance, méthode convergente et asymptotiquement efficace, et une étude menée sur cet estimateur a permis de déduire son comportement sur des trajectoires de longueur faible avec deux types

d'échantillonnage : l'un est systématique avec un échantillonnage périodique, l'autre est aléatoire avec un échantillonnage aléatoire uniforme. Cette dernière analyse confirme en particulier la plus grande sensibilité (biais supérieur) de l'estimation du maximum de vraisemblance dans le cas systématique par rapport au cas aléatoire. Nous avons ensuite proposé une application à la segmentation multidimensionnelle d'une courbe discrétisée, paramétrée par le temps et échantillonnée de manière non régulière. Le principe est basé sur un indice de fusion du paramètre de Hurst estimé sur chacune des trajectoires marginales.

Au chapitre 5, les concepts présentés ont été appliqués à l'analyse locale du mouvement. Pour les méthodes factorielles et la distribution de Wishart, les applications ont été effectuées sur un corpus de mouvement de danse contemporaine, où les variables aléatoires sont les coordonnées de capteurs placés sur un danseur exécutant une figure chorégraphique. Nous avons défini la relation binaire comme une proximité temporelle entre les individus. Cette relation dépend d'un paramètre θ qui permet de déterminer quelles sont les liaisons entre les individus pour le graphe. Avec cette relation binaire, et au vue des propriétés établies au chapitre 2, la covariance relationnelle tient compte de l'ordre temporel des captures, et gomme les périodes d'immobilité du mouvement pour un θ suffisamment petit. L'espérance relationnelle joue le rôle de moyenne mobile, ce qui, en la retranchant de la variable aléatoire associée, permet de lui retirer sa tendance locale. L'ACP relationnelle a permis de donner une représentation synthétique d'un mouvement, de dégager des mouvements locaux et de mettre en évidence des corrélations non perceptibles avec l'ACP standard. Dans le cadre de l'analyse de tableaux à trois entrées, l'opérateur de covariance relationnelle identifie les mouvements par une signature de même taille de manière à les rendre comparables. L'analyse factorielle d'opérateurs a permis de donner une représentation graphique des mouvements et de donner une interprétation des opérateurs principaux. Trois cas de figure ont alors été considérés : normé, non normé, et relationnel normé. Les taux de reconnaissance par classification supervisée ont été calculés par validation croisée au plus proche centre de gravité. Globalement, les taux de reconnaissance avec les opérateurs de covariance relationnelle n'apportent pas de meilleurs résultats du fait de la mise en avant d'autres caractéristiques plus discriminatoires. Dans le cas non supervisé, les résultats ont été validés en mesurant la différence symétrique entre la partition estimée et celle de référence. Les écarts-types entre les différences symétriques et le critère BIC ont confirmé les problèmes classiques de dimensionnalité et d'initialisation de la procédure EM. Pour l'analyse fractale, nous avons proposé une illustration à l'aide d'un déplacement 2D d'un bar évoluant dans un aquarium et échantillonné irrégulièrement. Les coordonnées sont supposées être des portions de trajectoires de mouvements browniens fractionnaires. À partir de l'estimation locale du paramètre de Hurst, le déplacement est alors segmenté par extraction des comportements locaux du poisson. Les résultats ont été validés sur

différentes échelles de temps et sur les composantes décorrélées du signal.

6.2 Rappel des contributions

Dans cette section, nous rappelons brièvement les principales contributions de la thèse. Une approche unificatrice reliant des méthodes factorielles et probabilistes a été adoptée selon trois considérations :

1. Covariance relationnelle
2. Distribution de Wishart
3. Mouvement brownien fractionnaire

Dans le premier point, nous avons présenté le formalisme de la covariance relationnelle, qui tient compte d'une relation binaire sur les observations. La notion d'espérance conditionnellement au graphe de la relation binaire a été définie. Des propriétés ont été proposées afin de donner de nouvelles interprétations. La covariance relationnelle s'utilise dans l'analyse en composantes principales, l'analyse factorielle d'opérateurs et l'analyse discriminante d'opérateurs. En particulier, nous avons établi que l'ACP relationnelle est un cas particulier de l'ACP à noyaux et de l'ACP fonctionnelle pour lesquelles nous donnons respectivement le noyau et les fonctions de base associées, ainsi que leur schéma de dualité.

Nous avons proposé dans le second point une approche de classification par la distribution de Wishart. En particulier, la classification non supervisée est établie par la procédure EM pour matrices de covariance.

Pour le dernier point, une méthode de segmentation par estimation du paramètre de Hurst a été proposée. Le comportement et la validité des résultats de l'estimateur du maximum de vraisemblance a été étudié pour des trajectoires de longueur faible échantillonnées irrégulièrement.

Les concepts présentés sont appliqués à l'analyse du mouvement dansé (méthodes factorielles et distribution de Wishart, représentation graphique et classification) et à la biologie marine (mouvement brownien fractionnaire, segmentation multidimensionnelle).

6.3 Perspectives

Nous menons actuellement une étude sur la distribution d'échantillonnage de la variance et de la covariance relationnelles. À terme, cela permettra d'introduire dans l'approche "statistique géométrique" adoptée lors des analyses factorielles des considérations probabilistes. On pourra ainsi de manière classique bâtir des intervalles de confiance et mener des tests d'hypothèses à partir des observations, pour juger soit du caractère significatif de la liaison linéaire relationnelle entre variables, soit de l'homogénéité d'un ensemble de variances relationnelles, ou enfin de l'égalité d'une variance relationnelle à une valeur donnée. En particulier, si le graphe de la relation binaire est complet, l'estimateur de la variance relationnelle est biaisé au même titre que celui de la variance. Pour un graphe quelconque, savoir si l'estimateur de la variance relationnelle présente lui aussi un biais semble être une question difficile qu'il serait intéressant d'étudier. D'un autre côté, et compte tenu du fait que le graphe tend à limiter les liaisons entre les individus, on peut remarquer que l'estimateur tend à être plus robuste à la présence de points aberrants, sous réserve que leurs liaisons soient également limitées par la relation binaire.

Un autre point qui n'a pas été abordé est le problème de dimensionnalité en classification non supervisée par la procédure EM, associé à la convergence vers un minima local. D'une manière générale, et même avec l'étape intermédiaire de l'algorithme *stochastic expectation-maximization*, nous n'avons pas la convergence vers le minimum global. Pour palier à la contrainte de dimensionnalité, une étape de pré-traitement de type ACP fondée sur une méthode d'analyse multi-tableaux comme l'analyse factorielle multiple. En sélectionnant un nombre réduit de CP, cela permettrait de diminuer la dimension matricielle de manière significative et efficace au détriment de la perte d'information. De plus, nous avons par symétrie $\frac{p(p+1)}{2}$ paramètres libres pour les matrices de covariance. Dès lors, le nombre de paramètres à estimer peut être très grand comparé à la taille de l'échantillon. Celeux *et al.* [CG95] proposent l'introduction de modèles "contraints" en exprimant les matrices de covariance à partir de leur décomposition spectrale et sous des hypothèses liées à leur type, leur volume, leur forme et leur orientation. Le but est de réduire le nombre de ddl afin d'affaiblir le modèle pour une meilleure convergence de l'algorithme. À la base, cette étude a été menée pour le clustering de points dont la densité s'exprime comme un mélange de gaussiennes. Avec les mêmes notations que l'article, les matrices de covariance s'expriment sous la forme :

$$\Sigma_k = \lambda_k D_k A_k {}^t D_k \tag{6.1}$$

où $\lambda_k = |\Sigma_k|^{1/p}$, D_k est l'ensemble des vecteurs propres de Σ_k et A_k la matrice diagonale des valeurs propres triées et normalisées de façon à avoir $|A_k| = 1$. Les paramètres λ_k, D_k et A_k sont respectivement appelés *volume*, *forme* et *orientation* du $k$ème cluster.

On peut alors imposer diverses hypothèses afin de diminuer la complexité du modèle d'apprentissage : $\forall k,\ \lambda_k = \lambda,\ D_k = D$ ou encore $A_k = A$, paramètres constants à estimer. Cela nous donne au total un ensemble de 14 types de matrice [CD85]. Sous ces hypothèses, il serait intéressant d'étudier le comportement des méthodes présentées dans ce mémoire en terme de taux de classification.

D'un point de vue expérimental, nous avons considéré une relation binaire unique sur les opérateurs de covariance (θ constant pour tous les mouvements). Il s'agit maintenant d'étudier le sens et les interprétations dans le cas où la proximité temporelle varie d'un opérateur à l'autre. Les taux de reconnaissance sont comparables voire meilleurs dans certains cas que ceux obtenus avec une approche probabiliste de type modèles de Markov [CBV04, Che04], appliquée à ces mêmes données. Ces résultats nous incitent à appliquer les différentes méthodes de classification supervisée (factorielles et de type Wishart) à la segmentation multidimensionnelle. Nous envisageons en effet de considérer une succession de mouvements de danse contemporaine formant un ballet et d'appliquer une procédure de type "fenêtre glissante" en comparant l'opérateur calculé sur chaque fenêtre avec la matrice de covariance estimée sur chacune des classes. Le but est de repérer et reconnaître chaque mouvement de l'exécution chorégraphique.

Annexe A

Chapitre 2

A.1 Preuve du sens de la définition 3

On a :

$$\begin{aligned} < X, Y >_{\Delta} &= \Delta(X).Y = \mathcal{D}(X).Y - \mathcal{A}(X).Y \\ &= \int_{\Omega} \left(X(\omega)Y(\omega)deg(\omega) - X(\omega) \int_{\Omega} Y(\omega')g(\omega,\omega')dP(\omega') \right) dP(\omega) \end{aligned} \tag{A.1}$$

Par définition de $deg(\omega)$, on obtient :

$$\Delta(X).Y = \iint_{\Omega\times\Omega} X(\omega)(Y(\omega) - Y(\omega'))g(\omega,\omega')dP(\omega')dP(\omega) \tag{A.2}$$

D'après le théorème de Fubini :

$$\begin{aligned} \frac{1}{\kappa} < X, Y >_{\Delta} &= \frac{1}{2\kappa}(\Delta(X).Y + \Delta(X).Y) \\ &= \frac{1}{2\kappa} \iint_{\Omega\times\Omega} X(\omega)(Y(\omega) - Y(\omega'))g(\omega,\omega')dP(\omega')dP(\omega) \\ &\quad + \frac{1}{2\kappa} \iint_{\Omega\times\Omega} X(\omega')(Y(\omega') - Y(\omega))g(\omega',\omega)dP(\omega')dP(\omega) \end{aligned} \tag{A.3}$$

Enfin, comme le graphe est symétrique :

$$\begin{aligned}\frac{1}{\kappa} < X, Y >_{\Delta} &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} X(\omega)(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega')dP(\omega) \\ &\quad +\frac{1}{2\kappa}\iint_{\Omega\times\Omega} X(\omega')(Y(\omega')-Y(\omega))g(\omega,\omega')dP(\omega')dP(\omega) \\ &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} X(\omega)(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega')dP(\omega) \\ &\quad -\frac{1}{2\kappa}\iint_{\Omega\times\Omega} X(\omega')(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega')dP(\omega) \\ &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega)dP(\omega') \\ &= Cov_g(X,Y)\end{aligned} \tag{A.4}$$

A.2 Démonstrations des propriétés

A.2.1 Propriété 2

Soit g un graphe quelconque :

$$\begin{aligned}&Cov_g(X,Y) = \\ &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega)dP(\omega') \\ &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))Y(\omega)g(\omega,\omega')dP(\omega)dP(\omega') \\ &\quad -\frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))Y(\omega')g(\omega,\omega')dP(\omega)dP(\omega')\end{aligned} \tag{A.5}$$

Ce qui donne en permutant les indices du second terme :

$$\begin{aligned}Cov_g(X,Y) &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))Y(\omega)g(\omega,\omega')dP(\omega)dP(\omega') \\ &\quad -\frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega')-X(\omega))Y(\omega)g(\omega',\omega)dP(\omega)dP(\omega') \\ Cov_g(X,Y) &= \frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))Y(\omega)g(\omega,\omega')dP(\omega)dP(\omega') \\ &\quad +\frac{1}{2\kappa}\iint_{\Omega\times\Omega} (X(\omega)-X(\omega'))Y(\omega)g(\omega',\omega)dP(\omega)dP(\omega')\end{aligned} \tag{A.6}$$

Donc :

$$Cov_g(X,Y) = \frac{1}{\kappa}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))Y(\omega)\left(\frac{g(\omega,\omega')+g(\omega',\omega)}{2}\right)dP(\omega)dP(\omega') \tag{A.7}$$

Le graphe g^* qui à (ω,ω') associe $\frac{(g(\omega,\omega')+g(\omega',\omega))}{2}$ est symétrique et on a :

$$\begin{aligned} Cov_g(X,Y) &= \frac{1}{\kappa}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))Y(\omega)g^*(\omega,\omega')dP(\omega)dP(\omega') && \text{(A.8)}\\ Cov_g(X,Y) &= \frac{1}{\kappa}\iint_{\Omega\times\Omega}(X(\omega')-X(\omega))Y(\omega')g^*(\omega,\omega')dP(\omega)dP(\omega') && \text{(A.9)}\\ &= -\frac{1}{\kappa}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))Y(\omega')g^*(\omega,\omega')dP(\omega)dP(\omega') && \text{(A.10)}\end{aligned}$$

En additionnant les formules (A.8) et (A.10), on en déduit :

$$Cov_g(X,Y) = \frac{1}{2\kappa}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))g^*(\omega,\omega')dP(\omega)dP(\omega') \tag{A.11}$$

A.2.2 Propriété 3

On note par $\overline{deg}(\omega)$ le degré de la mesure ω pour $\overline{g}$. La connectivité du graphe complémentaire se calcule de la manière suivante :

$$\begin{aligned} \overline{\kappa} &= \int_\Omega \overline{deg}(\omega)dP(\omega) && \text{(A.12)}\\ &= \iint_{\Omega\times\Omega}(1-g(\omega,\omega'))dP(\omega')dP(\omega)\\ &= 1-\iint_{\Omega\times\Omega}g(\omega,\omega')dP(\omega')dP(\omega)\\ &= 1-\int_\Omega deg(\omega)dP(\omega)\\ &= 1-\kappa\end{aligned}$$

D'après la linéarité de l'intégrale, il est clair que l'on a les formules :

$$\begin{aligned} Cov(X,Y) &= \frac{1}{2}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))dP(\omega)dP(\omega') && \text{(A.13)}\\ &= \frac{1}{2}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))g(\omega,\omega')dP(\omega)dP(\omega')\\ &\quad +\frac{1}{2}\iint_{\Omega\times\Omega}(X(\omega)-X(\omega'))(Y(\omega)-Y(\omega'))\overline{g}(\omega,\omega')dP(\omega)dP(\omega')\\ &= \kappa Cov_g(X,Y)+(1-\kappa)Cov_{\overline{g}}(X,Y)\end{aligned}$$

Donc :

$$< X,Y >_{\Delta_1} = < X,Y >_{\Delta} + < X,Y >_{\Delta_1-\Delta} \tag{A.14}$$

A.2.3 Propriété 4

D'après la propriété 3, on peut supposer sans perte de généralité la symétrie du graphe. En utilisant la définition 3, on en déduit :

$$\begin{aligned} Cov_g(X,Y) &= \frac{1}{\kappa} < X, Y >_\Delta \\ &= \frac{1}{\kappa} \iint_{\Omega\times\Omega} (X(\omega)Y(\omega) - X(\omega)Y(\omega'))g(\omega,\omega')dP(\omega)dP(\omega') \\ &= \frac{1}{\kappa} \int_\Omega X(\omega) \left(Y(\omega) - \frac{1}{deg(\omega)} \int_\Omega Y(\omega')g(\omega,\omega')dP(\omega') \right) deg(\omega)dP(\omega) \\ &= \frac{1}{\kappa} \int_\Omega X(\omega) \left(Y(\omega) - E(Y|g)(\omega) \right) deg(\omega)dP(\omega) \\ &= \int_\Omega X(\omega) \left(Y(\omega) - E(Y|g)(\omega) \right) dQ(\omega) \\ &= < X, Y - E(Y|g) >_Q \end{aligned} \tag{A.15}$$

A.2.4 Propriété 5

Il suffit d'appliquer le résultat de la propriété précédente pour X et Y variables aléatoires Q-centrées :

$$\begin{aligned} < X, Y >_\Delta &= \kappa < X, Y - E(Y|g) >_Q \\ &= \kappa < X, Y >_Q - \kappa < X, E(Y|g) >_Q \\ &= < Y, X >_\Delta \\ &= \kappa < Y, X - E(X|g) >_Q \\ &= \kappa < Y, X >_Q - \kappa < Y, E(X|g) >_Q \\ &= \kappa < X, Y >_Q - \kappa < E(X|g), Y >_Q \end{aligned} \tag{A.16}$$

Donc :

$$< E(X|g), Y >_Q = < X, E(Y|g) >_Q \tag{A.17}$$

A.2.5 Propriété 6

$$
\begin{aligned}
E_Q(E(X|g)) &= \int_\Omega E(X|g)(\omega)dQ(\omega) \qquad (A.18)\\
&= \int_\Omega \left(\frac{1}{deg(\omega)}\int_\Omega X(\omega)g(\omega,\omega')dP(\omega')\right)dQ(\omega)\\
&= \frac{1}{\kappa}\iint_{\Omega\times\Omega} X(\omega)g(\omega,\omega')dP(\omega')dP(\omega)\\
&= \frac{1}{\kappa}\int_\Omega X(\omega)\left(\int_\Omega g(\omega,\omega')dP(\omega')\right)dP(\omega)\\
&= \frac{1}{\kappa}\int_\Omega X(\omega)deg(\omega)dP(\omega)\\
&= \int_\Omega X(\omega)dQ(\omega) = E_Q(X)
\end{aligned}
$$

A.3 Éléments de théorie des graphes

La théorie des graphes est un domaine relativement ancien qui a débuté au 18ème siècle avec les travaux d'Euler afin de résoudre le célèbre problème des ponts de Königsberg. De nos jours, la théorie moderne a été développé par Berge [Ber62] et les applications se sont multipliées dans divers domaines tels que la chimie, la biologie, l'informatique et le traitement du signal. Ci-dessous, nous faisons un rappel sur quelques notions de base de la théorie des graphes.

Graphe

Un graphe simple est un ensemble noté (V, E) où V est un sous-ensemble fini de Ω constitué des éléments appelés *sommets* et E le sous-ensemble fini de $\Omega \times \Omega$ contenant les éléments appelés *arêtes*. Une arête est dite *réflexive* si elle connecte un sommet avec lui-même. La figure A.1(a) montre un exemple de représentation de graphe où les points symbolisent les sommets et les segments reliant certains points les arêtes. Deux sommets sont *adjacents* s'ils sont reliés par une arête. Si tous les sommets sont adjacents, alors le graphe est dit *complet*. Notons qu'il existe également les multi-graphes possédant des arêtes multiples mais ce type de graphe ne sera pas abordé dans ce travail.

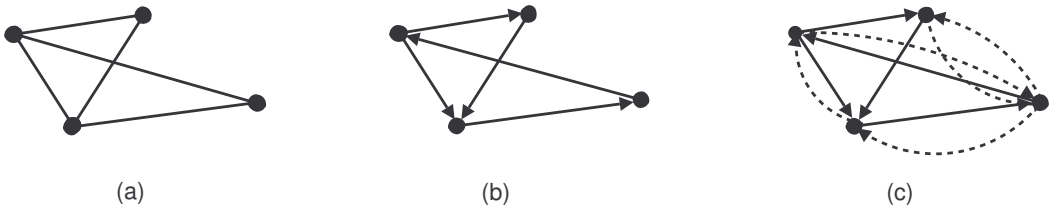

FIG. A.1 – (a) Graphe simple - (b) Graphe simple orienté - (c) Symétrisé.

Graphe orienté

Un graphe est dit orienté si on lui soumet un sens de parcours. C'est à dire qu'un sommet aura nécessairement un sommet *prédécesseur* et *successeur*. Ce type de graphe est symbolisé par des flèches orientant chacune des arêtes, comme illustré à la figure A.1(b). Chaque élément de E décrit le sommet origine et final d'une arête fléchée, qu'on appelle aussi *arc*. L'arc ayant pour extrémité d'origine ω et finale ω' sera noté $\{\omega, \omega'\} \in E$. Un *chemin* allant d'un sommet ω à ω' est une suite de sommets où passe ce chemin pour relier ω à ω'. La *longueur* d'un chemin correspond au nombre d'arêtes parcourues.

Graphe valué

Un graphe valué est un graphe orienté $G = (V, E)$ muni d'une fonction f dite *de coût* :

$$f \ : \ E \to \mathbb{R} \tag{A.19}$$

Cette fonction associe un réel $f(\{\omega, \omega'\})$ à l'arête $\{\omega, \omega'\}$ du graphe, représentant un poids que l'on affecte en fonction des sommets correspondants.

Graphe symétrique

Un graphe orienté est symétrique si pour tout couple (ω, ω') de sommets du graphe, on a la relation :

$$(\{\omega, \omega'\} \in E) \Longrightarrow (\{\omega', \omega\} \in E) \tag{A.20}$$

La figure A.2 en donne une illustration, où chaque couple de sommets en liaison pour le graphe est relié par deux arcs opposés. De même, le *symétrisé* d'un graphe G est le graphe noté $G + {}^tG$ obtenu avec l'ajout minimum d'arcs tels que $G + {}^tG$ soit symétrique.

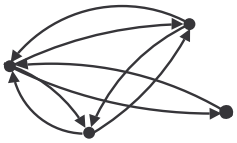

FIG. A.2 – Graphe symétrique.

À la figure A.1, les arêtes en pointillés ont été rajoutées au graphe (b) de manière à obtenir son symétrisé (c).

Composante connexe d'un graphe

La composante connexe d'un sommet ω d'un graphe $G = (V, E)$ est le sous-ensemble de V constitué des sommets pouvant être atteints à partir de ω. Par exemple, le graphe de la figure A.1 n'a aucune composante connexe. Un graphe est dit *connexe* si il existe un chemin reliant n'importe quel couple de sommets.

A.4 Éléments de théorie des processus

La théorie des processus [CM77, Sap90] est un domaine se situant entre les probabilités et les statistiques. Les processus permettent de modéliser l'évolution d'un phénomène aléatoire en général dépendant du temps. Les thèmes d'application sont très nombreux : traitement du signal, économétrie, traffic ethernet, télécommunication, ... Dans cette section, nous énonçons certaines des notions de base de la théorie des processus aléatoires.

Processus

Une *fonction aléatoire* (ou processus aléatoire) X est définit comme une suite de variables aléatoires indicées par un ensemble $\mathbb{T}$. Le processus aléatoire X est donc une application d'un ensemble $\Omega \times \mathbb{T}$ sur $\mathbb{R}$:

$$\begin{array}{rccc} X & : & \mathbb{T} \times \Omega & \rightarrow & \mathbb{R} \\ & & (t, \omega) & \mapsto & X(t, \omega) \end{array} \tag{A.21}$$

L'ensemble Ω est l'ensemble des observations et $\mathbb{T}$ l'ensemble sur lequel sont indicées les variables. Pour $t \in \mathbb{T}$ fixé, la quantité $X(t, .)$, notée plus simplement X_t, est une

variable aléatoire définie sur Ω. Si ω est une mesure fixée de Ω, la quantité $X(.,\omega)$, notée $X(\omega)$, est appelée *trajectoire* du processus. Une fonction aléatoire est donc à la fois une famille de variables aléatoires indicées par $\mathbb{T}$ et une famille de trajectoires indicées par Ω.

Les fonctions aléatoires permettent de modéliser un très grand nombre de phénomènes qui extrait de manière pondérée un ensemble de mesures. En général, le but est de soit établir une prévision robuste à plus ou moins long-terme des variables, soit de repérer les tendances, les saisonnalités ou les chocs structurels des trajectoires, ou encore de déterminer la causalité des phénomènes en fonction des réalisations des variables précédentes.

Dans ce travail, $\mathbb{T}$ est un ensemble temporel contenant les dates de capture d'un phénomène, et l'on écrira donc *processus stochastique* plutôt que fonction aléatoire. Un processus stochastique est dit *gaussien* si toute sous-famille finie de variables aléatoires est un vecteur aléatoire gaussien.

Accroissements

On appelle *accroissement d'ordre τ de X* le processus X^τ défini sur $\Omega \times \mathbb{T}$ obtenu en faisant la différence des variables aléatoires de X espacées de τ pour $\mathbb{T}$:

$$\begin{array}{rccl} X^\tau & : & \mathbb{T}\times\Omega & \rightarrow & \mathbb{R} \\ & & (t,\omega) & \mapsto & X_t^\tau(\omega) = X_{t+\tau}(\omega) - X_t(\omega) \end{array} \tag{A.22}$$

Si $\tau = 1$, on écrira simplement accroissement sans mentionner l'ordre un.

Fonction d'autocorrélation

La fonction d'*autocovariance* du processus X est donnée par :

$$\begin{array}{rccl} \gamma_t(h) & : & \mathbb{T}\times\mathbb{R} & \rightarrow & \mathbb{R} \\ & & (t,h) & \mapsto & \gamma_t(h) = Cov(X_t, X_{t-h}) \end{array} \tag{A.23}$$

La fonction d'*autocorrélation* est la fonction d'autocovariance normalisée par les écarts-type de X_t et X_{t-h}

$$\begin{array}{rccl} \rho_t(h) & : & \mathbb{T}\times\mathbb{R} & \rightarrow & \mathbb{R} \\ & & (t,h) & \mapsto & \rho_t(h) = \frac{\gamma_t(h)}{\sqrt{\gamma_t(0)\gamma_{t-h}(0)}} \end{array} \tag{A.24}$$

L'autocorrélation caractérise les dépendances internes entre les variables aléatoires du processus. Elle admet un maximum local en l'origine puisque $\rho_t(0) = 1$.

Stationnarité

Un processus est *stationnaire* au second ordre si ses deux premiers moments (espérance et autocovariance) sont indépendants de $\mathbb{T}$:

$$\begin{aligned} E(X_t) &= \mu < \infty \\ \gamma_t(h) &= \gamma(h) \end{aligned} \tag{A.25}$$

Ainsi, la stationnarité implique que la fonction d'autocovariance (et d'autocorrélation) est une fonction paire et qu'elle ne dépend que de l'écart h entre les variables aléatoires. Si la fonction d'autocorrélation admet un autre maximum local en dehors de l'origine, alors celle-ci est nécessairement périodique et a donc une infinité de maximums locaux. Plus généralement, on définit la stationnarité stricte si tous les moments du processus à tout ordre fini sont indépendants du temps. Dans le cas d'un processus gaussien, la stationnarité au second ordre et la stationnarité stricte coïncident car le processus est entièrement déterminé par ses deux premiers moments.

Ergodicité

Dans la pratique, on ne réalise qu'un ensemble restreint de mesures que l'on assimile ensuite à une trajectoire comme une suite de réalisations des variables aléatoires. La notion de processus dans son intégralité est donc purement théorique et l'on est amené à faire des hypothèses garantissant une certaine représentativité de la trajectoire. En fait, il est assez naturel de considérer que les caractéristiques du processus peuvent se retrouver dans une seule de ses trajectoires. La théorie ergodique permet de s'assurer de ce phénomène de manière raisonnable. L'ergodicité est très importante, voire indispensable, sous peine de se retrouver extrêmement limiter dans l'interprétation. L'*hypothèse d'ergodicité*, issue de la théorie ergodique, suppose que les moyennes d'ensemble coïncident avec les moyennes temporelles :

$$\frac{1}{n}\int_{-n/2}^{n/2} X(t,\omega)dt \underset{n\to+\infty}{\longrightarrow} E(X_t) \tag{A.26}$$

De la même façon, si les variables aléatoires du processus stationnaire sont centrées réduites, alors :

$$\frac{1}{n}\int_{-n/2}^{n/2} X(t+h,\omega)X(t,\omega)dt \underset{n\to+\infty}{\longrightarrow} \rho(h) \tag{A.27}$$

Ainsi, un processus est *ergodique* si toutes ses caractéristiques peuvent être déterminées à partir d'une seule de ses trajectoires [Sap90]. Une trajectoire y de l'accroissement s'exprime à partir de la trajectoire observée $(x_1, \ldots, x_n)$:

$$y = (x_n - x_{n-1}, \ldots, x_2 - x_1) \tag{A.28}$$

La fonction d'autocorrélation empirique $\widehat{\rho}_t(h)$, qui donne une estimation de ρ_t, est définie de la manière suivante :

$$\widehat{\rho}_t(h) = \frac{\sum_{i=1}^{n-h}(x_i - \overline{x})(x_{i+h} - \overline{x})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} \tag{A.29}$$

La fonction d'autocorrélation empirique est calculée en fonction de la moyenne empirique $\overline{x}$ de la série et la disposition du nuage des points (x_i, x_{i+h}) donne une indication sur la valeur de $\widehat{\rho}_t(h)$. Plus l'allure du nuage est linéaire, plus la valeur est proche de ± 1, où le signe dépend de la pente de la droite de régression.

A.5 Éléments de théorie de l'estimation

La théorie de l'estimation s'appuie sur la statistique inférentielle, dont le but est de déduire des résultats statistiques d'une population entière à partir d'un échantillon observé. Le principe est de calculer la quantité souhaitée à partir de l'échantillon puis d'évaluer son "potentiel" à fournir des conclusions fiables concernant l'ensemble de la population. Cette section a pour but de formuler l'ensemble des notions de base pour effectuer cette démarche.

Un estimateur T est une variable aléatoire dont chaque réalisation est donnée par une valeur, appelée *estimation*, calculée sur un échantillon de référence. En tant que variable aléatoire, T possède donc une espérance et une variance que l'on cherche à mettre en valeur grâce au biais, à l'efficacité et à la robustesse de l'estimateur. Comme l'estimateur est issu d'un échantillon de taille réduite et parfois non représentatif, sa valeur peut être entachée d'une erreur par rapport à la valeur théorique. Cette erreur est quantifiée par le *biais* de l'estimateur qui se définit comme la différence entre la valeur théorique et son espérance, soit $T - E(T)$. Si $T - E(T) = 0$, alors l'estimateur est dit *sans biais*.

En théorie, les meilleurs estimateurs ont pour caractéristiques d'être sans biais et de variance minimale. Dans ce cas, ils sont dits *efficaces*. Plus précisément, l'estimateur T non biaisé est efficace si pout tout autre estimateur T^*, on a :

$$Var(T) < Var(T^*) \tag{A.30}$$

Même si la population est de taille infinie, il n'est pas toujours possible d'avoir accès à la valeur théorique de la variance. Néanmoins, la variance d'un estimateur non biaisé est toujours inférieure à une valeur seuil, appelée *borne de Cramer-Rao*, égale à l'inverse de l'information de Fisher [Sap90].

Si T_n est un estimateur de T sur un échantillon de taille n, alors l'estimateur est *convergent* s'il converge en probabilité vers sa valeur théorique :

$$\forall \varepsilon > 0\,, \quad P(|T_n - T| > \varepsilon) \underset{n \to +\infty}{\longrightarrow} 0 \tag{A.31}$$

La *précision* d'un estimateur est liée à sa variance et la convergence de l'estimateur sera rapide si celui-ci est "précis", c'est à dire de variance faible. Enfin, un estimateur est dit *robuste* si sa valeur n'est pas perturbée par l'ajout dans la population de points "aberrants", c'est à dire des valeurs extrêmes par rapport au comportement initial de l'échantillon.

Annexe B

Chapitre 3

B.1 Obtention des composantes principales d'une ACP

Soit $(\mathbb{X}, M, D)$ un triplet statistique, où $\mathbb{X}$ est un tableau $n \times p$ contenant en colonne p variables aléatoires discrétisées notées X_i. Le produit scalaire sur les individus (resp. variables) est induit par la métrique M (resp. D). La notion d'information est liée à celle de variance et covariance entre les variables aléatoires et l'on cherche les composantes principales (CP) qui en donnent la meilleure représentation. Les CP forment une base de l'espace vectoriel engendré par les variables aléatoires X_i. Chaque CP est une combinaison linéaire des variables aléatoires initiales. Pour j fixé :

$$\exists!\ u = {}^t(u_1, \cdots, u_p) : C_j = \sum_{i=1}^{p} \sum_{j=1}^{p} M_{ij} u_j X_i \tag{B.1}$$

Le problème se résume à déterminer les coefficients u_{ij}. Expliquons le raisonnement pour la première CP. Avec les mêmes notations, on a :

$$C_1 = \sum_{i=1}^{p} \sum_{j=1}^{p} M_{ij} u_j X_i = \mathbb{X} M u \tag{B.2}$$

Un moyen simple d'augmenter la variance de C_1 est d'accroître la valeur relative des composantes de u sans toutefois jouer sur sa norme. On impose donc une contrainte de normalisation $\|u\|_M = {}^t u M u = 1$. Le problème s'exprime ainsi : C_1 étant de variance maximale, trouver un vecteur $u \in \mathbb{R}^n$ solution de

$$\max_{\|u\|_M = 1} Var(\mathbb{X} M u) = \max_{\|u\|_M = 1} ({}^t u M {}^t\mathbb{X}) D (\mathbb{X} M u) = \max_{\|u\|_M = 1} {}^t u M V M u \tag{B.3}$$

Pour déterminer la première CP, il s'agit donc de maximiser la forme quadratique ${}^tuMVMu$ sous la contrainte $\|u\|_M = 1$. On utilise pour cela un multiplicateur de Lagrange λ. Le lagrangien $\mathfrak{d}$ s'écrit :

$$\mathfrak{d} = {}^tuMVMu - \lambda\,({}^tuMu - 1) \tag{B.4}$$

Pour que u maximise $\mathfrak{d}$, alors nécessairement :

$$\frac{\partial \mathfrak{d}}{\partial u} = 0 \quad \text{et} \quad \frac{\partial \mathfrak{d}}{\partial \lambda} = 0 \tag{B.5}$$

Comme M est symétrique, la dérivation vectorielle nous permet d'écrire :

$$\frac{\partial \mathfrak{d}}{\partial u} = 2MVMu - 2\lambda Mu = 0 \tag{B.6}$$

Donc :

$$\frac{\partial \mathfrak{d}}{\partial u} = 0 \Longleftrightarrow MVMu = \lambda Mu \tag{B.7}$$

En multipliant à gauche les deux termes par M^{-1} il vient :

$$\frac{\partial \mathfrak{d}}{\partial u} = 0 \Longleftrightarrow VMu = \lambda u \tag{B.8}$$

La solution est donc le vecteur propre de VM associé à la plus grande valeur propre λ. On en déduit ensuite C_1 en effectuant la somme des variables X_i pondérée par les éléments du vecteur solution u. Pour obtenir les autres CP, la procédure est réitérée en rajoutant dans B.4 la contrainte d'orthogonalité mutuelle entre les vecteurs propres.

Bibliographie

[ABL84] T. Aluja-Banet and L. Lebart. Local and partial principal component analysis and correspondence analysis. *Proceedings in Computational Statistics (COMPSTAT)*, pages 113–118, 1984.

[ABNT91] T. Aluja-Banet and R. Nonnell-Torrent. Local principal components analysis. *Statistics and Operations Research Transactions Journal*, 15(3) :267–278, 1991.

[AC99] J.K. Aggarwal and Q. Cai. Human motion analysis : A review. *Computer Vision and Image Understanding*, 73 :428–440, 1999.

[ACLS98] J.K. Aggarwal, Q. Cai, W. Liao, and B. Sabata. Nonrigid motion analysis : Articulated and elastic motion. *Computer Vision and Image Understanding*, 70 :142–156, 1998.

[ACLV00] A. Ayache, S. Cohen, and J. Lévy-Véhel. The covariance structure of multifractional brownian motion with application to long range dependence. *International Conference on Acoustics Speech, and Signal Processing (ICASSP)*, 6 :3810–3813, 2000.

[AHSE95] H.H. Andersen, M. Hojbjerre, D. Svrensen, and P.S. Ericksen. Linear and graphical models (for the multivariate complex normal distribution). 1995.

[AM85] N. Alon and V.D. Milman. λ_1, isoperimetric inequalities for graphs, and superconcentrators. *Journal of Combinatorial Theory Series B*, 38 :73–88, 1985.

[AOVE05] H. Abdi, A.J. O'Toole, D. Valentin, and B. Edelman. DISTATIS : The analysis of multiple distance matrices. *Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society Conference*, 3 :42–42, 2005.

[AS96] P. Abry and F. Sellan. The wavelet-based synthesis for fractional brownian motion. *Applied and Computational Harmonic Analysis*, 3 :377–383, 1996.

[Bar99] J.M. Bardet. Un test d'autosimilarité pour les processus gaussiens à accroissements stationnaires. *Compte-rendu de l'Académie des Sciences*, Série I Mathématiques :521–526, 1999.

[BD91] P.J. Brockwell and R.A. Davis. *Time Series : Theory and Methods*. Springer, Second edition, 1991.

[BE90] H. Benali and B. Escofier. Analyse factorielle lissée et analyse factorielle des différences locales. *Revue de Statistique Appliquée*, 38 :55–76, 1990.

[Bel61] R. Bellman. *Adaptative Control Processes : A Guided Tour*. Princeton University Press, Princeton, USA, 1961.

[Ben04] S. Benhamou. How to reliably estimate the tortuosity of an animal's path : straightness, sinuosity, or fractal dimension ? *Journal of Theoretical Biology*, 229 :209–220, 2004.

[Ber62] C. Berge. *The Theory of Graphs and its Applications*. Wiley, New York, 1962.

[Ber94] J. Beran. *Statistics for Long-Memory Processes*. Chapman and Hall, New York, USA, 1994.

[BHK97] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs fisherfaces : Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19 :711–720, 1997.

[BKdL96] J.M.F. Ten Berge, H.A.L. Kiers, and J. de Leeuw. Some uniqueness results for PARAFAC2. *Psychometrika*, 61, 1996.

[BN02] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems 14*, 2002.

[BN03] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15 :1373–1396, 2003.

[BN06] F. Baudoin and D. Nualart. Notes on the two-dimensional fractional brownian motion. *The Annals of Probability*, 34 :159–180, 2006.

[Bol98] B. Bollodas. *Graph Theory*. Springer Verlag, 1998.

[Bri75] D.R. Brillinger. *Time Series : Data Analysis and Theory*. New York, USA, 1975.

[Bro98] R. Bro. *Multi-way Analysis in the Food Industry. Models, Algorithms, and Applications*. PhD thesis, University of Amsterdam, 1998.

[Cas06] J. Castaing. *Méthodes PARAFAC pour la séparation de signaux*. PhD thesis, Université de Cergy-Pontoise, 2006.

[CBV04] F. Chenevière, S. Boukir, and B. Vachon. A HMM-based dance gesture recognition. *International Workshop on Systems, Signals and Image Processing (IWSSIP)*, pages 322–326, 2004.

[CCS04] A. Carbone, G. Castelli, and H.E. Stanley. Time-dependent hurst exponent in financial time series. *Physica A*, 344 :267–271, 2004.

[CD85] G. Celeux and J. Diebolt. The SEM algorithm : a probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2 :73–92, 1985.

[CG95] G. Celeux and G. Govaert. Gaussian parsimonious clustering models. *Pattern Recognition*, 28 :781–793, 1995.

[Che04] F. Chenevière. *Contribution à l'analyse du mouvement dansé*. PhD thesis, Université de La Rochelle, 2004.

[CM77] D.R. Cox and H.D. Miller. *The Theory of Stochastic Processes*. Chapman and Hall/CRC, 1977.

[Coe00] J.-F. Coeurjolly. Simulation and identification of the fractional brownian motion : a bibliographical and comparative study. *Journal of Statistical Software*, 5 :1–53, 2000.

[CP76] F. Cailliez and J.-P. Pagès. Introduction à l'analyse des données. *SMASH*, 1976.

[Cre93] N. Cressie. *Statistics for Spatial Data*. Wiley, New York, 1993.

[CS95] C. Cedras and M. Shah. Motion-based recognition : A survey. *Image and Vision Computing*, 13 :129–155, 1995.

[CSC93] P.-A. Cornillon, R. Sabatier, and D. Chessel. Analyse d'un cube sous double contrainte de voisinage. *In 49th session of the International Statistical Institute*, pages 285–286, 1993.

[Dah89] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *The Annals of Statistics*, 17 :1749–1766, 1989.

[Dan98] M. Van Dang. *Classification de données spatiales : Modèles probabilistes et critères de partitionnement*. PhD thesis, Université Technologique de Compiègne, 1998.

[DLR77] A.P. Dempster, N.M. Laird, and D.B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39 :1–38, 1977.

[DM03] A.B. Dieker and M. Mandjes. On spectral simulation of fractional brownian motion. *Probability in the Engineering and Informational Sciences*, 17 :417–434, 2003.

[EB04] A. Von Eye and G.A. Bogat. Testing the assumption of multivariate normality. *Psychology Science*, 46 :243–258, 2004.

[EG04] A. Von Eye and J.C. Gardiner. Locating deviations from multivariate normality. *Understanding Statistics*, 3 :313–331, 2004.

[EP98] B. Escofier and J. Pagès. *Analyses factorielles simples et multiples*. Dunod, 1998.

[Esc80] Y. Escoufier. L'analyse conjointe de plusieurs matrices de données. *Biométrie et Temps, éditions Jolivet*, pages 59–76, 1980.

[Esc85] Y. Escoufier. Objectifs et procédures de l'analyse conjointe de plusieurs tableaux. Statistique et analyse de données :1–10, 1985.

[Esc89] B. Escofier. Multiple correspondence analysis and neighboring relation. *Data Analysis, Learning Symbolic and Numeric Knowledge*, pages 55–62, 1989.

[FFPL01] L. Ferro-Famil, E. Pottier, and J.-S. Lee. Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/alpha-wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 39 :2332–2342, 2001.

[Fie73] M. Fiedler. Algebraic connectivity of graphs. *Czechoslovak Mathematical Journal*, 23 :298–305, 1973.

[Fol82] Y. Le Foll. Pondération des distances en analyse factorielle. *Revue Statistiques et Analyse des Données*, 1 :13–31, 1982.

[For65] E. Forgy. Cluster analysis of multivariate data : Efficiency vs. interpretability of classifications. *Biometrics*, 21, 1965.

[Fou84] T. Foucart. *Analyse factorielle de tableaux multiples.* Masson, Paris, 1984.

[FT86] R. Fox and M.S. Taqqu. Large sample properties of parameter estimates for strongly dependent stationary time series. *The Annals of Statistics*, 14 :517–532, 1986.

[GCL03] J.B. Gao, Y. Cao, and J.-M. Lee. Principal component analysis of $1/f^{\alpha}$ noise. *Physics Letters A*, 314 :392–400, 2003.

[Gea54] R.C. Geary. The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5 :115–145, 1954.

[GN00] A.K. Gupta and D.K. Nagar. *Matrix Variate Distributions.* Chapman and Hall/CRC, Boca Raton, USA, 2000.

[Gow75] J.C. Gower. Generalized procrustes analysis. *Psychometrika*, 40 :33–51, 1975.

[Har70] R.A. Harshman. Foundations of the PARAFAC procedure : model and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Papers in phonetics*, 16 :1–84, 1970.

[Har72] R.A. Harshman. PARAFAC 2 : mathematical and technical note. *Working Paper in Phonetics*, 1972.

[Hay99] S. Haykin. *Neural Networks.* Prentice-Hall, Englewood, USA, 1999.

[HC84] D.M. Hawkins and N. Cressie. Robust kriging - a proposal. *Journal of the International Association for Mathematical Geology*, 16 :3–18, 1984.

[Her04] E. Herbin. *Processus (multi-)fractionnaires à paramètres multidimensionnels et régularité höldérienne*. PhD thesis, Paris XI Orsay, 2004.

[HHK+04] J.M. Halley, S. Hartley, A.S. Kallimanis, W.E. Kunin, J.J. Lennon, and S.P. Sgardelis. Uses and abuses of fractal methodology in ecology. *Ecology Letters*, 7 :254–271, 2004.

[HL96] R.A. Harshmana and M.E. Lundy. Uniqueness proof for a family of models sharing features of tucker's three-mode factor analysis and PARAFAC/CANDECOMP. *Psychometrika*, 61, 1996.

[HLSJ06a] S. Hidot, J.-Y. Lafaye, and C. Saint-Jean. ACP relationnelle pour l'analyse du mouvement : Application à la danse. *Congrès Reconnaissance des Formes et Intelligence Artificielle (RFIA)*, page 64, 2006.

[HLSJ06b] S. Hidot, J.-Y. Lafaye, and C. Saint-Jean. Analyse factorielle d'opérateurs pour l'étude du mouvement : Application à la danse. *Atelier "Fouille de données", Journées francophones Extraction et Gestion des Connaissances (EGC)*, pages 21–30, 2006.

[HLSJ06c] S. Hidot, J.-Y. Lafaye, and C. Saint-Jean. Discriminant factor analysis for movement recognition : application to dance. *Proceedings of the International Conference on Computer Vision and Graphics (ICCVG'06), Special issue in Journal of Machine Graphics and Vision*, 15(3-4) :391–399, 2006.

[HLSJ07] S. Hidot, J.-Y. Lafaye, and C. Saint-Jean. Propriétés et interprétation de la covariance relationnelle en ACP. *Revue Traitement du Signal*, 24(1) :29–37, 2007.

[Hor61] P. Horst. Relation among m sets of measures. *Psychometrika*, 26 :129–149, 1961.

[Hot33] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24 :417–441, 498–520, 1933.

[HSJL] S. Hidot, C. Saint-Jean, and J.-Y. Lafaye. Étude expérimentale de l'influence d'un échantillonnage irrégulier dans l'estimation du paramètre de Hurst. *À paraître dans la Revue de Statistique Appliquée.*

[HSJL07] S. Hidot, C. Saint-Jean, and J.-Y. Lafaye. Classification de signaux multidimensionnels utilisant la distribution de wishart : Application à la reconnaissance de mouvements. *Actes du Groupement de Recherche en Traitement du Signal et des Images (GRETSI)*, 2007.

[Hub00] P. Hubert. The segmentation procedure as a tool for discrete modeling of hydrometeorological regimes. *Stochastic environmental research and risk assessment*, 14 :297–304, 2000.

[IL94] J. Istas and G. Lang. Variations quadratiques et estimation de l'exposant de hölder local d'un processus gaussien. *Compte-rendu de l'Académie des Sciences, Série I Mathématiques*, 319 :201–206, 1994.

[JDM00] A.K. Jain, R.P.W. Duin, and J. Mao. Statistical pattern recognition : A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 :4–37, 2000.

[JHJ96] R. Jennane, R. Harba, and G. Jacquet. Estimation de la qualité des méthodes de synthèse du mouvement brownien fractionnaire. *Revue Traitement du Signal*, 13 :289–302, 1996.

[JHJ01] R. Jennane, R. Harba, and G. Jacquet. Méthodes d'analyse du mouvement brownien fractionnaire : théorie et résultats comparatifs. *Revue Traitement du Signal*, 18 :419–436, 2001.

[Jol86] I.T. Joliffe. *Principal Component Analysis*. Springer Verlag, 1986.

[Kah85] J.P. Kahane. *Some random series of functions*. Cambridge Studies in Advanced Mathematics 5, 2nd Edition, 1985.

[Kam96] A. Kamont. On the fractional anisotropic wiener field. *Probability and Mathematical Statistics*, 16 :85–98, 1996.

[KCHP93] E. Keogh, S. Chu, D. Hart, and M. Pazzani. Segmenting time series : A survey and novel approach. *World Scientific Publishing Company*, 1993.

[KdL80] P.M. Kroonenberg and J. de Leeuw. Principal components analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, 45 :69–97, 1980.

[Keh04] Ath. Kehagias. A hidden markov model segmentation procedure for hydrological and environmental time series. *Stochastic environmental research and risk assessment*, 18 :117–130, 2004.

[Kri51] D.G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of Metallurgical and Mining Society of South Africa*, 52 :119–139, 1951.

[Lai04] D. Lai. Estimating the hurst effect and its application in monitoring clinical trials. *Computational Statistics and Data Analysis*, 45 :549–562, 2004.

[Leb69] L. Lebart. Analyse statistique de la contiguïté. *Institut de Statistiques de l'Université de Paris*, 28 :81–112, 1969.

[Leh80] E.L. Lehmann. Efficient likelihood estimators. *The American Statistician*, 34 :233–235, 1980.

[LGA+99] J.-S. Lee, M.R. Grunes, T.L. Ainsworth, L.-J. Du, D.L. Schuler, and S.R. Cloude. Unsupervised classification using polarimetric decomposition and

the complex wishart classifier. *IEEE Transactions on Geoscience and Remote Sensing*, 37 :2249–2258, 1999.

[LMP04] L. Lebart, A. Morineau, and M. Piron. *Statistique exploratoire multidimensionnelle*. Dunod, 2004.

[LOKS86] T. Lundahl, W. Ohley, S. Kay, and R. Siffert. Fractional brownian motion : A maximum likelihood estimator and its application to image texture. *IEEE Transactions on Medical Imaging*, 5 :152–161, 1986.

[Man83] B.B. Mandelbrot. *The Fractal Geometry of Nature*. Freeman, New York, USA, 1983.

[Mar80] K.V. Mardia. Tests of univariate and mutivariate normality. *In P.R. Krishnaiah (Ed.), Handbook of statistics*, 1 :279–320, 1980.

[Mat65] G. Matheron. *La théorie des Variables Régionalisées et ses Applications*. Masson, Paris, 1965.

[MG01] T.B. Moeslund and E. Granum. A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding*, 81 :231–268, 2001.

[MN68] B.B. Mandelbrot and J.W. Van Ness. Fractional brownian motions, fractional noises and applications. *SIAM Review*, 10 :422–437, 1968.

[Moh91] B. Mohar. The laplacian spectrum of graphs. *In Graph Theory, Combinatorics and Applications*, 2 :871–898, 1991.

[Mor48] P.A.P. Moran. The interpretation of statistical maps. *Journal of the Royal Statistical Society, series B (Methodological)*, 10 :243–251, 1948.

[MW03] J. Masaro and C.S. Wong. Wishart distributions associated with matrix quadratic forms. *Journal of Multivariate Analysis*, 85 :1–9, 2003.

[NJW01] A.Y. Ng, M. Jordan, and Y. Weiss. On spectral clustering : Analysis and an algorithm. *In Advances in Neural Information Processing Systems 14*, 2001.

[Nor94] I. Norros. A storage model with self-similar input. *Queueing Systems*, 16 :387–396, 1994.

[Pel98] R.F. Peltier. *Processus stochastiques fractals avec applications en finance*. PhD thesis, Université Paris VI, 1998.

[Per01] F. Perales. Human motion analysis and synthesis using computer vision and graphics techniques. state of art and applications. *In Proceedings of the fifth World Multi-Conference on Systemics, Cybernetics and Informatics*, 2001.

[PHBJ+01] E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren, and A. Bonami. nth order fractional brownian motion and fractional gaussian noise. *IEEE Transactions on Signal Processing*, 49 :1049–1059, 2001.

[PLV94] R.F. Peltier and J. Lévy-Véhel. A new method for estimating the parameter of fractional brownian motion. Research report 2396, INRIA Rocquencourt, 1994.

[PLV95] R.F. Peltier and J. Lévy-Véhel. Multifractional brownian motion : Definition and preliminary results. Research report 2645, INRIA Rocquencourt, 1995.

[PP06] K.B. Petersen and M.S. Pedersen. The matrix cookbook. *http ://www.est.uc3m.es/daejin/files/MatrixCookBook.pdf*, 2006.

[PPLV02] B. Pesquet-Popescu and J. Lévy-Véhel. Stochastic fractal models for image processing. *Signal Processing Magazine, IEEE Publication*, 19 :48–62, 2002.

[PST00] D. Potts, G. Steidl, and M. Tasche. *Fast Fourier transforms for nonequispaced data : A tutorial*, chapter 12, pages 253–274. Birkhäuser, 2000.

[PY86] J. Pitman and M. Yor. Asymptotic laws of planar brownian motion. *The Annals of Probability*, 14 :733–779, 1986.

[Rao64] C.R. Rao. The use and interpretation of principal component analysis in applied research. *Sankhya serie A*, 26 :329–357, 1964.

[RE76] P. Robert and Y. Escoufier. A unifying tool for linear multivariate statistical methods. *Applied Statistics*, 25 :257–265, 1976.

[RMNFF07] A. Reigber, M.Jäger, M. Neumann, and L. Ferro-Famil. Polarimetric fuzzy K-means classification with consideration of spatial context. *Proceedings of POLinSAR*, 2007.

[Row03] D.B. Rowe. *Multivariate Bayesian statistics : models for source separation and signal unmixing*. Chapman and Hall/CRC, 2003.

[RS97] J.O. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Series in Statistics, 1997.

[RV00] G.V. Raz and B. Van Veen. Blind equalization and identification of nonlinear and IIR systems- a least squares approach. *IEEE transactions on Signal Processing*, 48 :192–200, 2000.

[RW06] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[Sab87] R. Sabatier. *Methodes factorielles en analyse des donnees : approximations et prise en compte de variables concomitantes*. PhD thesis, Université de Montpellier 2, 1987.

[Sam69] Jr. J. Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers, C–18*, pages 401–409, 1969.

[Sap90] G. Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 1990.

[Sca82] J.D. Scargle. Studies in astronomical time series. II. statistical aspects of spectral analysis of unevenly spaced data. *Astrophysical Journal*, 263 :835–853, 1982.

[Sch78] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6 :461–464, 1978.

[Sel95] F. Sellan. Synthèse de mouvements browniens fractionnaires à l'aide de la transformation par ondelettes. *Compte-rendu de l'Académie des Sciences, Série I Mathématiques*, 321 :351–358, 1995.

[SGW03] N. Scafetta, L. Griffin, and B.J. West. Hölder exponent spectra for human gait. *Physica A*, 328 :561–583, 2003.

[SNC03] H. Skriver, A.A. Nielsen, and K. Conradsen. Evaluation of the wishart test statistics for polarimetric SAR data. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 2 :699–701, 2003.

[Spi58] F. Spitzer. Some theorems concerning 2-dimensional brownian motion. *Transactions of the American Mathematical Society*, 87 :187–197, 1958.

[SSM96] B. Schölkopf, A. Smola, and K.-R. Müller. Non linear component analysis as a kernel eigenvalue problem. Research report 44, Max-Planck-Institute für biologische Kybernetik, 1996.

[ST94] G. Samorodnitsky and M.S. Taqqu. *Stable non-Gaussian random processes : Stochastic models with infinite variance*. Chapman and Hall, New York, USA, 1994.

[ST96] D.A. Spielman and S.-H. Teng. Spectral partiting works : Planar graphs and finite element meshes. Research report UCB/CSD-96-898, EECS Department, University of California, Berkeley, 1996.

[SW96] D.L. Swets and J.J. Weng. Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18 :831–836, 1996.

[SWM+04] I.S. Stanimirova, B. Walczak, D.L. Massart, V. Simeonov, C.A. Saby, and E. Di Crescenzo. STATIS, a three-way method for data analysis. Application to environmental data. *Chemometrics and Intelligent Laboratory Systems*, 73 :219–233, 2004.

[TCC02] J. Thioulouse, D. Chessel, and S. Champely. Multivariate analysis of spatial patterns : a unified approach to local and global structures. *Environmental and Ecological Statistics*, 27 :1–14, 2002.

[TM04] F. Thullier and H. Moufti. Multi-joint coordination in ballet dancers. *Neuroscience Letters*, 369 :80–84, 2004.

[Tuc58] L.R. Tucker. Determination of parameters of a functional relationship by factor analysis. Psychometrika :19–23, 1958.

[Tuc64] L.R. Tucker. The extension of factor analysis to three-dimensional matrices. *Contribution to Mathematical Psychology, H. Gulliksen and N. Frederiksen*, pages 109–127, 1964.

[Tur96] P. Turchin. Fractal analyses of animal movement : a critique. *Ecology*, 77 :2086–2090, 1996.

[Vis05] M. Visani. *Vers de nouvelles approches discriminantes pour la reconnaissance automatique de visages*. PhD thesis, Institut National des Sciences Appliquées de Lyon, 2005.

[Vit67] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13 :260–269, 1967.

[VV00] A. Vidacs and J.T. Virtamo. Parameter estimation of geometrically sampled fractional brownian traffic. *Proceedings of INFOCOM*, 3, 2000.

[War85] D. Wartenberg. Multivariate spatial correlation : a method for exploratory geographical analysis. *Geographical Analysis*, 17 :263–283, 1985.

[WC94] A. Wood and G. Chan. Simulation of stationary Gaussian processes in $[0,1]^d$. *Journal of computational and graphical statistics*, 3 :409–432, 1994.

[Wei99] Y. Weiss. Segmentation using eigenvectors : a unifying view. *In proceedings IEEE, International Conference on Computer Vision (ICCV)*, pages 975–982, 1999.

[Wis28] J. Wishart. The generalized product-moment distribution in samples from multivariate population. *Biometrika, 20A*, pages 127–139, 1928.

[Woo66] A. Wook. A note on square roots of positive operators. *SIAM Review*, 8 :100–102, 1966.

[Yue78] C.K. Yuen. Quadratic windowing in the segment averaging method for power spectrum computation. *Technometrics*, 20 :195–200, 1978.

[YZ02] Y.-K. Yu and Y.-C. Zhang. On the anti-wishart distribution. *Physica A*, 312 :1–22, 2002.

[YZFY04] J. Yang, D. Zhang, A.F. Frangi, and J.Y. Yang. Two-dimensional PCA : A new approach to appearance-based face representation and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26 :131–137, 2004.