



Recalage de signaux et analyse de variance fonctionnelle par ondelettes. Applications au domaine biomédical

Jérémie Bigot

► To cite this version:

Jérémie Bigot. Recalage de signaux et analyse de variance fonctionnelle par ondelettes. Applications au domaine biomédical. Autre [q-bio.OT]. Université Joseph-Fourier - Grenoble I, 2003. Français.
NNT : . tel-00003344v2

HAL Id: tel-00003344

<https://theses.hal.science/tel-00003344v2>

Submitted on 29 Jan 2004

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 DE L'UNIVERSITÉ JOSEPH
FOURIER

SPÉCIALITÉ
MATHÉMATIQUES APPLIQUÉES

Présentée par
JÉRÉMIE BIGOT

Pour obtenir le titre de

DOCTEUR DE L'UNIVERSITÉ JOSEPH FOURIER

Sujet de la thèse :

RECALAGE DE SIGNAUX ET ANALYSE DE VARIANCE
FONCTIONNELLE PAR ONDELETTES
APPLICATIONS AU DOMAINE BIOMÉDICAL

Soutenue le 2 septembre 2003 devant le jury composé de

- Madame Valérie Perrier, ENSIMAG-INPG (Président)
- Monsieur Steve Marron, UNC-Chapel Hill (Rapporteur)
- Monsieur Georges Oppenheim, Paris-Sud (Rapporteur)
- Monsieur Stéphane Mallat, CMAP-Ecole Polytechnique (Examinateur)
- Monsieur Anestis Antoniadis, Grenoble I (Directeur de thèse)

Thèse préparée au sein du Laboratoire de Modélisation et Calcul, Equipe SMS.

Remerciements

Je tiens à exprimer mes plus vifs remerciements à Steve Marron (Professeur, UNC-Chapel Hill) et Georges Oppenheim (Professeur, Paris-Sud) pour avoir accepté la lourde tâche que représente la lecture de ce document et de l'avoir rapporté. Je souhaite également les remercier pour les nombreuses et intéressantes remarques qui ont permis l'achèvement d'une nouvelle version. Je souhaite témoigner ma profonde gratitude envers Stéphane Mallat (Professeur, CMAP-Ecole Polytechnique) et Valérie Perrier (Professeur, ENSIMAG-INPG) pour avoir accepté de faire partie des membres du jury. Les discussions que nous avons pu avoir durant ma thèse ont été très formatrices et je les remercie d'avoir dévoilé un intérêt particulier et un esprit critique pour le travail que j'ai effectué.

Je tiens également à remercier les professeurs James O.Ramsay (McGill University), Alois Kneip (Johannes-Gutenberg-Universität of Mainz), Theo Gasser (University of Zurich), Ian L Dryden (School of Mathematical Sciences, University of Nottingham) et Laurent Younes (CMLA, ENS Cachan) pour les discussions particulièrement enrichissantes que nous avons eues, qui m'ont permis de prendre un recul significatif par rapport à mon travail.

Je tiens également à exprimer ma plus vive reconnaissance à Chris Glasbey (BIOSS, Edinburgh) pour m'avoir accueilli pendant les derniers mois de ma thèse et pour avoir accepté que je puisse travailler dans le laboratoire qu'il dirige. Je remercie également Graham Horgan (BIOSS, Edinburgh) de l'intérêt qu'il a porté aux méthodes que j'ai développées pendant ma thèse. Mes remerciements vont également à Yolanda Munoz Maldonado pour m'avoir permis d'utiliser les données de la ganglioside.

Enfin, je souhaite remercier très chaleureusement Anestis Antoniadis (Professeur, Grenoble I) qui a bien voulu m'encadrer pendant ce travail. Je ne sais pas trop comment résumer en quelques phrases tout le plaisir et le privilège que j'ai eus de pouvoir travailler avec lui aussi bien d'un point de vue scientifique que pour l'amitié qu'il m'a témoignée. Je tiens à le remercier tout particulièrement pour l'investissement qu'il a bien voulu mettre dans l'encadrement de cette thèse en étant toujours disponible pour répondre à mes très nombreuses interrogations scientifiques, et pour m'avoir donné la possibilité de collaborer avec de nombreux chercheurs aussi bien en France qu'à l'étranger. Sa bonne humeur et sa passion pour la recherche ont également été toujours un élément moteur, et je lui suis particulièrement reconnaissant pour ses très nombreux encouragements et pour avoir cru en mon travail.

Je tiens aussi à remercier mes parents pour l'aide qu'ils m'ont apportée les quelques jours qui ont précédé la soutenance, en particulier pour l'impression de la thèse et la préparation du pot d'après soutenance. La rédaction de ce document ne s'est pas faite sans quelques moments de découragement et d'énerverment, et je ne sais pas trop comment remercier en quelques mots à quel point la présence de Marie a été encourageante et m'a aidé pendant les derniers mois de ce travail...

Je remercie enfin toutes les personnes qui m'ont entouré pendant cette thèse (les thésards du LMC, des laboratoires de l'IMAG, du laboratoire BIOSS, ma soeur Sarah, et tous les amis de Grenoble, Tours et ailleurs...) et tout particulièrement ceux qui m'ont fait le plaisir de venir assister à la soutenance.

Résumé

Cette thèse porte sur le recalage de signaux à partir de l'alignement de leurs landmarks, pour la comparaison d'ensembles de courbes ou d'images. Après une revue des techniques de recalage qui existent dans la littérature, une approche nonparamétrique est proposée pour estimer les landmarks d'une fonction 1D bruitée à partir des lignes de maxima d'ondelettes et de zero-crossings de sa transformée continue en ondelettes. Un nouvel outil, l'intensité structurelle, est introduit pour représenter les positions des points caractéristiques d'une courbe sous forme d'une densité de probabilité. Cette méthode conduit à une nouvelle technique de mise en correspondance automatique des landmarks de deux fonctions. L'approche envisagée dans le cas 1D est étendue au cas 2D en utilisant des décompositions en wedgelets/platelets pour détecter les contours d'une image. De nombreuses simulations et des problèmes réels d'analyse de variance fonctionnelle servent d'illustration des méthodes proposées.

Mots-clés : recalage automatique de signaux, landmarks, transformée continue en ondelettes, zero-crossings, maxima d'ondelettes, wedgelets, platelets, détection de contours, représentation temps-échelle, estimation nonparamétrique.

Abstract

This thesis focuses on landmark-based registration of curves and images for the comparison of multiple sets of signals. After a detailed review of some important registration techniques, a nonparametric approach is proposed to estimate the landmarks of a 1D function observed with noise, from the wavelet maxima and the zero-crossings lines of its continuous wavelet transform. A new tool, the structural intensity, is introduced to represent the locations of the typical features of a curve via a probability density function. This method yields to a new technique to automatically put into correspondence the landmarks of two functions. The approach developed for 1D signals is extended on a 2D setting by using wedgelets/platelets decompositions to detect the edges of an image. The usefulness of these methods is documented by various simulations and applications to real examples involving functional analysis of variance.

Keywords : automatic landmark-based registration, continuous wavelet transform, zero-crossings, wavelet maxima, wedgelets, platelets, edge detection, scale-space representation, nonparametric estimation.

Table des matières

1	Introduction	11
1.1	Recalage de signaux	14
1.2	Approches temps-échelle pour détecter les landmarks d'un signal	17
1.3	Plan de la thèse	19
2	Alignement de courbes et d'images	23
2.1	Définition du recalage de deux signaux	23
2.2	Recalage de courbes	25
2.2.1	Alignement à partir de landmarks	25
2.2.2	Alignement global	30
2.3	Recalage d'images	31
2.3.1	Alignement à partir de landmarks	32
2.3.1.1	Transformations paramétriques	35
2.3.1.2	Transformations nonparamétriques	35
2.3.2	Mise en correspondance automatique de deux ensembles de landmarks	41
2.3.3	Alignement global	44
2.3.4	Active Shape Models - Active Appearance Models	46
3	Une approche par ondelettes pour détecter les landmarks d'un signal 1D	49
3.1	Un modèle de régression nonparamétrique	50
3.2	Ondelettes et structure d'un signal	52
3.2.1	La transformée en ondelettes continue	52
3.2.2	Zero-crossings d'une transformée en ondelettes	53
3.2.3	Maxima d'ondelettes	53
3.2.4	Quelques propriétés des lignes de maxima d'ondelettes et de zero-crossings	56
3.2.4.1	Les bonnes propriétés temps-échelles de la Gaussienne et de ses dérivées	56
3.2.4.2	Convergence asymptotique des ondelettes B-Spline	61
3.2.4.3	Intensité structurelle des maxima d'ondelettes et des zero-crossings	63
3.3	Modèle de régression nonparamétrique et modèle de bruit blanc	68

3.4	Estimation des maxima d'ondelettes	76
3.4.1	Détection des maxima d'ondelettes	78
3.4.1.1	Choisir un seuil	78
3.4.1.2	Estimation des maxima d'ondelettes quand α est inconnu	82
3.4.2	Plusieurs singularités	83
3.5	Estimation des zero-crossings	84
3.5.1	Détection des zero-crossing	85
3.5.2	Convergence en probabilité	88
3.5.3	Distribution asymptotique des estimateurs	91
3.6	Estimation des landmarks	93
3.7	Simulations	98
3.7.1	Estimation des lignes de zero-crossings	99
3.7.2	Estimation des lignes de maxima d'ondelettes	103
3.7.3	Bagging pour supprimer les estimations erronées	106
3.7.4	Résultats des simulations	113
3.7.5	Quelques exemples réels	123
3.7.5.1	Détection de transitions brutales	123
3.7.5.2	Détection de maxima locaux significatifs	123
4	Recalage automatique de courbes	131
4.1	Mise en correspondance automatique de deux ensembles de landmarks	132
4.1.1	Approches déjà existantes	132
4.1.2	Une méthode automatique	134
4.1.2.1	Alignement des intensités structurelles par DTW	136
4.1.3	Implémentation pratique et simulations	142
4.1.3.1	Discrétisation du problème variationnel	142
4.1.3.2	Un exemple réel	143
4.1.3.3	Simulations	145
4.2	Recalage d'un ensemble de courbes	155
4.2.1	Alignement des données de la ganglioside	162
4.2.1.1	Pré-traitement des données	165
4.2.1.2	Recalage à partir des maxima	165
4.2.2	Alignement des données d'orthèse	167
4.3	Analyse de variance fonctionnelle	169
4.3.1	Tests d'hypothèses dans un modèle de FANOVA	175
4.3.2	Application à deux exemples réels	177
4.3.2.1	Données d'orthèses	177
4.3.2.2	Données de la ganglioside	179
4.4	Discussions	181

5 Détection de contours dans une image et alignement de signaux 2D	187
5.1 De nouvelles représentations adaptées aux contours	191
5.1.1 Détection de contours avec des wedgelets	193
5.1.2 Détection de contours avec des platelets	198
5.2 Combiner la détection de contours et l'alignement de landmarks	200
5.2.1 Alignement de wedgelets	200
5.2.2 Application à des images réelles	203
5.3 Discussions	207
6 Conclusion et perspectives	213
6.1 Alignement de courbes	213
6.2 Alignement d'images	215

Chapitre 1

Introduction

L'objectif de cette thèse est le développement d'outils nécessaires à la comparaison de plusieurs ensembles de signaux. Lorsque que l'on étudie un processus biologique ou physique chez différents sujets, on remarque généralement que les types de courbes ou d'images observées présentent des caractéristiques similaires. Un problème important consiste alors à déterminer la forme typique du processus observé, ou bien à tester s'il existe des différences significatives parmi deux sous-ensembles de sujets. Par exemple, lors de l'observation d'un échantillon de courbes de croissance chez des garçons et chez des filles, il peut s'agir de savoir quelle courbe représente le mieux la forme typique du processus de croissance, et si sa forme caractéristique présente des variations entre les garçons et les filles (voir Gasser et Kneip [39]). L'estimation des variations au sein d'une population à partir de données fonctionnelles, peut se résoudre à partir de l'*Analyse de Variance Fonctionnelle* (FANOVA) qui généralise au cas de fonctions, les techniques d'*Analyse de Variance* (ANOVA) qui est l'un des outils les plus utilisés en statistique appliquée, pour déterminer comment les variations d'une variable observée peuvent être expliquées par l'observation d'autres quantités. Toutefois, afin de pouvoir comparer des objets présentant des caractéristiques similaires, il est nécessaire de trouver un système référentiel commun pour les représenter. En effet, les courbes ou les images à analyser peuvent ne pas être observées sur les mêmes domaines. De plus, il peut exister des variations locales ou globales entre ces signaux (dues par exemple au processus d'acquisition) qu'il peut être nécessaire de corriger, car ceux-ci peuvent conduire à des conclusions erronées. Le travail effectué dans cette thèse vise donc à construire un ensemble d'outils pour *synchroniser* un ensemble de signaux afin de les représenter dans un référentiel bien adapté à la comparaison de ces derniers. Il s'agit d'une étape de pré-conditionnement de données fonctionnelles généralement nécessaire avant l'application de toute autre procédure d'inférence statistique. Dans ce qui suit, la *synchronisation* d'un ensemble de signaux sera également appelée *alignement* ou *recalage* de fonctions. Les signaux considérés dans cette thèse seront des fonctions unidimensionnelles (1D, courbes) ou bidimensionnelles (2D, images). Afin d'illustrer plus précisément le problème de la synchronisation de signaux, nous proposons de présenter deux exemples réels qui seront étudiés plus en détail par la

suite :

- **Exemple 1D :** Les fonctions représentées dans la Figure 1.1 correspondent à une partie de l'ensemble de courbes étudié dans Munoz Maldonado *et. al.* [71]. Ces courbes représentent la distribution de la ganglioside dans les tissus cérébraux de groupes de rats d'âges distincts. Le but de l'étude conduite dans [71] est de comparer la distribution de la ganglioside, représentée par un profil de densité, selon l'âge des rats, dans 3 zones différentes du cerveau : Medulla (MD), Locus Coeruleus (LC) et Hippocampus (HY). Une description précise de la méthode d'acquisition de ces données peut être trouvée dans [71].

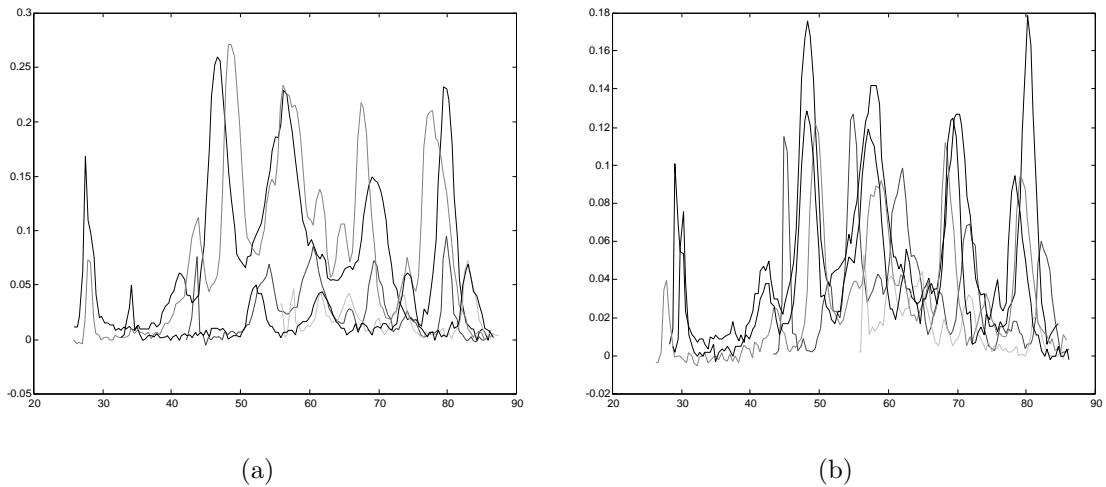


FIG. 1.1: Profil de densité de la ganglioside dans la région Medulla : (a) 5 Vieux rats (b) 5 Jeunes rats.

- **Exemple 2D :** La Figure 1.2 montre des images photographiques de deux espèces de poissons (haddock et whiting) que l'on souhaite discriminer. Ces images font partie d'un ensemble de données plus large qui se compose de 10 images de haddocks, et 10 images de whitings, analysé dans Glasbey et Mardia [43]. Ces données peuvent obtenue à l'URL :

<http://www.blackwellpublishers.co.uk/rss/>. Le but de la comparaison de cet ensemble d'images est de déterminer la forme caractéristique de chaque espèce, et de pouvoir quantifier les différences ou les similitudes entre les poissons observés.

On peut remarquer que les courbes de la Figure 1.1 sont observées sur des intervalles différents. Afin de pouvoir les comparer, il est nécessaire de représenter ces courbes sur un support commun. Supposons que l'on observe un ensemble de m courbes f_1, f_2, \dots, f_m sur m intervalles différents $[a_1, b_1], [a_2, b_2], \dots, [a_m, b_m]$ et que l'on souhaite déterminer leur forme moyenne \bar{f} . L'idée la plus simple, pour représenter ces courbes sur un même intervalle (par exemple $[0, 1]$), consiste à définir



(a)



(b)



(c)



(d)

FIG. 1.2: Images photographiques de deux espèces de poissons : (a) Haddock 0 (b) Haddock 1 (c) Whiting 0 (d) Whiting 1 (les images sont de taille 300×500 pixels).

les fonctions : $\tilde{f}_i(t) = f_i(a_i + t(b_i - a_i))$ pour $t \in [0, 1]$ et $i = 1, \dots, m$. Pour obtenir la forme caractéristique d'une courbe de cette ensemble, il semble alors naturel de définir $\tilde{f}(t) = \frac{1}{m} \sum_{i=1}^m \tilde{f}_i(t)$, $t \in [0, 1]$. Un exemple de cette approche, que nous appellerons **recalage linéaire**, est donné dans la Figure 1.3 avec des données simulées et $m = 5$. La courbe moyenne ainsi obtenue (Figure 1.3c) ne donne pas une représentation satisfaisante de la forme typique d'une courbe de la Figure 1.3a : toutes ces courbes sont en effet clairement composées de deux sinusoïdes séparées par un “saut”, précédé d'un maximum global, ce qui n'est plus le cas de la courbe de la Figure 1.3c (le saut est moins important et il est précédé d'un maximum local).

De même, les images de la Figure 1.2 ont déjà été alignées globalement mais il existe encore des petites différences locales entre les positions des poissons qui doivent être corrigées. Ces différences sont dues à la façon dont les poissons ont été placés avant d'être photographiés. Elles doivent donc être supprimées si l'on souhaite, par exemple, comparer la forme des poissons (dans ce cas là, il semble essentiel que les contours extérieurs des poissons soient alignés, ou se trouvent relativement à la même position).

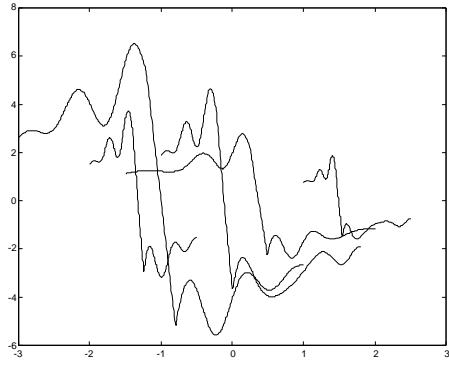
1.1 Recalage de signaux

Le problème de la synchronisation d'un ensemble de signaux se situe dans le contexte du recalage de structures 1D ou 2D. Formellement, le problème de l'alignement de deux signaux f_1 et f_2 (que nous considérerons par la suite comme des fonctions réelles 1D ou 2D), consiste à déterminer une transformation u telle que :

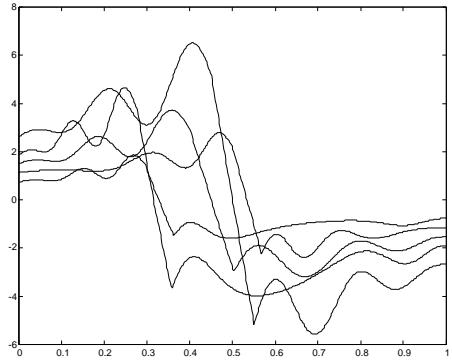
- $f_1 \approx f_2 \circ u$: la transformation (ou déformation) u doit être telle que l'écart entre les fonctions f_1 et $f_2 \circ u$ soit le plus petit possible au sens d'une certaine mesure de dissimilarité entre deux signaux.
- u soit suffisamment régulière afin que les caractéristiques de la fonction f_2 soient conservées avec la déformation. Généralement, il est souhaitable que u soit inversible et différentiable.

Lorsque l'on souhaite recaler un ensemble de m signaux f_1, \dots, f_m ($m \geq 2$), le but est de trouver m transformations u_1, \dots, u_m suffisamment régulières, telles que les fonctions déformées $f_1 \circ u_1, \dots, f_m \circ u_m$ soient alignées de manière optimale par rapport à une certaine mesure de dissimilarité. Il existe une très vaste littérature sur ce sujet. Des revues détaillées et de très nombreuses références peuvent être trouvées dans Younes [95], Glasbey et Mardia [42], Brown [10], et Lester et Arridge [55]. Parmi différentes applications du recalage de signaux nous pouvons citer : la reconnaissance de formes ou de visages, la construction d'atlas du cerveau, le recalage d'images multi-modales (couplage imagerie fonctionnelle et anatomique) ou la caractérisation des variations au sein d'une population qui est l'application envisagée dans cette thèse.

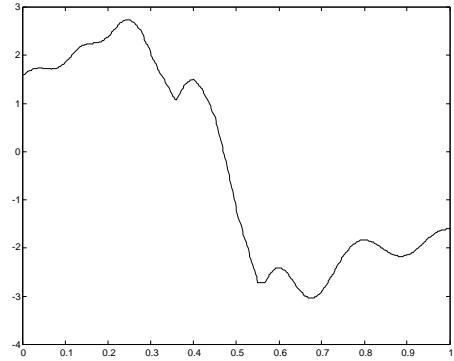
Deux grandes classes de méthodes peuvent être distinguées pour le recalage de signaux :



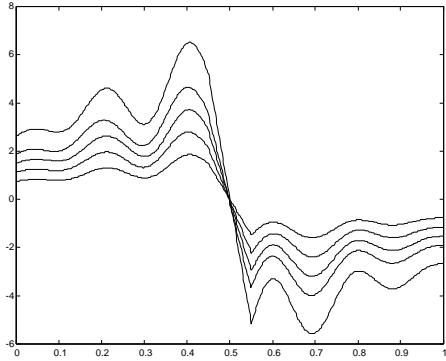
(a)



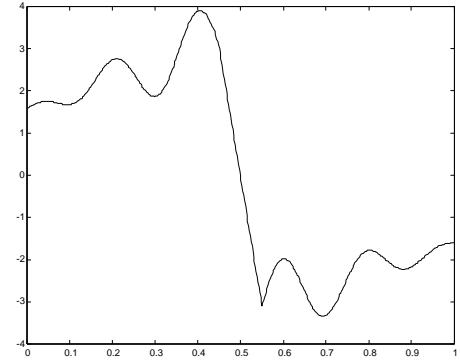
(b)



(c)



(d)



(e)

FIG. 1.3: Estimation de la forme moyenne d'un ensemble de courbes : (a) 5 courbes présentant des caractéristiques similaires, observées sur des intervalles différents. (b) Recalage linéaire (c) Moyenne des courbes après recalage linéaire (d) Recalage non-linéaire (e) Moyenne des courbes après recalage non-linéaire.

- **Alignement à partir de landmarks** : nous désignerons par landmarks un ensemble de points caractéristiques de la forme d'un signal. Pour une courbe il peut s'agir de la position de maxima, de minima, de points d'inflexion ou encore de singularités. Pour une image représentant un visage, les positions des yeux, des oreilles, du contour de la bouche sont des exemples classiques de landmarks en 2D. Les méthodes d'alignement à partir de landmarks visent à calculer des transformations qui mettent en correspondance les landmarks qui sont communs à un ensemble de signaux devant être recalés. Les landmarks des signaux qui caractérisent les mêmes structures, se trouvent ainsi à la même position après déformation.
- **Alignement global** : ces méthodes visent à déterminer des transformations suffisamment régulières qui alignent l'ensemble des points d'un signal au sens d'une certaine distance entre deux courbes ou deux images (par exemple $\int(f_1 - f_2 \circ u)^2$).

Dans cette thèse, nous avons choisi de développer de nouvelles méthodes d'*alignement à partir de landmarks*. Un exemple de recalage de signaux à partir de leurs landmarks est donné dans la Figure 1.3. Chaque courbe de la Figure 1.3a possède 5 maxima et 5 minima. Pour chacune de ces courbes, une transformation non linéaire a été calculée pour synchroniser ces 10 extrema sur les même positions. Le résultat de la déformation de ces 5 courbes à l'aide de ces transformations est donné dans la Figure 1.3d. Le calcul de la moyenne des courbes après un recalage *non linéaire* donne une bien meilleure représentation de la forme typique d'une courbe que celle obtenue avec un recalage *linéaire* (comparer les Figures 1.3c et 1.3e).

L'alignement de signaux à partir de landmarks comporte plusieurs étapes successives :

1. Définition des points caractéristiques d'un signal qui seront utilisés comme landmarks.
2. Extraction des landmarks d'un signal à partir d'une série d'observations de ce dernier. L'estimation des landmarks d'un signal est généralement compliquée par la présence de bruit dont les fluctuations peuvent produire des structures qui ne correspondent pas à de "vrais" points caractéristiques du signal.
3. Détermination des landmarks qui sont communs à un ensemble de signaux et des landmarks qui doivent se correspondre.
4. Calcul des transformations qui alignent les landmarks qui doivent être mis en correspondance.
5. Déformation des signaux à l'aide des transformations obtenues par alignement des landmarks.

Les deux points les plus délicats de ces méthodes sont les étapes 2 et 3. Généralement, elles sont effectuées manuellement ce qui peut être fastidieux quand le nombre de signaux à analyser est important. Au cours de cette thèse, nous avons donc cherché à construire des méthodes automatiques pour estimer les landmarks communs à un

ensemble de signaux (éventuellement observés avec du bruit) et les mettre en correspondance.

En pratique, les signaux observés sont généralement bruités. Ainsi, une fois que les déformations pour synchroniser un ensemble de signaux ont été déterminées, il est nécessaire d'estimer ces signaux avant de les aligner. Les méthodes à *base d'ondelettes*, qui permettent une représentation creuse d'une fonction, se sont révélées être des outils puissants pour le débruitage de signaux. Tout au long de cette thèse, nous utiliserons donc des estimateurs à base d'ondelettes lorsqu'il sera nécessaire de débruiter un signal. Toutefois, nous ne donnerons pas une description précise de ces outils. Une revue détaillée des méthodes existantes, ainsi qu'une comparaison de leur performances numériques sur de nombreux types de fonctions peuvent être trouvées dans Antoniadis, Bigot et Sapatinas [3].

1.2 Approches temps-échelle pour détecter les landmarks d'un signal

Lorsque l'on regarde la forme d'un signal, on peut généralement distinguer des structures globales qui nous permettent d'identifier grossièrement le type d'objet qu'il représente. Par exemple, on reconnaît immédiatement que les images de la Figure 1.2 sont des photographies de poissons d'après la forme du contour extérieur de ces derniers. Toutefois, il existe des structures plus fines et plus localisées dans un signal qui permettent de le décrire plus précisément. Ainsi, en inspectant plus attentivement la forme des nageoires et de la queue des poissons de la Figure 1.2, on peut noter des différences significatives entre les haddocks et les whittings. Cette constatation sous-entend l'idée que l'information dans un signal peut être hiérarchisée en la représentant à différents niveaux de détails allant d'une approximation grossière et globale, à des représentations plus précises et plus localisées. Le principe d'une telle représentation hiérarchique est au cœur de l'*analyse temps-échelle* des signaux : le problème de la détermination des structures d'un signal peut se résoudre plus facilement si celles-ci sont décomposées à différents niveaux d'approximations qui permettent de distinguer les caractéristiques globales, des caractéristiques plus locales d'un signal. L'*analyse temps-échelle* de signaux a été largement utilisée en vision par ordinateur (voir le livre de Lindeberg [58] pour une revue détaillée). Les propriétés mathématiques d'approximation et d'estimation des structures d'un signal par des *décompositions multi-échelles* ont été étudiées à partir des années 1980 dans les travaux de Meyer [66], Daubechies [28] et Mallat [63] par l'introduction des *ondelettes*.

Afin d'expliquer le principe de caractérisation des structures d'un signal par une approche temps-échelle, nous présentons brièvement la *transformée en ondelettes continue* (CWT) d'un signal 1D. Une ondelette ψ est une fonction de moyenne nulle

(nous supposerons que ψ est réelle) :

$$\int_{-\infty}^{+\infty} \psi(u) du = 0.$$

qui est dilatée par un paramètre d'échelle s , et translatée par x :

$$\psi_{x,s}(u) = \frac{1}{\sqrt{s}} \psi\left(\frac{u-x}{s}\right),$$

La transformée en ondelettes d'un signal f à l'échelle s et la position x est alors définie par :

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u) \psi_{x,s}(u) du.$$

Une transformée en ondelettes représente la "corrélation" entre le signal f et les atomes $\psi_{x,s}$. En appliquant la formule de Parseval, elle peut être également écrite :

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u) \psi_{x,s}(u) du = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \hat{f}(\omega) \hat{\psi}_{x,s}(\omega) d\omega,$$

où la transformée de Fourier d'un signal est définie par :

$$\hat{f}(\omega) = \int_{-\infty}^{+\infty} f(t) e^{-i\omega t} dt.$$

Le coefficient d'ondelette $W_s(f)(x)$ dépend donc des valeurs de $f(t)$ et $\hat{f}(\omega)$ dans le domaine temps-fréquence où l'énergie de $\psi_{x,s}$ et $\hat{\psi}_{x,s}$ est concentrée. Les coefficients d'ondelettes de grandes amplitudes, à différentes échelles et positions, sont donc reliés à la structure temps-fréquence d'un signal. Etant donné que ψ est de moyenne nulle, $W_s(f)(x)$ mesure la variation de f dans un voisinage du point x dont la taille est proportionnelle à s . Les transitions brutales dans un signal créent donc des coefficients d'ondelettes de grandes amplitudes. Mallat et Hwang [64] ont montré que, pour des fonctions ψ bien choisies, la propagation des grands coefficients d'ondelettes quand l'échelle s tend vers zéro permet de caractériser la régularité locale d'un signal. En particulier, les singularités d'une fonction peuvent être détectées en suivant la propagation le long des échelles des maxima locaux d'une transformée en ondelettes. De même, la propagation des zero-crossings de $x \mapsto W_s(f)(x)$ quand l'échelle s tend vers zéro, permet de détecter les extrema ou les points d'inflexion de f selon le type d'ondelette choisie. Les landmarks d'un signal 1D peuvent donc être caractérisés par la propagation des maxima en module et des zero-crossings de sa transformée continue en ondelettes. Un exemple de décomposition en ondelette à différentes échelles d'un signal 1D, comportant une singularité au point $x = 0.4$ et différents extrema, est donné dans la Figure 1.4, où ψ est la dérivée d'une Gaussienne i.e. $\psi = \theta'$ avec :

$$\theta(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}},$$

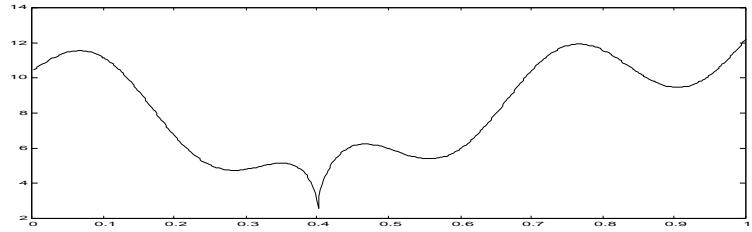
Les coefficients d'ondelettes de grande amplitude se concentrent effectivement autour de la singularité au point $x = 0.4$ quand s tend vers zéro (voir Figure 1.4b).

La transformée en ondelettes continue peut être étendue au cas bidimensionnel. Etant donné que les contours d'une image peuvent être définis comme les points où l'intensité de l'image présente des variations brutales, les coefficients de grandes amplitudes d'une transformée en ondelettes 2D, se concentrent autour de ceux-ci quand le niveau de résolution tend vers zéro (voir Mallat [63]). La propagation des grands coefficients d'une transformée en ondelettes 2D permet donc une caractérisation multi-échelles des points de singularité d'une image.

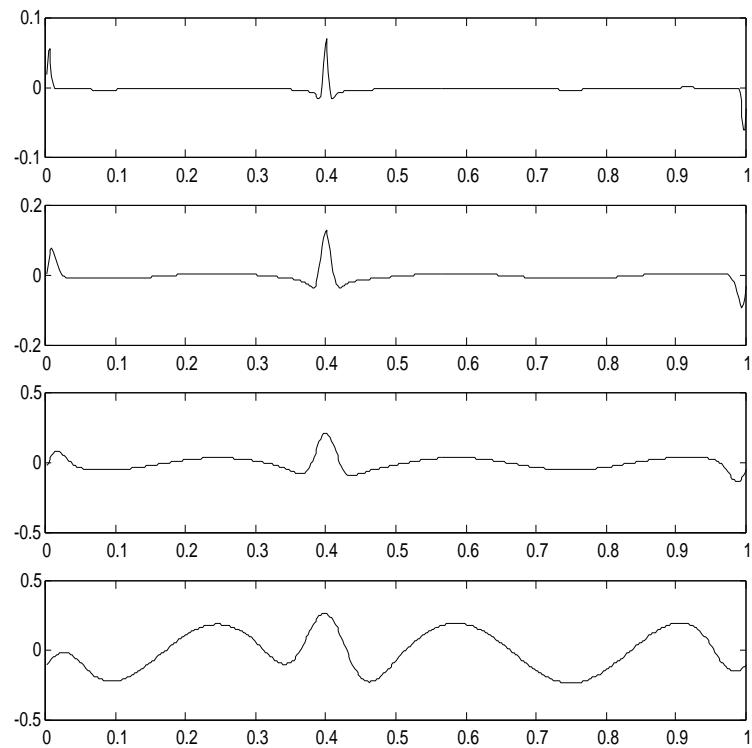
Les contours sont souvent les caractéristiques les plus importantes d'une image. De ce fait, la qualité visuelle du recalage de deux ou plusieurs images dépend principalement de la qualité de l'alignement de leur contours "communs" après déformation. Ceci nous a amené à considérer que les landmarks d'une image doivent être localisés le long des contours de celle-ci. Ainsi, comme pour le cas 1D, les landmarks d'une image peuvent être localisés en suivant la propagation des maxima locaux de sa transformée en ondelettes à différentes échelles. Toutefois, de nombreux coefficients d'ondelettes sont nécessaires pour représenter les contours d'une image (voir Figure 5.1 dans le Chapitre 5), ce qui peut conduire à un nombre de landmarks beaucoup trop important pour être utilisé dans un problème de recalage. Nous avons donc été amené à utiliser de nouvelles représentations multi-échelles spécialement adaptées à la géométrie des images. Dans ces nouvelles représentations, les contours d'une image jouent le rôle des points de singularités d'un signal 1D. Ces approches visent à améliorer les propriétés de compression et d'estimation des algorithmes basés sur des décompositions en ondelettes. Parmi les différentes méthodes récemment développées, nous pouvons citer les travaux de Donoho [29], Candes et Donoho [15], Donoho et Huo [33], Willet et Nowak [94], Le Pennec et Mallat [54], Cohen et Matei [25]. Afin de détecter automatiquement les contours d'une image, éventuellement contaminée par un bruit Gaussien, nous nous sommes plus particulièrement intéressés aux Wedgelets, introduites par Donoho [29], qui permettent une approximation des contours d'une image par des polygones à différents niveaux de résolution.

1.3 Plan de la thèse

Dans le chapitre 2, nous présentons de façon plus formelle le problème du recalage de deux signaux 1D ou 2D. Une revue détaillée des principales approches qui existent dans la littérature est proposée, en distinguant les méthodes d'alignement à partir de landmarks, des méthodes d'alignement global. Une distinction est également faite entre les approches paramétriques et nonparamétriques pour la forme de la fonction de déformation u entre deux signaux. La théorie de l'interpolation Spline est présentée en détail, et l'intérêt ainsi que les limitations de cette technique sont il-



(a)



(b)

FIG. 1.4: Transformée en ondelette continue d'un signal 1D où ψ est la dérivée d'une Gaussienne : (a) Signal $f_1(x) = 2 \sin(6\pi x) + 15|x - 0.4|^{2/5}$, $x \in [0, 1]$. (b) CWT de f_1 à différentes échelles dyadiques : du haut vers le bas $s = 2^{-7}, 2^{-6}, 2^{-5}, 2^{-4}$.

lustrés par quelques simulations. Nous donnons plusieurs exemples de recalage de signaux 1D et 2D, et un algorithme de mise en correspondance automatique de deux ensembles de landmarks (le TPS-RPM, développé par Chui et Rangarajan [22]) est également détaillé. Enfin, au fur et à mesure de la présentation de ces différents algorithmes de recalage, nous expliquons plus précisément la démarche adoptée dans cette thèse en soulignant l'intérêt des nouveaux résultats que nous avons cherchés à obtenir.

Dans le chapitre 3, nous expliquons pourquoi les ondelettes sont un outil puissant pour l'analyse de la structure locale d'un signal 1D. Nous rappelons comment la régularité Lipschitzienne d'une fonction peut être déterminée à partir de la propagation des maxima locaux de sa transformée continue en ondelettes. Nous montrons qu'il est possible de détecter les landmarks d'une fonction en suivant la propagation des zero-crossings et des maxima d'ondelettes quand le niveau de résolution tend vers zéro. Les propriétés temps-échelles des lignes de zero-crossings et des maxima d'ondelettes sont également étudiées en détail en utilisant la notion de *causalité* d'une représentation multi-échelles. Afin d'identifier les limites de ces lignes, un nouvel outil est introduit, l'*intensité structurelle*, qui calcule "la densité" des positions des zero-crossings et des maxima d'ondelettes à différentes échelles. Il est montré que les "modes de cette densité" correspondent à la position des landmarks d'un signal. Une approche nonparamétrique est ensuite envisagée pour estimer les zero-crossings et les maxima d'ondelettes, et donc les landmarks, d'un signal 1D observé avec un bruit Gaussien. Les propriétés asymptotiques des différents estimateurs sont étudiées et illustrées à l'aide de simulations. Combiné avec une méthode de Bagging, nous montrons que l'intensité structurelle est un outil efficace pour supprimer les erreurs d'estimation de landmarks dues aux fluctuations du bruit. Quelques exemples réels nécessitant la détection de landmarks servent d'illustration des méthodes proposées.

Le chapitre 4 est consacré à l'alignement à partir de landmarks de signaux 1D observés avec un bruit Gaussien. A partir de l'estimation de l'intensité structurelle, une nouvelle méthode d'alignement automatique de deux signaux 1D à partir de leurs landmarks est proposée, et quelques propriétés asymptotiques de cette approche sont étudiées. L'intérêt de cette méthode est démontré sur quelques exemples réels et à l'aide de simulations, pour le problème de l'alignement de signaux observés avec un bruit Gaussien. Une étude numérique est également proposée pour comparer notre méthode d'alignement automatique à partir de landmarks avec un algorithme d'alignement global. Cette approche peut également s'étendre au cas de l'alignement de plusieurs signaux. Deux exemples réels (provenant du domaine biomédical) d'estimation des variations au sein d'une population sont étudiés. La méthodologie s'appuie à la fois sur le recalage de signaux 1D et sur une technique d'*analyse de variance fonctionnelle* (FANOVA) par ondelettes récemment développée par Abramovich *et al.* [1].

La détection de contours dans des images est étudiée dans le chapitre 5. Il s'agit

d'une partie plus algorithmique que les chapitres précédents. Nous précisons comment la transformée en ondelettes 1D peut être étendue au cas de signaux 2D. Les contours d'une image peuvent être caractérisés par la propagation des grands coefficients d'ondelettes le long des échelles. Toutefois, nous n'avons pas cherché à étendre les résultats d'estimation des maxima d'ondelettes obtenus pour des signaux 1D au cas d'images éventuellement observées avec du bruit. Pour détecter les contours d'une image, nous avons choisi d'utiliser les Wedgelets, introduites par Donoho [29], qui sont des outils d'analyse multi-échelles adaptés à la géométrie des images. Nous avons également utilisé les Platelets, introduites par Willet et Nowak [94], qui sont une généralisation des Wedgelets. Willet et Nowak [94] ont utilisé cet outil pour estimer des images médicales observées avec un bruit Poissonnien, et nous avons donc adapté les Platelets au cas d'images contaminées par un bruit Gaussien. Nous montrons à l'aide de simulations que ces nouveaux outils sont bien adaptés à la détection de contours pour des images régulières par morceaux qui est le modèle *sous-jacent* de ces techniques. Pour le cas d'images réelles, la détection automatique de contours est plus délicate. Toutefois, nous montrons comment combiner l'approximation d'une image à partir de Platelets/Wedgelets, avec l'algorithme de mise en correspondance automatique de landmarks en 2D développé par Chui et Rangarajan [22]. L'intérêt de cette nouvelle méthode de mise en correspondance des "contours communs" de deux images est illustré par des exemples réels.

Chapitre 2

Alignement de courbes et d'images

Le but de ce chapitre est de présenter le problème de l'alignement de signaux 1D ou 2D. Une revue de quelques méthodes existant dans la littérature est proposée et quelques algorithmes pour calculer la déformation entre deux signaux sont présentés. Nous distinguerons les approches qui mettent en correspondance deux ensembles de landmarks, de celles qui recherchent un alignement global entre deux objets. Dans ce qui suit, nous supposerons qu'une courbe (signal 1D) ou une image (signal 2D) est une fonction $f : \Omega \rightarrow \mathbb{R}$ où Ω désigne soit un sous ensemble de \mathbb{R} dans le cas d'une courbe, soit un sous ensemble de \mathbb{R}^2 dans le cas d'une image. Une définition plus précise de Ω sera donnée en fonction des méthodes de recalage présentées. Dans ce chapitre, nous aborderons uniquement le problème de l'alignement de deux signaux f_1 et f_2 en s'inspirant des notations et de la méthodologie proposées par Younes [95]. L'alignement de plusieurs courbes est discuté dans le chapitre 4 et celui de plusieurs images dans la chapitre 5. Afin de clarifier les notations, nous désignerons parfois par I_1 et I_2 deux images à mettre en correspondance.

2.1 Définition du recalage de deux signaux

L'ensemble Ω est un espace de départ qui peut être éventuellement déformé. Définissons $H(\Omega)$ comme l'ensemble des homéomorphismes sur Ω , c'est à dire l'ensemble des applications $u : \Omega \rightarrow \Omega$ qui sont continues, inversibles et telles que u^{-1} est continu. Etant donné deux objets f_1 et f_2 , le but du recalage de signaux est de trouver une transformation $u \in H(\Omega)$ telle que :

- u est suffisamment lisse.
- $f_1(x) \approx f_2(u(x))$, pour tout $x \in \Omega$.

La première contrainte a pour but de trouver une transformation u qui ne déforme pas trop le signal f_2 , alors que la seconde contrainte vise à synchroniser les signaux f_1 et $f_2 \circ u$. Ces deux contraintes sont antagonistes et peuvent conduire à des solutions dans des directions opposées. Il faut donc trouver une procédure qui soit un compromis entre régularité de la transformation u et synchronisation des signaux. De nombreux problèmes de recalage peuvent donc être formulé comme un problème

variationnel : trouver une transformation u qui minimise la fonctionnelle suivante :

$$L_{f_1, f_2}(u) = D(f_1, f_2 \circ u) + \lambda S(u), \quad (2.2.1)$$

où S est une mesure de la régularité de u , D quantifie la qualité de l'alignement des signaux f_1 et $f_2 \circ u$, et λ est un paramètre de régularisation qui exprime l'équilibre entre une transformation suffisamment lisse et une transformation qui conduit à un bon alignement. D peut être une mesure entre quelques points caractéristiques (landmarks) de f_1 et f_2 à mettre en correspondance, une mesure locale de corrélation entre deux signaux, ou une métrique entre f_1 et $f_2 \circ u$. De nombreux choix pour S sont possibles qui correspondent à des conditions spécifiques de régularité pour u . Par exemple, dans certains algorithmes, la solution optimale u est recherchée parmi l'ensemble des difféomorphismes de Ω , c'est à dire l'ensemble des homéomorphismes de Ω dont l'inverse est différentiable. Selon la stratégie de recherche et la méthode utilisée, u peut également appartenir à d'autres espaces fonctionnels. Afin de présenter différents algorithmes de recalage, nous avons choisi de distinguer les méthodes d'alignement à partir de landmarks qui reviennent à un problème d'interpolation entre deux ensembles de points de Ω , des méthodes d'alignement global lorsque l'on ne sait pas à priori quelle partie d'un signal doit correspondre à une partie donnée de l'autre. Les méthodes de recalage de courbes et d'images seront présentées séparément et nous expliquerons dans quel contexte se situe l'approche développée dans cette thèse.

Notons que distinguer les méthodes qui alignent quelques points caractéristiques de deux signaux, de celles qui recherchent un alignement global n'est pas la seule possibilité de classification des méthodes de recalage. Dans Glasbey et Mardia [42], une revue des algorithmes d'alignement est proposée en distinguant les modèles *paramétriques* et *nonparamétriques* pour la transformation u . Dans Brown [10], les méthodes de recalage sont classifiées à partir des critères suivants :

- un *espace de caractéristiques*, qui correspond à des caractéristiques particulières des signaux f_1 et f_2 à mettre en correspondance.
- un *espace de recherche* qui est l'ensemble des transformations susceptibles d'être utilisées.
- une *métrique de similarité* qui mesure l'optimalité de la transformation u pour l'ensemble de caractéristiques choisi.
- une *stratégie de recherche* qui contrôle le calcul de la déformation optimale.

Dans Lester et Arridge [55], les méthodes sont classifiées en trois groupes principaux qui correspondent à trois stratégies hiérarchiques différentes :

- augmentation de la *complexité des données*, à un niveau grossier d'alignement, seules les structures globales des signaux sont prises en compte et des détails plus fins sont progressivement incorporés dans le processus de recalage.
- augmentation de la *complexité de la transformation*, la complexité de la transformation peut être contrôlée en développant u dans une base fonctionnelle

appropriée (e.g. ondelettes) afin d'organiser les déformations locales et globales de manière hiérarchique, ou bien en incorporant progressivement les landmarks à aligner.

- augmentation de la *complexité du modèle*, la stratégie est d'utiliser des transformations dont la complexité augmente avec la progression du processus de recalage.

2.2 Recalage de courbes

Dans cette partie, nous supposerons que $\Omega = [0, 1]$.

2.2.1 Alignement à partir de landmarks

Un ensemble de landmarks est une série ordonnée de points dans $[0, 1]$ qui sont associées à des points caractéristiques des courbes f_1 et f_2 . Extrema, points d'inflexions et singularités sont des exemples possibles de landmarks. Le but des méthodes de recalage à partir de landmarks est de trouver une transformation qui fait correspondre ces points caractéristiques. La première difficulté de cette approche est la détermination des paires de landmarks qui doivent être associés. Cette étape peut être réalisée manuellement (généralement par un expert ou à partir d'informations *a priori*, voir Munoz Maldonado *et. al.* [71] ou bien les nombreux exemples donnés dans Ramsay et Silverman [77]). Il n'existe pas à notre connaissance d'approche automatique pour décider quelles paires de landmarks doivent être associés lorsque l'on souhaite aligner deux signaux 1D. Une technique de mise en correspondance automatique a donc été développée au cours de cette thèse. Celle-ci est présentée dans le chapitre 4. Remarquons toutefois que Gasser et Kneip [39] ont proposé une méthode pour déterminer les points caractéristiques qui sont communs à un ensemble de m courbes. Leur approche consiste à extraire tout d'abord les landmarks de chacune des courbes, et calculer ensuite la densité (avec un estimateur à noyau) de l'ensemble des positions des points caractéristiques de toutes ces fonctions. Etant donné que les landmarks qui sont communs à toutes les courbes doivent se trouver relativement à la même position, les modes de la densité ainsi obtenue reflètent donc les positions des landmarks qui doivent être mis en correspondance. Toutefois, cette approche suppose que l'on dispose d'un nombre suffisamment important de courbes et n'est pas applicable pour $m = 2$.

Afin de présenter le principe de l'alignement à partir de points caractéristiques, nous supposerons dans ce chapitre que l'on dispose de deux ensembles de N landmarks $(\tau_{1,1}, \dots, \tau_{1,N})$ et $(\tau_{2,1}, \dots, \tau_{2,N})$ obtenus à partir de la forme des courbes f_1 et f_2 . Le but de l'alignement à partir de landmarks est alors de trouver une application u , strictement croissante, telle que pour tout $i = 1, \dots, N : \tau_{2,i} = u(\tau_{1,i})$ (*alignement exact*) or $\tau_{2,i} \approx u(\tau_{1,i})$ (*alignement approché*). Généralement, il est aussi imposé que $u(0) = 0$ et $u(1) = 1$. Ainsi, calculer la transformation u revient à un problème de

régression sous contrainte monotone. Il existe donc de nombreuses manière de définir une application qui aligne deux ensembles de landmarks. De nombreuses références peuvent être trouvées dans Younes [95] et le livre de Ramsay et Silverman [77]. Nous proposons de détailler quelques méthodes pouvant être utilisées :

- **Lissage Spline sous contraintes**

Dans de nombreux problèmes de régression nonparamétrique, il peut être imposé que la courbe obtenue après lissage satisfasse certaines contraintes (e.g. monotonie). Une approche générale du problème de lissage sous contrainte a été proposée par Mammen *et. al.* [59] où de nombreuses références sur ce sujet peuvent être trouvées. La plupart des travaux sur des problèmes de régression nonparamétrique sous contraintes, ont été effectués dans le contexte des Splines (remarquons que le lissage Spline sous contraintes monotone est également discuté dans Mammen et Thomas-Agnan [60]). Le lissage Spline correspond à la minimisation d'un problème de moindres carrés pénalisé, tel que celui donné ci-dessous (2.2.2). Les contraintes sont facilement introduites en minimisant sur un espace réduit de fonctions. Etant donné deux ensembles de landmarks $(\tau_{1,1}, \dots, \tau_{1,N})$ et $(\tau_{2,1}, \dots, \tau_{2,N})$, une approche possible pour trouver une transformation u qui aligne ces deux ensembles de landmarks est de chercher le minimum de :

$$\|h\|^2 = \frac{1}{N} \sum_{i=1}^N (\tau_{2,i} - h(\tau_{1,i}))^2 + \lambda \int_0^1 \{h'(t)\}^2 dt, \quad (2.2.2)$$

parmi toutes les applications $h : [0, 1] \mapsto [0, 1]$ strictement monotones et dérivables, où λ est un paramètre de régularisation.

- **Une équation différentielle pour des fonctions monotones**

Ramsay [75] a proposé de simplifier le problème de régression nonparamétrique sous contrainte monotone, en considérant des déformations d'une forme particulière. La notation $D^m h$ correspond à la dérivée d'ordre $m > 0$ de la fonction h , alors que D^{-1} correspond à l'opérateur d'intégration partielle $D^{-1}h(t) = \int_0^t h(x)dx$. La classe des déformations discutée dans Ramsay [75] est l'ensemble des fonctions h pour lesquelles $\ln D h$ est dérivable et $D \ln D h = D^2 h / Dh$ est de carré intégrable (au sens de Lebesgue). Ces conditions assurent que la fonction h est strictement monotone et que sa dérivée première est lisse et bornée presque partout. Ramsay a montré que chaque fonction h de ce type peut s'écrire $h = \beta_1 + \beta_2 D^{-1} \exp[D^{-1}w]$, où w est une fonction de carré intégrable et β_1, β_2 sont des constantes arbitraires.

Pour h dans cette ensemble particulier de fonctions, la déformation u qui aligne deux ensembles de landmarks $(\tau_{1,1}, \dots, \tau_{1,N})$ et $(\tau_{2,1}, \dots, \tau_{2,N})$ est définie comme la

fonction qui minimise la fonctionnelle F suivante :

$$F(h) = \frac{1}{N} \sum_{i=1}^N (\tau_{2,i} - \beta_1 - \beta_2 m(\tau_{1,i}))^2 + \lambda \int_0^1 \{w(t)\}^2 dt, \quad (2.2.3)$$

où λ est un paramètre de lissage, $m(t) = (D^{-1} \exp D^{-1} w)(t)$ et $h(t) = \beta_1 + \beta_2 m(t)$, $t \in [0, 1]$. Le principale avantage apporté par cette forme paramétrique pour la déformation u est le passage d'un problème de régression nonparamétrique sous contrainte, du type (2.2.2), à celui du calcul sans contrainte de la fonction w . Les détails de l'implémentation pratique de cette approche peuvent être trouvés dans Ramsay [75], et Ramsay et Li [76]. Les programmes Matlab pour créer ce type de fonctions sont disponibles à partir du site ftp :

<http://www.psych.mcgill.ca/faculty/ramsay/ramsay.html>.

Remarquons que pour les deux problèmes (2.2.2) et (2.2.3), $\lambda = 0$ correspond au cas de l'alignement exact de landmarks. Un exemple d'alignement de deux fonctions à partir de leur landmarks est donnée dans la Figure 2.1a. Ces deux fonctions sont de forme similaires mais décalées dans le temps. Notons qu'elles sont également d'amplitude différentes mais que la méthode d'alignement que nous avons choisie ne vise pas à corriger cette différence. Chaque fonction a 5 maxima et 5 minima. Nous avons choisi d'associer le i -ème maxima de g_1 avec le i -ème maxima de g_2 , et le i -ème minima de g_1 avec le i -ème minima de g_2 pour $i = 1, \dots, 5$. Le résultat de l'alignement, en mettant en correspondance les extrema de chacune des fonctions et en utilisant la méthode de Ramsay avec $\lambda = 0.001$, est donné dans la Figure 2.1b (nous avons également imposé $u(0) = 0$ et $u(1) = 1$). Les extrema communs des deux courbes sont alors correctement alignés. La déformation obtenue est représentée dans la Figure 2.1c.

Remarque 2.1 En général, nous disposons seulement d'observations discrétisées des fonctions f_1 et f_2 . Dans la suite, la déformation d'un signal (ici $f_2 \circ u$) sera déterminée à partir d'une interpolation linéaire des observations de celui-ci.

En pratique, les courbes f_1 et f_2 sont généralement inconnues et on observe plutôt deux suites d'observations qui sont des versions bruitées de celles-ci. Dans la suite, nous supposerons donc que deux signaux inconnus f_1 et f_2 (avec $f_j : [0, 1] \rightarrow \mathbb{R}$, $j = 1, 2$) sont observés avec du bruit aux mêmes instants $t_i = i/n$ régulièrement espacés sur l'intervalle $[0, 1]$:

$$y_{j,i} = f_j(t_i) + \sigma_j \epsilon_{j,i}, \quad j = 1, 2, \quad i = 1, \dots, n,$$

où $\epsilon_{j,i}$ ($j = 1, 2$) sont des variables de loi normale, indépendantes et identiquement distribuées (i.i.d.), de moyenne nulle et de variance 1, et σ_j sont des niveaux de bruits inconnus. Le problème que nous considérons est celui de l'alignement des fonctions f_1 et f_2 , c'est à dire celui de l'estimation d'une déformation qui aligne les landmarks des courbes f_1 et f_2 . Supposons qu'il soit possible de trouver une transformation u

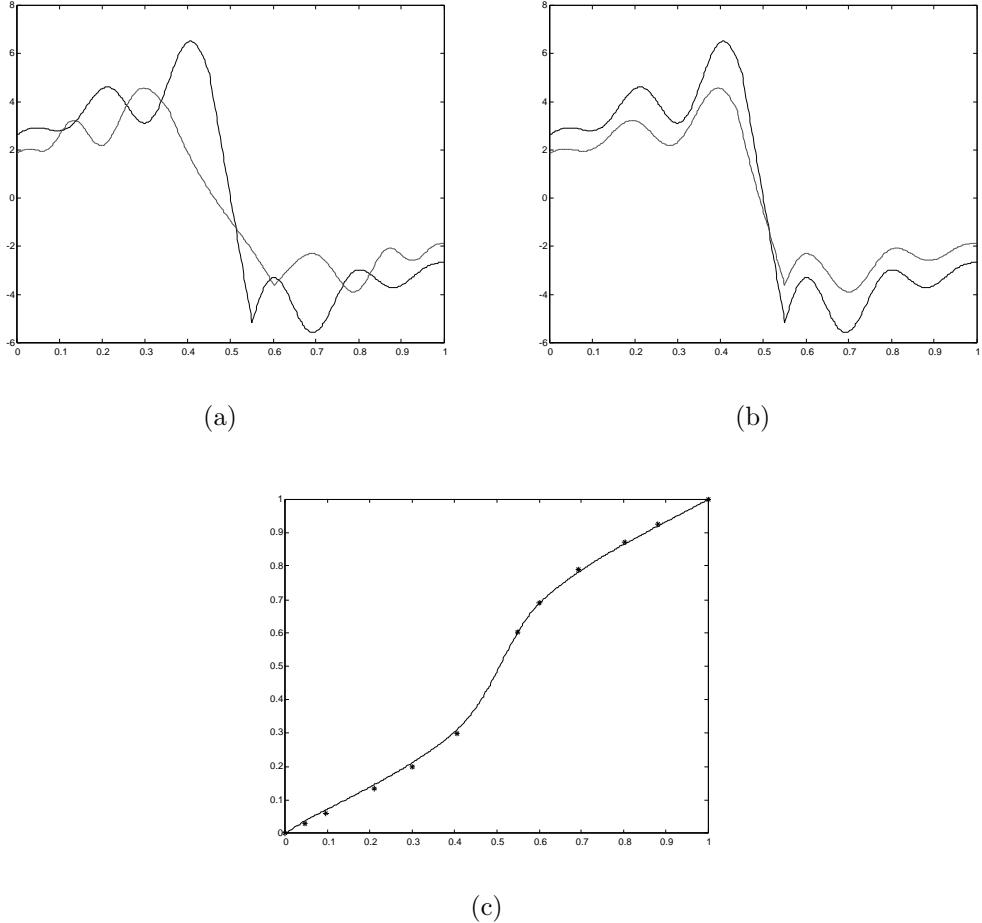


FIG. 2.1: (a) Deux fonctions similaires à recaler : g_1 (trait plein) et g_2 (traits pointillés) (b) Résultat de l'alignement de g_2 sur g_1 à partir de leur extrema avec la méthode de Ramsay ($\lambda = 0.001$) (c) Déformation u obtenue à partir des extrema de g_1 et g_2 . Les paires de landmarks mis en correspondance sont représentées par les points (*).

qui aligne les courbes f_1 et f_2 à partir de leur landmarks, quand celles-ci sont observées sans bruit. Une idée naturelle pour estimer la fonction u est de tout d'abord estimer les landmarks des courbes bruitées f_1 et f_2 , et d'utiliser ensuite l'une des méthodes d'alignement à partir de landmarks décrites précédemment. Ce problème a été étudié d'un point de vue statistique par Kneip et Gasser [51] en utilisant des estimateurs à noyau pour extraire les landmarks d'une fonction observée avec du bruit. Dans cette thèse, nous avons développé une approche temps-échelle avec des ondelettes pour extraire les landmarks d'un signal bruité (voir chapitre 3) qui nous a conduit à proposer une méthode automatique pour décider quelles paires de landmarks doivent être associés lors de la comparaison de deux signaux (voir chapitre 4).

Kneip et Gasser [51] ont proposé un cadre théorique pour étudier les propriétés asymptotiques des estimateurs des déformations nécessaires à la synchronisation d'un ensemble de courbes bruitées. Soit H_N l'ensemble de tous les vecteurs $\tau = (\tau_1, \dots, \tau_N)' \in [0, 1]^N$ avec $\tau_i < \tau_j$ pour tout $i < j$, et $C^1([0, 1])$ l'ensemble de toutes les fonctions continûment dérивables sur $[0, 1]$. En s'inspirant des notations dans [51], nous dirons qu'un opérateur $G : H_N^2 \rightarrow C^1([0, 1])$ est *un opérateur de déformation* si les conditions suivantes sont satisfaites :

- (i) Pour tout $(\tau_1, \tau_2) \in H_N^2$, $G_{(\tau_1, \tau_2)}(.)$ est une fonction réelle strictement croissante. De plus, pour tout $t \in [0, 1]$, $G_{(., .)}(t)$ est continue.
- (ii) Pour tout $\tau_1 = (\tau_{1,1}, \dots, \tau_{1,N})' \in H_N$ et tout $\tau_2 = (\tau_{2,1}, \dots, \tau_{2,N})' \in H_N$, soit :

$$G_{(\tau_1, \tau_2)}(\tau_{1,i}) = \tau_{2,i}, \text{ for all } i = 1, \dots, N,$$

et dans ce cas G est appelé *un opérateur de déformation exacte*, soit :

$$G_{(\tau_1, \tau_2)}(\tau_{1,i}) \approx \tau_{2,i}, \text{ for all } i = 1, \dots, N,$$

et dans ce cas G est appelé *un opérateur de déformation inexacte*.

Soit $\tau_1 = (\tau_{1,1}, \dots, \tau_{1,N})$ et $\tau_2 = (\tau_{2,1}, \dots, \tau_{2,N})$ deux ensembles de N landmarks à mettre en correspondance pour les courbes f_1 et f_2 . Pour un opérateur de déformation G , $u(.) = G_{(\tau_1, \tau_2)}(.)$ est appelée une fonction de déformation pour synchroniser les courbes f_1 et f_2 . En pratique, les landmarks et les fonctions de déformations doivent être estimés. Supposons que l'on dispose d'une méthode pour estimer les landmarks de deux fonctions inconnues (voir chapitre 3) et pour décider quels points caractéristiques doivent être associés (voir chapitre 4), alors, pour un opérateur de déformation donné G , un estimateur \hat{u} de la fonction de déformation u est obtenu par :

$$\hat{u}(.) = G_{(\hat{\tau}_1, \hat{\tau}_2)}(.),$$

où $\hat{\tau}_1 = (\hat{\tau}_{1,1}, \dots, \hat{\tau}_{1,\hat{N}})$ et $(\hat{\tau}_2 = \hat{\tau}_{2,1}, \dots, \hat{\tau}_{2,\hat{N}})$ dénotent deux ensembles de landmarks à mettre en correspondance, estimés à partir des courbes bruitées f_1 et f_2 respectivement.

2.2.2 Alignement global

Soit F une fonction réelle définie sur $]0, \infty[\times \mathbb{R} \times \mathbb{R}$. Une autre approche pour déterminer une fonction de déformation pour aligner deux courbes f_1 et f_2 , est de chercher le minimum, parmi toutes les fonctions croissantes et continûment dérivable $u : [0, 1] \rightarrow [0, 1]$, de la fonctionnelle :

$$L_{f_1, f_2}(u) = \int_0^1 F(u'(x), f_1(x), f_2 \circ u(x)) dx.$$

La fonction F et donc la fonctionnelle L_{f_1, f_2} doivent satisfaire des conditions spécifiques afin de garantir que la solution de cette minimisation corresponde à une solution de notre problème d'alignement. Nous ne les détaillerons pas ici et renvoyons à Trouvé et Younes [85] pour de plus amples références sur l'alignement de structures 1D par des méthodes variationnelles. Le problème de la minimisation de la fonctionnelle $L_{f_1, f_2}(u)$ par programmation dynamique temporelle (dynamic time warping, abrégé DTW par la suite) a été étudiée d'un point de vue statistique par Wang et Gasser ([91],[92]) en utilisant des estimateurs à noyau lorsque les fonctions f_1 et f_2 sont observées avec un bruit Gaussien. Afin de donner une illustration de ces méthodes variationnelles, nous allons expliquer brièvement la méthode proposée dans [91] et [92]. Quelques simulations numériques sont effectuées dans le chapitre 4 pour comparer la technique d'alignement automatique à partir de landmarks développée au cours de cette thèse avec l'algorithme donné dans [92].

L'approche développée par Wang et Gasser ([91],[92]) est la suivante : soit $f_1, f_2 : [0, 1] \rightarrow \mathbb{R}$ deux fonctions qui sont continûment dérivables. Une fonction de déformation pour aligner f_1 et f_2 est définie comme le minimum parmi toutes les fonctions strictement croissantes et continûment dériviales $u : [0, 1] \rightarrow [0, 1]$ de la fonctionnelle :

$$L_{f_1, f_2, \alpha}(u) = \int_0^1 \alpha^2 \left(\frac{f_1(x)}{\|f_1\|_\infty} - \frac{f_2(u(x))}{\|f_2\|_\infty} \right)^2 + (1 - \alpha)^2 \left(\frac{f'_1(x)}{\|f'_1\|_\infty} - \frac{f'_2(u(x))}{\|f'_2\|_\infty} \right)^2 + 2\phi(u(x)) dx,$$

où $\alpha \in [0, 1]$ et ϕ est une fonction qui pénalise les déformations qui ne sont pas réalistes. Les dérivées f'_1 et f'_2 sont introduites dans cette fonctionnelle afin d'aligner les extrema communs des courbes f_1 et f_2 , ce qui est souvent souhaitable dans de nombreux problèmes d'alignement. La normalisation par la norme sup $\|\cdot\|_\infty$ a pour but de réduire les différences d'amplitude entre les courbes. Différents choix pour ϕ sont discutés dans [91],[92]. Dans l'algorithme proposé dans [92], les auteurs proposent de prendre $\phi(u(x)) = 0.001 \frac{(u(x)-x)^2}{1+(u(x)-x)^2}$ afin de pénaliser les fonctions qui sont trop irrégulières. Les variations trop brusques de u sont également pénalisées en imposant que $|u(x) - x| \leq 1/2$. Le paramètre α représente le compromis entre alignement des courbes et alignement de leur dérivées. En pratique, les auteurs suggèrent de résoudre le problème variationnel ci-dessus par programmation dynamique (voir chapitre 4 pour plus de détails sur cette technique) pour trois valeurs de $\alpha = 0.3, 0.5, 0.7$ et de choisir celle qui minimise le coût $L_{f_1, f_2, \alpha}(u)$. Cette technique est également étudiée d'un point de vue statistique lorsque les courbes sont

observées avec un bruit Gaussien et lissées à l'aide d'estimateurs à noyau (voir l'étude numérique conduite dans le chapitre 4 pour évaluer les performances de cette approche). Afin d'illustrer l'intérêt pratique de cette méthode, nous avons appliqué cet algorithme aux courbes g_1 et g_2 que nous avons précédemment alignées à partir de leurs landmarks (voir Figure 2.1). Le résultat de l'alignement de g_1 et g_2 par DTW, ainsi que la fonction de déformation obtenue sont donnés dans la Figure 2.2) . Cette méthode donne un alignement visuellement très satisfaisant des courbes g_1, g_2 (en particulier de leurs extrema).

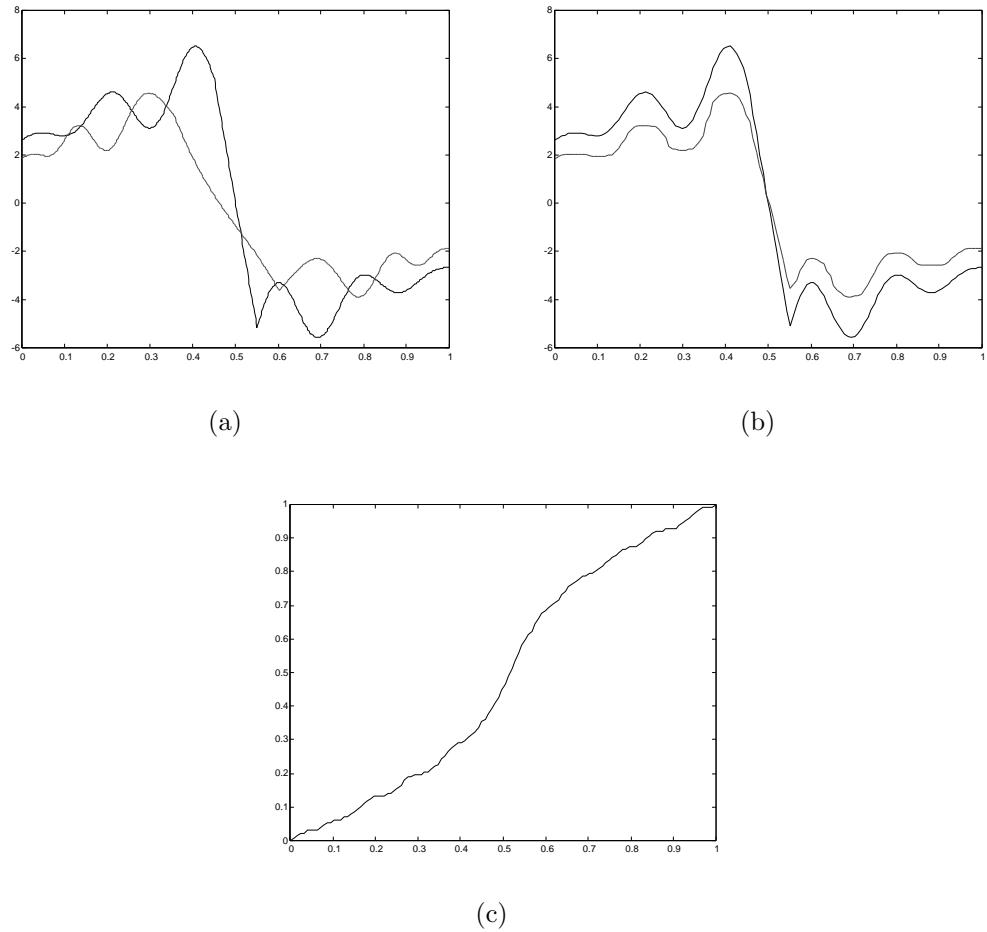


FIG. 2.2: (a) Deux fonctions similaires à recaler : g_1 (trait plein) et g_2 (traits pointillés) (b) Résultat de l'alignement de g_2 sur g_1 par DTW (c) Déformation u obtenue par DTW.

2.3 Recalage d'images

Dans cette partie, nous supposerons que Ω est un ouvert de \mathbb{R}^2 et proposons de présenter quelques techniques pour aligner deux images $I_1, I_2 : \Omega \rightarrow \mathbb{R}^+$. Les

notations et la démarche adoptées s'inspirent de Younes [95].

2.3.1 Alignement à partir de landmarks

Comme dans le cas 1D, un ensemble de landmarks est une série de points dans Ω qui représente des points caractéristiques d'une image (par exemple positions des yeux, des oreilles ou de la bouche dans un visage). L'identification des caractéristiques importantes d'une image peut être faite manuellement (généralement par un expert ou à partir d'information a priori) ou bien de manière automatique (voir chapitre 5). De nombreux exemples de landmarks pour le recalage d'images peuvent être trouvés dans le livre de Dryden et Mardia [35]. Afin d'illustrer le recalage d'images à partir de landmarks, nous avons identifié manuellement 18 landmarks dans chacun des deux visages de la Figure 2.3 et 16 landmarks pour les deux poissons de la Figure 2.4. Dans ces deux exemples, nous avons imposé que les coins des images ne soient pas déplacés lors de la déformation. Les images de la Figure 2.3 correspondent à un problème de reconnaissance de visages. Ces images sont extraites de la base de données de visages ORL. Cette base de données a été utilisée dans le contexte d'un projet de reconnaissance de visages. Pour plusieurs sujets, les images ont été prises à différents instants en modifiant l'éclairage, les expressions du visage (yeux fermés / ouverts, avec sourire / sans sourire) et des détails du visage (lunettes / sans lunettes). Cette base de données peut être obtenue à l'URL :

<http://www.uk.research.att.com/facedatabase.html>.

Nous noterons $P_N(\Omega)$ l'ensemble Ω^N composé de tous les N -uples (x_1, \dots, x_N) avec $x_i \in \Omega$. Le but de l'alignement de landmarks en 2D est le même qu'en 1D : trouver une transformation u qui met en correspondance deux ensembles de points caractéristiques de deux images. Cependant, là encore, la première difficulté de cette méthode est de définir les paires de landmarks qui doivent se correspondre. Généralement, il est supposé que l'on dispose de deux ensembles de N landmarks ordonnés et "marqués", dans le sens où le i -ème landmark dans un ensemble doit être associé au i -ème landmark de l'autre ensemble (voir Dryden et Mardia [35]). Le marquage des landmarks peut être fait manuellement mais cette approche peut être fastidieuse et n'est pas reproductible (en particulier si le nombre d'images à aligner est important). De plus lorsque deux images sont observées avec du bruit, le problème du marquage est compliqué par la présence de landmarks dus aux fluctuations du bruit et le fait que de nombreux points dans un ensemble peuvent ne pas avoir d'homologue dans l'autre. Dans cette partie nous présentons donc un algorithme (TPS-RPM pour Thin Plate Spline - Robust Point Matching) développé par Chui et Rangarajan [22] pour automatiquement aligner deux ensembles de landmarks (qui ne sont pas nécessairement de même taille).

Toutefois, avant d'expliquer cette technique, nous proposons de passer en revue quelques méthodes pour calculer une fonction de déformation qui aligne deux ensemble de N landmarks ordonnés et marqués (v_1, \dots, v_N) et (y_1, \dots, y_N) appartenant à

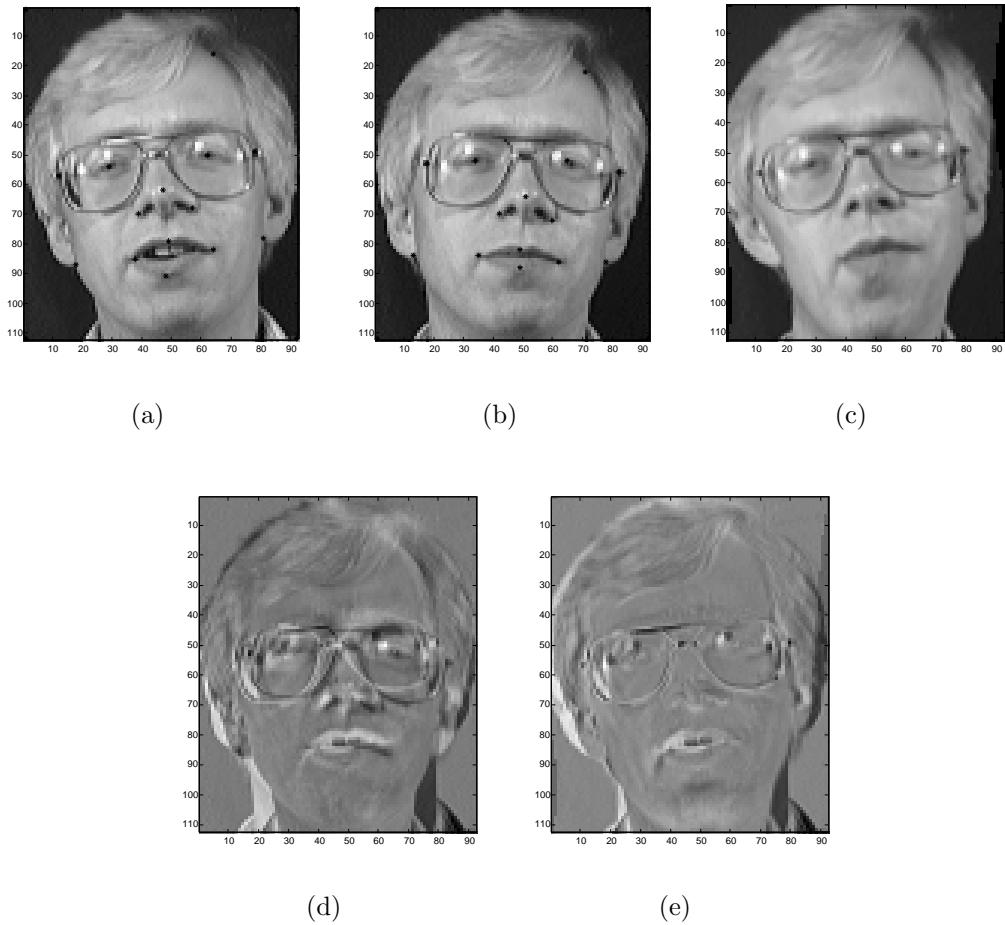


FIG. 2.3: Alignement de visages : (a) Visage 21 (taille : 112×92). (b) Visage 22 (taille : 112×92). Les landmarks utilisés pour aligner ces deux images sont représentés par les points (*). (c) Déformation du visage 22. Différence pixel par pixel entre les deux visages avant (d) et après (e) l'alignement.

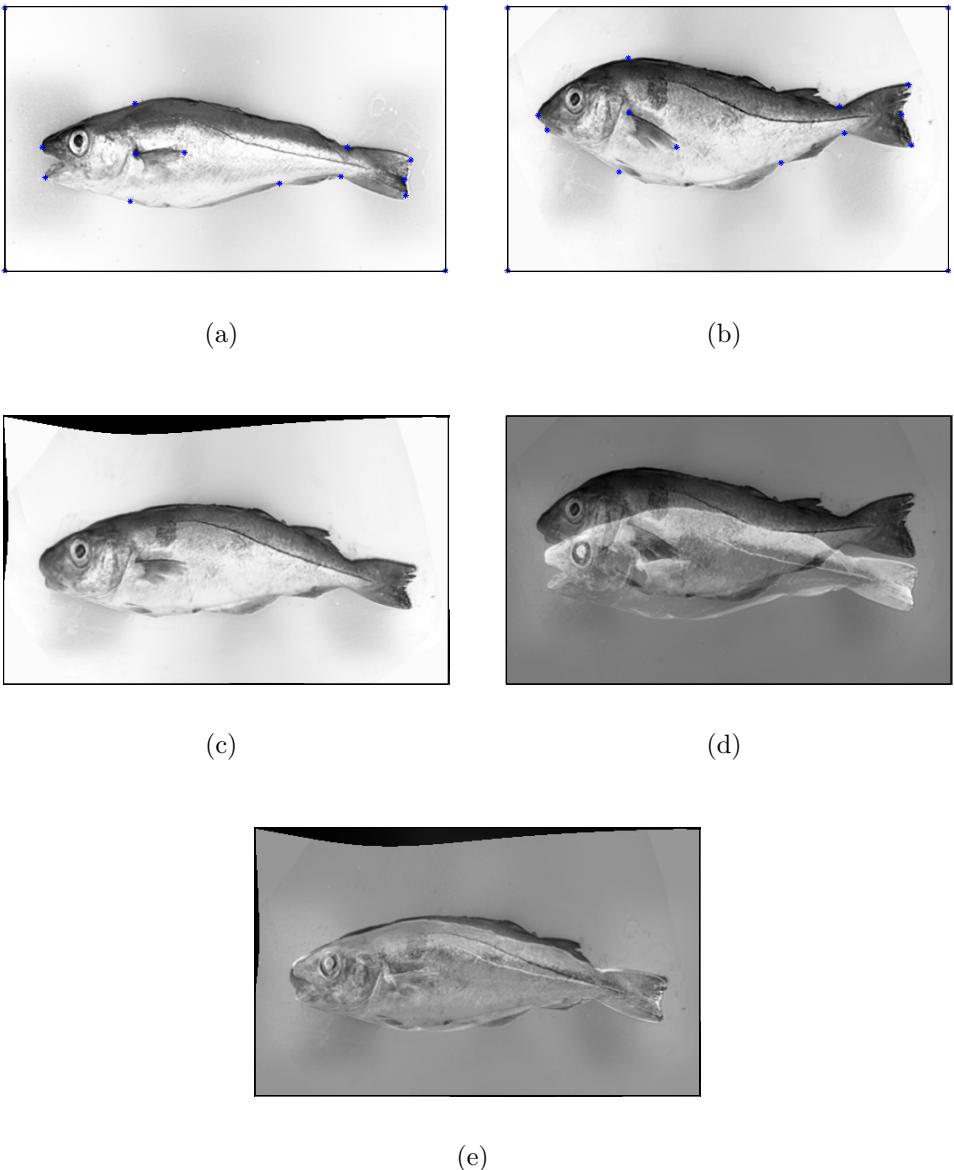


FIG. 2.4: Alignement de poissons : (a) Whiting (taille : 300×500). (b) Haddock (taille : 300×500). Les landmarks utilisés pour aligner ces deux images sont représentés par les points (*). (c) Déformation du haddock. Différence pixel par pixel entre les deux poissons avant (d) et après (e) l'alignement.

$P_N(\Omega)$. Soit $H(\Omega)$ l'ensemble des homéomorphismes de Ω , c'est à dire l'ensemble des applications $u : \Omega \rightarrow \mathbb{R}$ qui sont continues et dont l'inverse est continu. Comme en 1D, le but est de trouver une application $u \in H(\Omega)$ telle que pour tout $i = 1, \dots, N$: $y_i = u(v_i)$ (*alignement exact*) or $y_i \approx u(v_i)$ (*alignement approché*).

2.3.1.1 Transformations paramétriques

Dans beaucoup d'applications, la déformation ne doit pas être une application plus sophistiquée qu'elle n'a besoin d'être et u doit donc appartenir à un espace de déformations particulières. Deux espaces importants qui sont reliés à l'analyse de forme sont l'espace de forme de Kendall (voir [50]), qui est l'ensemble des similitudes euclidiennes (translations, rotations et dilatations), et l'espace de forme affine (i.e. l'ensemble des transformations affines qui est l'espace le plus général des transformations linéaires). Ces deux espaces sont étudiés en détail dans Dryden [34], dans le but d'aligner deux ensembles de landmarks au moyen d'une régression. Ces deux espaces caractérisent des transformations dites rigides dans le sens où les caractéristiques géométriques d'un ensemble de landmarks sont conservées lors de la déformation. Une revue des transformations paramétriques pour associer deux ensembles de landmarks peut également être trouvée dans Glasbey et Mardia [42], où de nombreuses références et d'exemples pratiques sur ce thème sont disponibles.

2.3.1.2 Transformations nonparamétriques

Lorsqu'il n'existe pas suffisamment d'informations sur la déformation sous-jacente, le choix d'une forme paramétrique pour u n'est pas évident et celui-ci donnera généralement de mauvais résultats en présence de déformations locales et non-rigides. Dans ce cas, une forme nonparamétrique est plus appropriée pour la fonction u . Nous présenterons deux méthodes de lissage : la “Thin-Plate Spline” (TPS) introduite par Bookstein [8] et les Splines d'interpolation géodésiques qui est une technique développée par Camion et Younes [17] pour garantir que la solution obtenue u est un difféomorphisme.

- **Thin-Plate Spline**

Avant de présenter l'interpolation TPS, nous rappelons quelques résultats concernant la théorie des Splines. Soit $c_1, \dots, c_N \in \mathbb{R}$ et $(v_1, \dots, v_N) \in \Omega$. Le lissage Spline correspond à la recherche d'une fonction $h : \Omega \rightarrow \mathbb{R}$ qui minimise :

$$\sum_{i=1}^N (h(v_i) - c_i)^2 + \lambda \|h\|_L, \quad (2.2.4)$$

où la régularité de h est évaluée au moyen d'une norme du type :

$$\|h\|_L = \int_{\Omega} |Lh|^2 dv,$$

où L est un opérateur appliqué à h . Le cas de l'alignement exact correspond à $\lambda = 0$. Cette norme définit un espace de Hilbert de fonctions H_L , avec le produit scalaire :

$$\langle h, g \rangle_L = \int_{\Omega} LhLg dv.$$

Si l'application $h \mapsto h(v)$ (pour un $v \in \Omega$ donné) est continue pour la norme $\|h\|_L$, alors le théorème de représentation de Riesz implique que pour tout $v_i \in \Omega, i = 1, \dots, N$, il existe un élément $f_{v_i} \in H_L$ tel que pour tout $h \in H_L$:

$$h(v_i) = \langle f_{v_i}, h \rangle_L.$$

Ainsi, les contraintes $\sum_{i=1}^N (h(v_i) - c_i)^2 = \sum_{i=1}^N (\langle f_{v_i}, h \rangle_L - c_i)^2$ dans (2.2.4) ne sont pas affectées si l'on remplace h par $h + g$ où g est orthogonal à tous les f_{v_i} , ce qui implique que la solution du problème (2.2.4) doit s'exprimer comme une combinaison linéaire des $f_{v_i}, i = 1, \dots, N$. Donc tout dépend de l'existence des fonctions f_{v_i} . Pour expliquer comment l'on peut calculer une expression analytique des f_{v_i} , nous introduisons l'opérateur dual L^* , tel que pour tout g et h à support compact dans Ω ,

$$\int_{\Omega} (Lh)gdv = \int_{\Omega} h(L^*g)dv.$$

Posons $K = L^*L$, nous avons alors que pour tout $x \in \Omega$:

$$h(x) = \int_{\Omega} Lf_x Lh dv = \int_{\Omega} f_x K h dv,$$

ce qui signifie que la fonction $(x, y) \mapsto f_x(y)$ est, par définition, le noyau de Green de l'opérateur K . Finalement, tout dépend si le noyau de Green de K est connu ou non (en supposant qu'il existe).

Un cas classique pour lequel il existe une formule explicite pour le noyau de Green est celui où L est une variante du Laplacien. Soit W_2 l'espace des fonctions : $\mathbb{R}^2 \rightarrow \mathbb{R}$ dont la dérivée seconde est de carré intégrable. Soit Δ l'opérateur Laplacien i.e. pour $f \in W_2$,

$$\Delta f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}.$$

Soit Δ^2 le Laplacien itéré, et posons $\phi(r) = r^2 \log(r)$. La fonction ϕ est telle que, pour n'importe quelle fonction régulière $f \in H$:

$$f(x) = \int_{\mathbb{R}^2} \phi(\|x - y\|^2) \Delta^2 f(y) dy,$$

où $\|\cdot\|$ dénote la norme euclidienne sur \mathbb{R}^2 . Afin de retrouver les considérations précédentes, on doit considérer le produit scalaire :

$$\langle f, g \rangle = \int_{\mathbb{R}^2} \Delta f \Delta g dy. \tag{2.2.5}$$

Nous avons alors $\langle f_x, g \rangle = g(x)$ où $f_x(y) = \phi(\|x - y\|^2)$. L'application directe de la théorie des Splines impliquerait que la solution du problème variationnel (2.2.4) pour $L = \Delta$ est de la forme :

$$h(v) = \sum_{i=1}^N w_i \phi(\|v - v_i\|^2).$$

Cependant, dans notre cas, la situation est légèrement différente car le produit scalaire (2.2.5) est dégénéré : $\langle f, f \rangle = 0$ quand f est linéaire. Ainsi, comme précisé dans Younes [95], il est possible de remplacer c_i par $c_i - a^T v_i - b$, pour tout $a \in \mathbb{R}^2$ et $b \in \mathbb{R}$, ce qui implique finalement que la solution optimale de (2.2.4) pour le Laplacien est de la forme :

$$h(v) = a^T v + b + \sum_{i=1}^N w_i \phi(\|v - v_i\|^2). \quad (2.2.6)$$

Revenons maintenant au problème de la mise en correspondance de deux ensembles de landmarks provenant de deux images. Rappelons que le problème est de trouver une déformation $u : \Omega \rightarrow \Omega$ suffisamment régulière telle que $y_i = u(v_i)$ ou $y_i \approx u(v_i), i = 1, \dots, N$. Pour résoudre ce problème, Bookstein [8] a proposé d'appliquer un lissage Spline à chaque composante de $y_i, i = 1, \dots, N$ quand L est le Laplacien. Alors, dans ce cas, chaque composante de la solution optimale est de la forme (2.2.6).

Une autre formulation du lissage Spline pour le cas 2D est donné dans Wahba [87]. Pour une fonction réelle $h : \Omega \rightarrow \mathbb{R}$, définissons :

$$J(h) = \int_{\Omega} \left(\left(\frac{\partial^2 h}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 h}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 h}{\partial y^2} \right)^2 \right)$$

La Thin-Plate Spline est la solution du problème de lissage suivant : étant donné deux ensembles de landmarks ordonnés et marqués (v_1, \dots, v_N) et (y_1, \dots, y_N) dans $P_N(\Omega)$, déterminer l'application $u = (u_x, u_y)$ qui minimise la fonctionnelle suivante :

$$L(u) = \sum_{i=1}^N \|y_i - u(v_i)\|^2 + \lambda(J(u_x) + J(u_y)), \quad (2.2.7)$$

où λ est un paramètre qui contrôle la régularité de u . Alors, pour un paramètre de régularisation donné λ et $v = (v_x, v_y) \in \Omega$, il existe une solution unique à (2.2.7) (voir Wahba [87]) qui peut être écrite comme :

$$u(v) = d_1 v_x + d_2 v_y + d_3 + \sum_{i=1}^N w_i \phi(\|v - v_i\|^2), \quad (2.2.8)$$

où $d_i, i = 1, \dots, 3$ sont des vecteurs 2×1 représentant le déplacement affine et $w_i, i = 1, \dots, N$ sont des vecteurs 2×1 représentant la transformation non-linéaire. Remarquons que c'est exactement la solution proposée par Bookstein [8] comme expliqué précédemment. En suivant les notations de Chui et Rangarajan [22], nous allons utiliser des coordonnées dites homogènes pour les deux ensembles de landmarks (un point y_i (reps. v_i) est représenté par un vecteur "allongé" $(1, y_{ix}, y_{iy})$ (resp. $(1, v_{ix}, v_{iy})$)), afin d'expliquer comment le problème de minimisation (2.2.8) peut être résolu en utilisant l'algèbre linéaire. Pour un vecteur "allongé" $v \in \Omega$, l'équation (2.2.8) peut être écrite,

$$u(v, d, w) = v.d + \phi(v).w, \quad (2.2.9)$$

où d est une matrice 3×3 représentant la transformation affine et w est une matrice de déformation $N \times 3$ qui représente la déformation non-affine. Si nous remplaçons l'expression de u (2.2.9) dans (2.2.7), la fonctionnelle $L(u)$ devient (voir Wahba [87]) :

$$L(d, w) = \|Y - Vd - \Phi w\|^2 + \lambda \text{trace}(w^T \Phi w), \quad (2.2.10)$$

où Y et V sont des versions concaténées des vecteurs "allongés" y_i et v_i , $i = 1, \dots, N$, et Φ est une matrice $K \times K$ dont les coefficients sont donnés par $\Phi_{i,j} = \phi(\|v_j - v_i\|^2); i, j = 1, \dots, N$. Comme expliqué dans [22], trouver directement la paire d, w qui minimise (2.2.10) est difficile, et Wahba [87] a donc proposé d'utiliser une décomposition QR pour séparer les espaces affine et non-affine :

$$V = [Q_1 Q_2] \begin{pmatrix} R \\ 0 \end{pmatrix},$$

où Q_1 et Q_2 sont des matrices orthogonales $N \times 3$ et $N \times (N - 3)$ respectivement. La matrice R est triangulaire supérieure. Ensuite, l'équation (2.2.10) devient :

$$\begin{aligned} L(\gamma, d) &= \|Q_2^T Y - Q_2^T \Phi Q_2 \gamma\|^2 + \|Q_1^T Y - R d - Q_1^T \Phi Q_2 \gamma\|^2 \\ &\quad + \lambda \gamma^T Q_2^T \Phi Q_2 \gamma, \end{aligned} \quad (2.2.11)$$

où $w = Q_2 \gamma$ et γ sont des matrices $(N - 3) \times 3$. La fonctionnelle (2.2.11) peut alors être tout d'abord minimisée par rapport à γ puis par rapport à d . La solution finale pour w et d est,

$$\begin{aligned} \hat{w} &= Q_2 (Q_2^T \Phi Q_2 + \lambda I_{N-3})^{-1} Q_2^T Y, \\ \hat{d} &= R^{-1} (Q_1^T V - \Phi \hat{w}). \end{aligned}$$

Remarquons que d'autres choix pour la fonction ϕ peuvent être faits à condition qu'elle puisse être associé à un opérateur L . Changer ϕ modifie le terme de pénalisation dans le problème variationnel (2.2.7), mais l'expression explicite de L n'est jamais nécessaire. Cependant, ϕ doit vérifier un certain nombre de conditions suffisantes qui sont spécifiées dans Younes [95]. De nombreux choix pour le noyau ϕ

sont donnés dans Glasbey et Mardia [42].

La Thin-Plate Spline permet d'obtenir une interpolation lisse entre les landmarks de deux images (i.e. $y_i \approx u(v_i)$) mais il n'y a aucune contrainte qui garantisse que la transformation globale soit bijective : des “plissemens” au voisinage desquels la transformation n'est pas inversible, peuvent apparaître comme on peut le constater dans la Figure 2.5 où une grille discrétisée du carré unité $[0, 1]^2$ est déformée par TPS. Dans cet exemple, la déformation n'est pas bijective, puisqu'après transformation certaines lignes de la grille dans la Figure 2.5 s'entrecroisent.

Deux exemples d'alignement d'images à partir de landmarks en utilisant l'interpolation TPS sont donnés dans les Figures 2.3 et 2.4. Afin de juger de la qualité de l'alignement, les différences pixel par pixel entre les deux images ont été représentées avant et après le recalage (la valeur zéro correspond à du gris). Nous pouvons constater que les images sont globalement bien alignées, mais que certaines déformations locales n'ont pas été prises en compte (au niveau des contours extérieurs des poissons par exemple).

Remarque 2.2 *Calcul de l'image déformée $I_2 \circ u$. Dans la suite nous supposerons que les images sont des applications de $[0, 1]^2$ dans \mathbb{R}^+ . Si (k_1, k_2) dénote la position d'un pixel sur une grille du carré $[0, 1]^2$ alors typiquement $u(k_1, k_2)$ ne correspond pas à la position d'un autre pixel sur cette grille. Dans tous les calculs de déformations d'images que nous présenterons, $I_2(u(k_1, k_2))$ sera alors obtenu par interpolation bilinéaire de I_2 . Si $u(k_1, k_2)$ se trouve en dehors du carré $[0, 1]^2$, nous posons $I_2(u(k_1, k_2))$ égal à zéro (ce qui correspond à des pixels noirs dans l'image).*

• Splines d'interpolation géodésique

Pour assurer que la solution d'un problème d'alignement de landmarks en 2D soit bijective, Camion et Younes [17] ont proposé de combiner la théorie des groupes de difféomorphismes, générés comme flots (solution d'une équation différentielle ordinaire, EDO) sur un ensemble Ω , avec des Splines d'interpolation pour garantir que la déformation qui aligne deux ensembles de landmarks soit un difféomorphisme. Leur méthode conduit à un algorithme numérique basé sur une minimisation d'énergie par descente de gradient et sur l'interpolation TPS.

Pour présenter leur approche, nous avons besoin d'expliquer comment le coût d'une déformation h (c'est à dire son énergie) peut être définie comme la somme de coûts infinitésimaux et de petites déformations. Soit d_1, \dots, d_n des petits déplacements sur Ω et posons $\psi_i = id + d_i$. La déformation totale après avoir successivement appliqué les transformations ψ_1, ψ_2, \dots est :

$$g_k = (id + d_k) \circ \dots \circ (id + d_1), \quad (2.2.12)$$

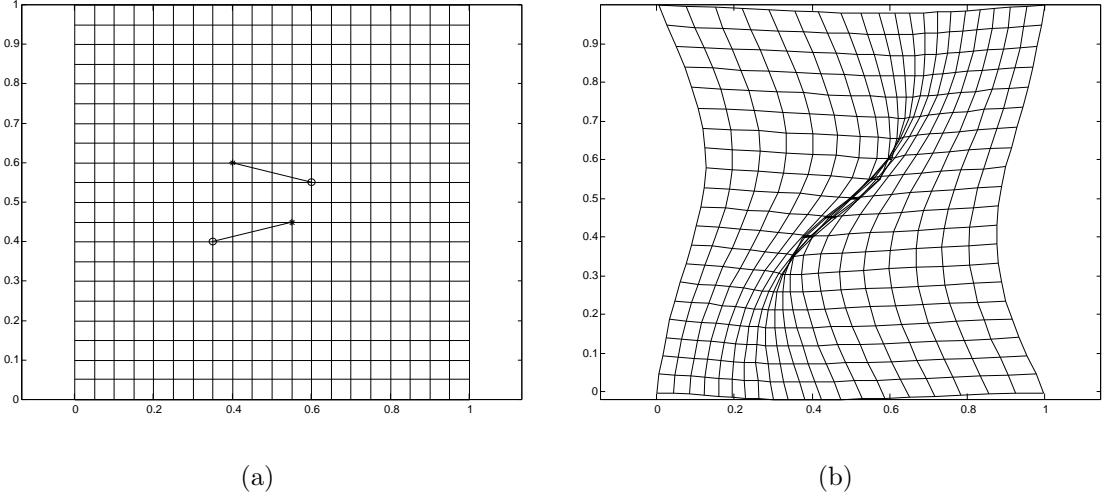


FIG. 2.5: Déformation d'une grille sur le carré unité $[0, 1]^2$: (a) Le point $v_1 = (0.4, 0.6)$ est associé au point $y_1 = (0.6, 0.55)$ et $v_2 = (0.55, 0.45)$ au point $y_2 = (0.35, 0.4)$, alors que les coins restent inchangés (b) Déformation par TPS : certaines lignes de la grille s'entrecroisent après la déformation.

ce qui donne

$$g_k - g_{k-1} = d_k \circ g_{k-1},$$

et nous ajoutons la contrainte $g_n = h$. Supposons qu'un coût $\|d\|$ puisse être associé à chaque petit déplacement et définissons l'énergie du processus de déformation comme :

$$\Gamma(d_1, \dots, d_n) = \|d_1\|^2 + \dots + \|d_n\|^2.$$

Faisons tendre n vers l'infini et remplaçons l'indice k par une variable continue $t \sim k/n$ dans $[0, 1]$. Introduisons $v(t, .)$ et $g(t, .)$ comme les limites de $(d_{[nt]})/n$ et $(g_{[nt]})$ respectivement, alors l'équation (2.2.12) devient :

$$\frac{\partial g}{\partial t}(t, .) = v(t, g(t, .)), \quad (2.2.13)$$

avec les conditions de bords $g(0, .) = id$ et $g(1, .) = h$. Le coût devient :

$$\Gamma(v) = \int_0^1 \|v(t)\|^2 dt.$$

Cette formule donne une nouvelle façon de définir le coût d'une déformation h : soit V_h l'ensemble de tous les déplacements dépendants du temps $t \mapsto v(t, .)$ dans Ω telle que la solution de l'EDO :

$$\frac{\partial y}{\partial t} = v(t, y),$$

avec la condition initiale $y_0 = x$, satisfasse $y(1) = h(x)$. Remarquons que connaître v permet de calculer g par intégration d'une EDO, ce qui donne un flot de difféomorphismes si v vérifie certaines conditions appropriées qui sont étudiées en détail dans Trouvé [84]. Ainsi, l'énergie de h est définie par :

$$E(h) = \inf_{v \in V_h} \int_0^1 \|v(t)\|^2 dt.$$

Revenons maintenant à la méthode de Camion et Younes [17]. Supposons que (v_1, \dots, v_N) et (y_1, \dots, y_N) sont deux ensembles de landmarks ordonnés et marqués dans $P_N(\Omega)$, et soit $q_i(t), i = 1, \dots, N$ des courbes dépendant du temps dans Ω (i.e des trajectoires de landmarks). Pour calculer un difféomorphisme, d'énergie minimale, qui aligne ces deux ensembles, Camion et Younes [17] ont suggéré de minimiser l'énergie suivante :

$$\int_0^1 \int_{\Omega} |Lv(t)|^2 dx dt + \sum_{i=1}^N \int_0^1 \left| \frac{dq_i}{dt}(t) - v(t, q_i(t)) \right|^2 dt,$$

sur l'ensemble de tous les vélocités dépendantes du temps v sur Ω et de toutes les courbes $q_1(.), \dots, q_N(.)$ telles que $(q_1(0), \dots, q_N(0)) = (v_1, \dots, v_N)$ et $(q_1(1), \dots, q_N(1)) = (y_1, \dots, y_N)$. Quand les trajectoires $q_i(.), i = 1, \dots, N$ sont fixées, la solution optimale v peut être calculée explicitement si le noyau de Green de L^*L est connu. Les détails de l'implémentation sont donnés dans Camion et Younes [17], et leur simulations numériques montrent les Splines d'interpolation géodésique sont bijectives, et sont relativement régulières par rapport à l'interpolation TPS qui crée des “plissements”.

2.3.2 Mise en correspondance automatique de deux ensembles de landmarks

Une des difficultés majeures des méthodes d'alignement à partir de landmarks est la définition des paires de points caractéristiques qui doivent se correspondre. De nombreux points dans un des ensembles peuvent ne pas avoir d'homologue dans l'autre ensemble, et le problème de la correspondance peut être compliqué par l'existence de landmarks erronés dus à la présence de bruit. Nous présentons ici une méthode générale d'alignement de points par une transformation non-rigide développée par Chui et Rangarajan [22] (une présentation précise et une revue détaillée des méthodes de recalage de landmarks en 2D sont données dans leur article). En particulier, Chui et Rangarajan [22] comparent différentes approches qui résolvent uniquement le problème de la correspondance, ou déterminent à la fois le calcul de la déformation et des correspondances entre deux ensembles de landmarks. Les performances de l'algorithme détaillé dans [22] (appelé TPS-RPM pour Thin Plate Spline - Robust Point Matching) montrent clairement la supériorité de leur approche sur d'autres méthodes existant dans la littérature (e.g. l'algorithme itéré des points les plus proches (ICP) (Besl et McKay [5]) qui n'est pas robuste

en présence de landmarks erronés ; ou Cross et Hancock [27] qui ont développé un algorithme similaire au TPS-RPM mais qui calcule les correspondances seulement pour des transformations rigides). De plus, l'intérêt du TPS-RPM est démontré sur des exemples réels qui nécessitent des tâches de recalage à partir de landmarks. Nous allons donc maintenant expliquer leur approche de manière un peu plus précise pour le cas 2D.

Supposons que nous ayons deux ensembles de points V et X dans \mathbb{R}^2 que nous noterons $\{v_a, a = 1, \dots, K\}$ et $\{x_i, i = 1, \dots, N\}$ respectivement. Pour résoudre à la fois le problème de la correspondance et celui du calcul de la déformation entre V et X , l'approche décrite dans [22] consiste à minimiser l'énergie suivante :

$$E(M, u) = \sum_{i=1}^N \sum_{a=1}^K m_{ai} \|x_i - u(v_a)\|^2 + \lambda \|Lu\|^2 + T \sum_{i=1}^N \sum_{a=1}^K m_{ai} \log m_{ai} - \zeta \sum_{i=1}^N \sum_{a=1}^K m_{ai}, \quad (2.2.14)$$

où :

- $m_{ai} \in [0, 1]$ peut être interprété comme la probabilité d'associer v_a avec x_i . La matrice de correspondance $M = (m_{ai})$ de taille $K \times N$ permet des “associations partielles” entre les deux ensembles V et X . C'est l'idée de base du *Softassign* qui remplace une correspondance binaire entre deux ensembles de caractéristiques par une correspondance floue. Afin de prendre en compte les landmarks erronés (dus au bruit) et ceux qui n'ont pas d'homologue, une $N+1$ -ème ligne et une $K+1^{th}$ -ème colonne supplémentaires sont ajoutées à la matrice M dont les coefficients doivent satisfaire les contraintes : $\sum_{i=1}^{N+1} m_{ai} = 1$ pour $a = 1, \dots, K$ et $\sum_{a=1}^{K+1} m_{ai} = 1$ pour $i = 1, \dots, N$.
- $T \sum_{i=1}^N \sum_{a=1}^K m_{ai} \log m_{ai}$ est un terme d'entropie qui contrôle le flou de la correspondance, où T est appelé le paramètre de température. Lorsque la température T tend progressivement vers zéro, M devient une matrice binaire (i.e. les coefficients valent 0 ou 1, voir [22]). Ce processus, appelé *Deterministic Annealing*, est une heuristique classique dans de nombreux problèmes d'optimisation (voir [22] pour plus de détails).
- λ est un paramètre de régularisation, L un opérateur et $\|Lu\|^2$ une mesure de régularité.
- $\zeta \sum_{i=1}^N \sum_{a=1}^K m_{ai}$ est un terme de contrôle de robustesse qui évite le rejet de trop de points comme landmarks erronés ou sans homologue. Il peut donc être interprété comme une information a priori sur le pourcentage de ce type de landmarks dans chaque ensemble de points.

L'énergie (2.2.14) peut être minimisée en estimant alternativement la matrice de correspondance M et la transformation u en diminuant progressivement la température T . Etant donné qu'il n'existe pas de manière claire pour estimer le paramètre ζ ,

Chui et Rangarajan [22] ont proposé de considérer des variables erronées qui agissent comme le centre d'une classe avec une large variance T_0 , telle que tous les points qui n'ont pas d'homologue (i.e. quand $\sum_{i=1}^N m_{ai}$ ou $\sum_{a=1}^K m_{ai}$ sont trop petits) sont placés dans cette classe. Par ailleurs le paramètre de lissage λ est fixé à $\lambda = \lambda^{init}T$ où λ^{init} est une constante, de sorte que les transformations globales et rigides sont favorisées à de larges températures, et que des déformations plus locales et non-rigides sont prises en compte quand T tend vers zéro et que la correspondance devient binaire. Pour être un peu plus précis, leur méthode est donc un schéma d'estimation dual, avec un processus qui comprend deux étapes :

- **Etape 1 : Estimation de la correspondance** : Pour les points $\{v_a, a = 1, \dots, K\}$ et $\{x_i, i = 1, \dots, N\}$, m_{ai} est la probabilité que le point x_i appartienne à une classe Gaussienne centrée au point $u(v_a)$ et de variance T ,

$$m_{ai} = \frac{1}{T} e^{-\frac{(x_i - u(v_a))^T (x_i - u(v_a))}{2T}},$$

et les coefficients correspondant aux landmarks sans homologue sont donnés par :

$$\begin{aligned} m_{K+1,i} &= \frac{1}{T_0} e^{-\frac{(x_i - v_{K+1})^T (x_i - v_{K+1})}{2T_0}}; i = 1, \dots, N, \\ m_{a,N+1} &= \frac{1}{T_0} e^{-\frac{(x_{N+1} - u(v_a))^T (x_{N+1} - u(v_a))}{2T_0}}; a = 1, \dots, K, \end{aligned}$$

où v_{K+1} et x_{N+1} sont les centres de gravité des deux ensembles de points V et X respectivement, et sont les centres de deux classes Gaussiennes de variance T_0 où les landmarks sans homologue sont placés comme expliqué précédemment. La normalisation suivante des lignes et des colonnes est répétée plusieurs fois afin de satisfaire les contraintes pour la matrice M :

$$\begin{aligned} m_{ai} &= \frac{m_{ai}}{\sum_{b=1}^{K+1} m_{bi}}; i = 1, \dots, N, \\ m_{ai} &= \frac{m_{ai}}{\sum_{j=1}^{N+1} m_{aj}}; a = 1, \dots, K. \end{aligned}$$

Le nombre d'itérations de cet algorithme de normalisation est fixé à l'avance (10 en pratique), et est indépendant de la température et du nombre de points.

- **Etape 2 : Estimation de la transformation** : Après avoir éliminé le terme indépendant de u dans (2.2.14), un problème simplifié d'interpolation est implémenté afin d'éliminer les landmarks sans homologue :

$$\min_u E(u) = \min_u \sum_{a=1}^K \|y_a - u(v_a)\|^2 + \lambda T \|Lu\|^2,$$

où

$$y_a = \sum_{i=1}^N m_{ai} x_i.$$

Les variables y_a sont les nouvelles positions estimées des points de l'ensemble X qui correspondent au point v_a . Remarquons que quelques calculs supplémentaires sont nécessaires pour prendre en compte les landmarks sans homologue (si $\sum_{i=1}^N m_{ai}$ est trop petit) et les éliminer. La solution dépend de la forme de la fonction u . Dans [22], elle est calculée pour l'interpolation TPS, mais le terme de régularisation dans l'énergie TPS (2.2.10) est légèrement modifié pour incorporer une pénalisation de la transformation affine :

$$E(d, w) = \|Y - Vd - \Phi w\|^2 + \lambda_1 \text{trace}(w^T \Phi w) + \lambda_2 \text{trace}[d - I]^T[d - I], \quad (2.2.15)$$

où I est la matrice identité. Pour les raisons données précédemment, les paramètres λ_1 et λ_2 sont diminués avec la température T : $\lambda_i = \lambda_i^{init} T$, $i = 1, 2$, où λ_2^{init} est une constante qui est beaucoup plus petite que λ_1^{init} afin de donner plus de liberté à la transformation affine. En pratique, λ_1^{init} est fixée à 1 et λ_2^{init} à 0.01.

A chaque température T , le processus de mise à jour alterné de la matrice de correspondance M et du calcul de la transformation u est répété 5 fois. Ensuite, la température est diminuée, et le processus entier est répété selon le schéma suivant :

- **Annealing** : Commencer avec $T_{init} = T_0$, la température T est ensuite progressivement réduite selon $T^{new} = rT^{old}$, ($r \in [0, 1]$ est le taux de décroissance). A chaque nouvelle température, le processus de mise à jour alterné est répété, puis la température est diminuée, et l'ensemble de la procédure est exécuté à nouveau jusqu'à ce qu'une certaine température T_{final} soit atteinte. Dans [22], il est recommandé de poser $r = 0.93$. La température T_0 est égale à la racine carrée de la plus grande distance entre toutes les paires de points de V et X . La température finale T_{final} est égale à la racine carré de la distance moyenne entre les voisins les plus proches à l'intérieur de l'ensemble V (i.e l'ensemble qui est déformé). L'interprétation de ce choix est qu'à T_{final} , les classes Gaussiennes pour tous les points dans V se "chevauchent" à peine les unes avec les autres.

Une implémentation sous Matlab de leur algorithme est disponible à l'URL : <http://www.cise.ufl.edu/anand/students/chui/research.html> et des exemples d'utilisation de cette méthode de recalage seront donnés dans le chapitre 5.

2.3.3 Alignement global

Nous terminons ce chapitre par une courte revue de quelques méthodes d'alignement global d'images. Lorsque l'on ne dispose pas de landmarks ou que ceux-ci n'apportent pas suffisamment d'information, une approche d'alignement global des pixels des images doit être utilisée. Soit I_1, I_2 deux images qui doivent être recalées. Le but des méthodes d'alignement global est de trouver une transformation u suffisamment régulière telle que I_1 et $I_2 \circ u$ soient suffisamment proches l'une de l'autre pour une certaine mesure de dissimilarité entre deux images. La solution u est généralement

définie comme le minimum d'un problème variationnel du type (2.2.1). Nous ne donnerons pas une revue détaillée des très nombreuses approches qui existent dans la littérature mais présenterons quelques grands groupes de méthodes d'alignement en indiquant quelques références.

- **Recalage élastique**

Le recalage élastique produit une fonction de déformation u qui minimise une distance entre I_1 et $I_2 \circ u$ (e.g. $D(I_1, I_2 \circ u) = \int_{\Omega} \|I_1(x) - I_2(u(x))\|^2 dx$), en imposant des contraintes sur l'ensemble possible de transformations (généralement au travers d'une norme du type $S(u) = \int_{\Omega} |Lu|^2 dx$ où L est un opérateur appliqué à u). La déformation peut être développée dans une base fonctionnelle bien choisie qui peut diagonaliser l'opérateur L et qui représente généralement la déformation u par un ensemble de coefficients qui sont ordonnés depuis des niveaux de détails grossiers vers des niveaux plus fins (les coefficients aux basses fréquences sont des descriptions globales de la transformations, alors que les coefficients à des fréquences plus élevées sont liés à des détails plus fins et plus locaux de la déformation). Une exemple de cette approche avec des ondelettes peut être trouvé dans Amit *et. al.* [2] et beaucoup d'autres références sont données dans Lester et Arridge [55]. L'implémentation numérique de ces techniques est explicitée dans Younes [95].

- **Formulation visqueuse**

La méthode utilisée dans Trouvé [84] consiste à minimiser pour v :

$$L(v) = \int_0^1 \int_{\Omega} |Lv|^2 dx dt + \lambda \int_{\Omega} \|I_1(x) - I_2 \circ g_v(1, x)\|^2 dx,$$

où v est une vitesse dépendante du temps sur Ω , et pour tout $x \in \Omega$, $g_v(t, x)$ est la valeur au temps t de la solution de l'EDO : $\frac{\partial y}{\partial t} = v(t, y(t))$ avec la condition initiale $y(0) = x \in \Omega$. Un exemple de recalage d'image avec minimisation de $L(v)$ par descente de gradient est proposée dans Younes [95].

- **Approche Bayésienne**

Dans une formulation Bayésienne, l'image I_1 est vue comme un modèle de référence à aligner avec une seconde image I_2 , et une distribution a priori $p(u)$ est attribuée à u qui dépend de notre connaissance a priori de la transformation sous-jacente. La fonction de déformation u est alors estimée en maximisant la densité a posteriori :

$$p(u|I_1, I_2) = p(I_2|I_1, u)p(u),$$

ce qui est équivalent à minimiser une vraisemblance pénalisée. Ainsi, une approche Bayésienne peut généralement être formulée comme un problème de recalage élastique

où la distribution a priori sur u impose des contraintes de régularité pour la déformation, et la vraisemblance $p(I_2|I_1, u)$ mesure la qualité de l'alignement entre les deux images. Dans Glasbey et Mardia [42], de nombreuses références sont proposées sur ce sujet et une courte description d'un algorithme de Metropolis-Hastings pour calculer u est donnée. Glasbey et Mardia [43] ont également proposé de formuler le recalage d'image comme un problème de vraisemblance pénalisée :

$$P(I_2|I_1, u, \xi, C, \lambda) = L(I_2|I_1, u, \xi) - \lambda S(u, C),$$

où L est la vraisemblance de I_2 qui dépend de la transformation u et des paramètres ξ , S est une mesure positive de distorsion de u , égale à zéro si u appartient à un ensemble "nul" pré-spécifié de transformations C , et λ est un paramètre de régularisation. Le modèle utilisé pour I_2 est un modèle d'image de Fourier-von Mises plutôt qu'un modèle Gaussien afin de prendre en compte à la fois l'intensité et les contours dans l'alignement des images. Le critère de mesure de distorsion S dépend de l'application et reflète donc notre connaissance a priori de la transformation.

D'autres approches statistiques pour le recalage d'image utilisent une mesure statistique qui compare les valeurs de l'intensité des pixels de I_1 et I_2 , telle que la covariance croisée (cross-corrélation) ou sa généralisation l'information mutuelle (voir Glasbey et Mardia [42] et Lester et Arridge [55] pour une revue de ces techniques).

2.3.4 Active Shape Models - Active Appearance Models

Nous avons décrit précédemment des méthodes de recalage qui n'utilisent pas d'information *a priori* sur les ensembles d'images devant être alignées. Ces approches visent à identifier dans les signaux des structures locales (e.g. contours, singularités ou des régions) qu'il s'agit de regrouper uniquement à partir des données observées, sans s'appuyer sur un modèle global indiquant ce que l'on peut espérer trouver dans les images. Nous voudrions préciser qu'il existe des techniques qui diffèrent de façon significative de ces approches que l'on pourrait qualifier de "bas niveau".

Les méthodes basées sur des modèles d'images utilisent un modèle *a priori*, qui représente ce qui peut être observé dans une image, et cherchent à trouver le meilleur alignement entre le modèle et les données dans une nouvelle image. En utilisant ce type de modèles, l'interprétation d'une image peut donc être formulé comme un problème de recalage : étant donné une nouvelle image à interpréter, des structures peuvent être localisées et identifiées en ajustant les paramètres du modèle de telle sorte que ce dernier génère une "image fictive" qui soit aussi proche que possible de l'image réelle. Deux approches statistiques, pour lesquelles un modèle est construit à partir de l'analyse d'un ensemble d'apprentissage, sont particulièrement utilisées :

- les Modèles de Forme Active (Active Shape Models) manipulent un modèle de forme pour décrire la position des landmarks d'une nouvelle image.

- les Modèles d’Apparence Active (Active Appearance Models) manipulent un modèle capable de synthétiser de nouvelles images du type des objets analysés, afin de déterminer les paramètres du modèle qui génèrent une image de synthèse aussi proche que possible de la nouvelle image.

Les approches décrites ci-dessus nécessitent tout d’abord l’aide d’un expert capable d’identifier des landmarks pour toutes les images de l’ensemble d’apprentissage, de telle sorte que chaque landmark représente un point distinct qui soit présent dans chacune des images. Ces méthodes comportent ensuite deux étapes : la construction d’un modèle statistique de forme ou d’apparence à partir de l’ensemble d’apprentissage, puis la construction d’algorithmes permettant d’ajuster les paramètres du modèle pour interpréter une nouvelle image. Une revue détaillée de ces techniques ainsi que de nombreuses références sur leurs applications à des problèmes de recalage peuvent être trouvées dans Cootes et Taylor [?].

L’inconvénient principal de ces méthodes est que le marquage des landmarks d’un ensemble d’images est souvent imprécis et peut se révéler être une étape fastidieuse. De plus il s’agit généralement d’une approche manuelle qui n’est donc pas reproductible et qui est spécifique au type d’images analysées (e.g. reconnaissance de visages).

Chapitre 3

Une approche par ondelettes pour détecter les landmarks d'un signal 1D

L'alignement de courbes à partir de landmarks a été étudié d'un point de vue statistique par Kneip et Gasser [51] en utilisant des estimateurs à noyaux pour estimer les positions des points caractéristiques d'une fonction régulière (c'est à dire au moins une ou deux fois continûment dérivable). Au début du chapitre 4, nous détaillons la méthodologie employée par Kneip et Gasser [51] et expliquons pourquoi cette approche, assez théorique, est difficilement utilisable en pratique.

Dans ce chapitre, nous proposons d'utiliser la transformée en ondelettes continue d'un signal 1D pour déterminer ses landmarks. Les ondelettes ont démontré avec succès leurs bonnes propriétés de localisation des structures d'un signal, car elles permettent une procédure de zoom pour l'étude des propriétés locales d'une fonction en réduisant progressivement un paramètre d'échelle (voir e.g. Mallat [63], Chapitre 6)). Regarder un signal à différents niveaux d'approximation pour caractériser sa structure locale, a été largement utilisé dans la littérature temps-échelle (Lindeberg [58]). Une approche temps-échelle dans le contexte de l'estimation non-paramétrique de courbes à l'aide d'estimateurs à noyau, a également été proposée par Chaudhuri et Marron [20] pour l'analyse statistique des caractéristiques d'un ensemble de données fonctionnelles. Dans le contexte du traitement du signal à base d'ondelettes, il a été montré par Mallat et Hwang [64], qu'il est possible de caractériser la structure locale d'un signal en suivant la propagation le long des échelles des zero-crossings et des maxima de sa transformée en ondelettes continue, ce qui nous a donc conduit à choisir cette approche pour détecter les landmarks d'une fonction.

Dans ce chapitre, une approche nonparamétrique est proposée pour estimer à différentes échelles, les zero-crossings et les maxima d'ondelettes d'un signal 1D observé avec du bruit. Cependant, il n'existe pas d'expression analytique des lignes

de zero-crossings et de maxima d'ondelettes dans le plan temps-échelle, et nous disposons généralement seulement d'une représentation visuelle de celles-ci. Un nouvel outil, appelé l'*intensité structurelle*, est donc introduit pour identifier les limites de ces lignes lorsqu'elle se propagent aux fines échelles. L'intensité structurelle correspond en quelque sorte à la “densité” des zéros et des maxima (en module) d'une représentation en ondelette le long de plusieurs échelles. Il est montré que les modes des intensités structurelles sont localisés au voisinage des landmarks du signal correspondant. De plus, lors de l'estimation des zero-crossings et des maxima d'ondelettes à différents niveaux de résolution, l'intensité structurelle est également un outil efficace pour éliminer les erreurs d'estimation dues aux fluctuations du bruit.

Dans la section suivante, nous définissons tout d'abord les types de landmarks qui seront estimés. Puis, nous rappelons la définition de la transformée en ondelettes continue d'un signal, et expliquons comment la régularité locale d'une fonction peut être caractérisée par la décroissance de ses coefficients d'ondelettes le long des échelles. Une étude de quelques propriétés des lignes de zero-crossings et de maxima d'ondelettes est également proposée en utilisant la notion de *causalité* d'une représentation multi-échelles, et les intensités structurelles correspondantes sont introduites. Nous définissons ensuite le problème de l'estimation des landmarks d'un signal observé selon un modèle de bruit blanc. Puis, deux procédures de seuillage sont envisagées pour estimer les zero-crossings et les maxima d'ondelettes de la transformée en ondelettes continue d'un signal bruité, et les propriétés asymptotiques des estimateurs sont étudiées. Nous montrons ensuite comment les landmarks d'un signal peuvent être détectés à partir des intensités structurelles des zero-crossings et des maxima d'ondelettes estimés. Enfin, nous précisons comment utiliser notre approche en pratique. Les performances de notre méthode sont illustrées à l'aide de simulations et d'exemples réels.

3.1 Un modèle de régression nonparamétrique

Supposons qu'un signal inconnu f (avec $f : [0, 1] \rightarrow \mathbb{R}$) est observé avec du bruit à des instants discrets, régulièrement espacés $t_i = i/n$:

$$y_i = f(t_i) + \sigma\epsilon_i, \quad i = 1, \dots, n,$$

où ϵ_i ($i = 1, \dots, n$) sont des variables de loi normale indépendantes et identiquement distribuées (i.i.d.) de moyenne nulle et de variance 1, et σ un paramètre de bruit inconnu. Le problème que nous considérons dans ce chapitre est la détection des landmarks (i.e. les points caractéristiques) du signal inconnu f . Dans notre approche, deux types de landmarks seront considérés :

- Type I : les points x_0 où la dérivée m -ème de f est singulière i.e. Lipschitz $0 \leq \alpha < 1$ (pour un entier $m \geq 0$).
- Type II : les points y_0 où la fonction f est r fois continûment dérivable au voisinage de y_0 et telle que la dérivée r -ème de f a un zéro au point y_0 .

Les landmarks de type I correspondent à la présence d'une singularité dans l'une des dérivées de la fonction f . Les singularités dans un signal peuvent être caractérisées de nombreuses manières. Par souci de simplicité, nous avons restreint notre étude à des singularités Lipschitziennes ponctuelles. Une fonction $f : \mathbb{R} \rightarrow \mathbb{R}$ est dite ponctuellement Lipchitzienne d'ordre $\alpha \geq 0$ au point x_0 , s'il existe une constante C_{x_0} et un polynôme P de degré $d = \lfloor \alpha \rfloor$ ($\lfloor \alpha \rfloor$ est la partie entière de α) tels que :

$$\forall x \in \mathbb{R}, |f(x) - P(x - x_0)| \leq C_{x_0}|x - x_0|^\alpha. \quad (3.3.1)$$

La régularité Lipschitzienne de f au point x_0 , notée $\alpha(x_0)$, est le supremum des α pour lesquels l'équation (3.3.1) est vérifiée. Une fonction est uniformément Lipchitzienne d'ordre α sur $[a, b]$ si elle vérifie (3.3.1) pour tout $x_0 \in [a, b]$ avec une constante C qui est indépendante de x_0 . Les exposants d'uniformité Lipchitzienne donnent une mesure globale de la régularité d'une fonction. En particulier, si une fonction f est uniformément Lipschitzienne $\alpha > m$ pour un entier $m \geq 0$ au voisinage de x_0 , alors f est nécessairement m fois continûment dérivable dans ce voisinage.

Si $0 \leq \alpha < 1$, alors $P(x - x_0) = f(x_0)$ et la condition Lipchitzienne (3.3.1) devient :

$$\forall x \in \mathbb{R}, |f(x) - f(x_0)| \leq C_{x_0}|x - x_0|^\alpha.$$

Une fonction qui est bornée mais discontinue au point x_0 est ponctuellement Lipchitzienne d'ordre 0 au point x_0 . Si la régularité Lipchitzienne est $\alpha < 1$ (au point x_0), alors f n'est pas dérivable au point x_0 et α caractérise le type de singularité. Plus généralement, nous dirons que la fonction f a une singularité d'ordre α au point x_0 si sa régularité Lipschitzienne au point x_0 est α . Ainsi, les landmarks de type I sont les points de $[0, 1]$ où f a une singularité d'ordre α .

Nous précisons que notre définition d'une singularité d'ordre α est différente de la caractérisation d'une singularité proposée par Raimondo [74]. Dans cet article, Raimondo considère qu'une fonction f a un α -cusp si f vérifie les hypothèses suivantes :

- Il existe un point $x_0 \in [a, b]$, $0 < a < b < 1$ et une constante positive $C_1 > 0$ tels que lorsque h tend vers zéro :

$$|f^{(m)}(x_0 + h) - f^{(m)}(x_0 - h)| \geq C_1|2h|^\delta, \quad (3.3.2)$$

- Il existe une constante positive C_2 telle que pour tout $x, y \in [0, 1]^2$ avec $x_0 \notin [x, y]$:

$$|f^{(m)}(x) - f^{(m)}(y)| \leq C_2|x - y|^\rho, \quad (3.3.3)$$

où $0 \leq \delta < \rho \leq 1$, m est un entier positif et $\alpha = m + \delta$. La fonction a donc un simple saut ou un cusp au point x_0 , et elle est régulière ailleurs (ρ caractérise la régularité de f aux points $x \neq x_0$).

Un cusp au point x_0 est un point où une fonction f est continue mais non dérivable, et telle que :

$$\begin{aligned} f'(x) &\rightarrow +\infty \quad \text{quand } x \rightarrow x_0 \quad \text{d'un côté (gauche ou droit),} \\ f'(x) &\rightarrow -\infty \quad \text{quand } x \rightarrow x_0 \quad \text{de l'autre côté.} \end{aligned}$$

Par exemple, la fonction $g(x) = |x - 0.5|^\alpha$ (avec $0 < \alpha < 1$) a un cusp au point $x_0 = 0.5$. Puisque g est dérivable pour tout $x \in [0, 1], x \neq 0.5$, elle vérifie la condition (3.3.3) avec $m = 0$ et $\rho = 1$. Cependant, pour tout $h \geq 0$, $|g(0.5+h) - g(0.5-h)| = 0$ ce qui implique que la condition (3.3.2) ne peut pas être vérifiée. Ainsi, la définition d'un point de transition brusque d'ordre α , donnée par Raimondo, n'est pas satisfaisante, puisqu'elle ne permet pas de prendre en compte le type de singularité de la fonction g .

La caractérisation d'un α -cusp donnée by Wang [89] est légèrement différente de celle suggérée par Raimondo. Wang considère qu'une fonction f a un α -cusp au point x_0 avec $0 \leq \alpha < 1$, s'il existe une constante K telle que, lorsque h tend vers zéro à gauche ou à droite :

$$|f(x_0 + h) - f(x_0)| \geq K|h|^\alpha. \quad (3.3.4)$$

Ainsi, dans l'une des preuves de l'appendix (voir Wang [89], p.395), on peut lire que si f a un α -cusp au point x_0 , elle ne peut pas être Lipschitzienne d'ordre α au point x_0 ce qui implique des propriétés spécifiques de décroissance des coefficients d'ondelettes de la fonction f au voisinage de x_0 . Cependant, cette affirmation n'est pas juste. La fonction $g(x) = |x - 0.5|^\alpha$ (avec $0 < \alpha < 1$) vérifie l'équation (3.3.4) avec $x_0 = 0.5$ et $K = 1/2$ bien qu'elle soit Lipschitzienne d'ordre α au point 0.5. Ainsi, la condition (3.3.4) n'est également pas satisfaisante pour la caractérisation des singularités d'ordre α .

3.2 Ondelettes et structure d'un signal

3.2.1 La transformée en ondelettes continue

Nous supposerons que nous travaillons avec une ondelette admissible réelle ψ avec r moments nuls ($r \in \mathbb{N}$). Nous supposerons également que l'ondelette ψ est à décroissance rapide et n'a pas plus de r moments nuls ce qui implique (voir Théorème 6.2 de Mallat [63]) qu'il existe θ à décroissance rapide telle que :

$$\psi(u) = (-1)^r \frac{d^r \theta(u)}{dt^r}, \quad \text{et} \quad \int_{-\infty}^{+\infty} \theta(u) du \neq 0. \quad (3.3.5)$$

De plus, nous supposerons que l'ondelette ψ est normalisée i.e. :

$$\int_{-\infty}^{+\infty} (\psi(u))^2 du = 1.$$

Par définition, la transformée en ondelette continue d'une fonction $f \in L^2(\mathbb{R})$ à une échelle donnée $s > 0$ est :

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u)\psi_s(u-x)du,$$

où $\psi_s(u) = \frac{1}{\sqrt{s}}\psi\left(\frac{u}{s}\right)$.

3.2.2 Zero-crossings d'une transformée en ondelettes

Supposons qu'une fonction $f \in L^2(\mathbb{R})$ est r fois continûment dérivable sur un intervalle $[a, b]$. En permutant le produit de convolution et l'opérateur de dérivation, la transformée en ondelettes de f peut être écrite :

$$W_s(f)(x) = f \star \bar{\psi}_s(x) = s^r f \star \frac{d^r \bar{\theta}_s}{dt^r}(x) = s^r \frac{d^r}{dt^r}(f \star \bar{\theta}_s)(x),$$

où $\bar{\psi}_s(x) = \frac{1}{\sqrt{s}}\psi\left(\frac{-x}{s}\right)$ et $\bar{\theta}_s(x) = \frac{1}{\sqrt{s}}\theta\left(\frac{-x}{s}\right)$.

Puisque θ est à décroissance rapide, nous avons :

$$\lim_{s \rightarrow 0} \frac{1}{\sqrt{s}} \bar{\theta}_s(x) = K\delta_0, \text{ avec } K = \int_{-\infty}^{+\infty} \theta(u)du \neq 0,$$

où δ_0 est la fonction de Dirac en 0. Ainsi, pour tout $x \in]a, b[$,

$$\lim_{s \rightarrow 0} \frac{W_s(f)(x)}{s^{r+1/2}} = \lim_{s \rightarrow 0} f^{(r)} \star \frac{1}{\sqrt{s}} \bar{\theta}_s(x) = K f^{(r)}(x). \quad (3.3.6)$$

Le terme *zero-crossings* sera utilisé pour décrire les points (z_0, s_0) dans le plan temps-échelle tels que $z \mapsto |W_{s_0}(f)(z)|$ a exactement un zéro au point $z = z_0$ au voisinage de z_0 . L'équation (3.3.6) prouve que les zero-crossings d'une transformée en ondelettes convergent vers les zéros de la dérivée r -ème de f qui sont localisés dans $[a, b]$ quand l'échelle $s \rightarrow 0$. Cette propriété montre que l'on peut détecter la position des extrema (resp. des points d'inflexion) d'une fonction en suivant la propagation des zero-crossings de sa transformée en ondelettes quand l'échelle s diminue, et quand l'ondelette mère a $r = 1$ (resp. $r = 2$) moment(s) nul(s).

3.2.3 Maxima d'ondelettes

Le terme *maxima d'ondelettes* est utilisé pour décrire les points (m_0, s_0) dans le plan temps-échelle tels que $z \mapsto |W_{s_0}(f)(z)|$ est localement maximum au point $z = m_0$. Ce maxima local doit être un maximum local strict dans un voisinage à droite ou à gauche de m_0 . Les théorèmes suivants (voir Mallat et Hwang [64]) relient la décroissance de l'amplitude de la transformée en ondelettes dans le plan temps-échelle, à la régularité locale d'une fonction.

Théorème 3.1 Soit ρ un réel tel que $0 \leq \rho \leq r$ où r est le nombre de moments nuls de l'ondelette ψ . Soit $f \in L^2(\mathbb{R})$. Si f est uniformément Lipschitzienne d'ordre ρ sur $[a, b]$, alors il existe une constante $A > 0$ telle que pour tout $x \in [a, b]$ et tout $s \geq 0$:

$$|W_s(f)(x)| \leq As^{\rho+1/2}. \quad (3.3.7)$$

Réciproquement, supposons que f est bornée et que $W_s(f)(x)$ vérifie (3.3.7) pour un réel ρ qui n'est pas un entier et tel que $0 < \rho < r$. Alors, f est uniformément Lipschitzienne d'ordre ρ sur $[a - \epsilon, b + \epsilon]$, pour tout $\epsilon > 0$.

Théorème 3.2 Soit α un réel tel que $0 \leq \alpha \leq r$ où r est le nombre de moments nuls de l'ondelette ψ . Soit $f \in L^2(\mathbb{R})$. Si f est ponctuellement Lipschitzienne d'ordre α au point x_0 , alors il existe une constante $A > 0$ telle que pour tous les points $x \in \mathbb{R}$ et toutes les échelles $s \geq 0$:

$$|W_s(f)(x)| \leq As^{\alpha+1/2} \left(1 + \left| \frac{x - x_0}{s} \right|^\alpha \right). \quad (3.3.8)$$

Réciproquement, soit α un réel non-entier tel que $0 < \alpha < r$. S'il existe une constante $A > 0$ et $\alpha' < \alpha$ tels que pour tous les points $x \in \mathbb{R}$ et toutes les échelles $s \geq 0$:

$$|W_s(f)(x)| \leq As^{\alpha+1/2} \left(1 + \left| \frac{x - x_0}{s} \right|^{\alpha'} \right), \quad (3.3.9)$$

alors f est ponctuellement Lipschitzienne d'ordre α au point x_0 .

Les conditions (3.3.8) et (3.3.9) sont légèrement différentes et plus facile à interpréter quand l'ondelette ψ est à support compact égal à $[-C, C]$. Le cône d'influence de x_0 dans le plan temps-échelle est l'ensemble des points (x, s) tels que x_0 est inclus dans le support de $\psi_{s,x}(u) = s^{-1/2}\psi((u - x)/s)$. Puisque le support de l'ondelette $\psi_{s,x}$ est égal à $[x - Cs, x + Cs]$, le cône d'influence de x_0 est défini par l'ensemble de points (x, s) tels que :

$$|x - x_0| \leq Cs,$$

ce qui est illustré dans la Figure 3.1.

Si x est dans le cône d'influence de x_0 , alors $W_s(f)(x)$ dépend de la valeur de f au voisinage de x_0 . Puisque $|x - x_0|/s \leq C$, les conditions (3.3.8) et (3.3.9) peuvent être écrites :

$$|W_s(f)(x)| \leq A's^{\alpha+1/2}, \text{ pour tout } |x - x_0| \leq Cs,$$

et pour tout $|x - x_0| > Cs$,

$$\begin{aligned} |W_s(f)(x)| &\leq A's^{1/2}|x - x_0|^\alpha, \text{ pour la condition (3.3.8),} \\ |W_s(f)(x)| &\leq A's^{\alpha-\alpha'+1/2}|x - x_0|^\alpha, \text{ pour la condition (3.3.9),} \end{aligned}$$

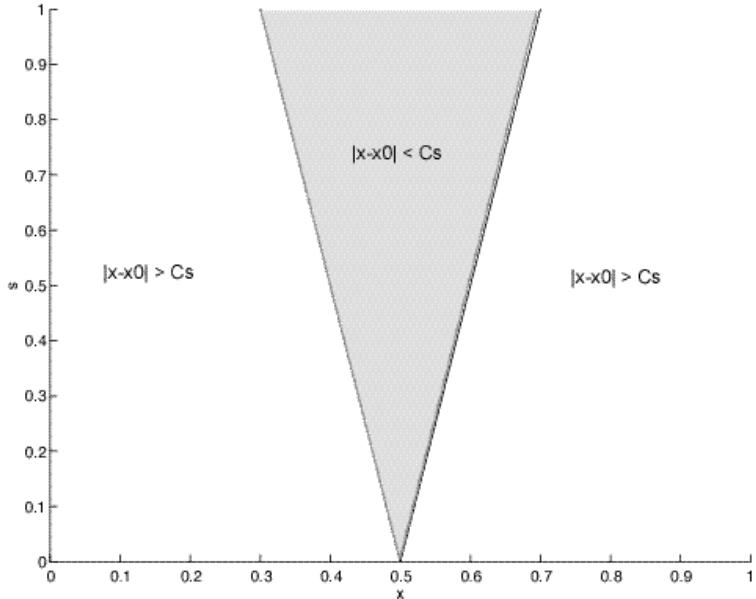


FIG. 3.1: Le cône d'influence de $x_0 = 0.5$ consiste en l'ensemble des points (x, s) pour lesquels le support de $u \mapsto \psi_s(u - x)$ intersecte $u = x_0$.

où $A' > 0$ est une constante qui est indépendante de la position de x et de l'échelle s .

Le théorème suivant (voir Mallat [63] (Chapter 6)) prouve que si $|W_s(f)(x)|$ n'a pas de maxima d'ondelettes aux fines échelles, alors f est localement régulière.

Théorème 3.3 *Supposons que l'ondelette ψ est C^r avec un support compact, et $\psi = (-1)^r \theta^{(r)}$ avec $\int_{-\infty}^{\infty} \theta(t) dt \neq 0$. Soit $f \in L^1[a, b]$. S'il existe $s_0 > 0$ telle que $|W_s(f)(x)|$ n'a pas de maximum local pour $x \in [a, b]$ et $s < s_0$, alors f est uniformément Lipschitzienne d'ordre r sur $[a + \epsilon, b - \epsilon]$, pour tout $\epsilon > 0$.*

Ainsi, une fonction f est singulière en un point x_0 seulement s'il existe une suite de maxima d'ondelettes $(x_p, s_p)_{p \in \mathbb{N}}$ tels que :

$$\lim_{p \rightarrow \infty} x_p = x_0 \text{ et } \lim_{p \rightarrow \infty} s_p = 0.$$

Donc, ce théorème suggère que les singularités d'une fonction peuvent être détectées en suivant la propagation des maxima d'ondelettes aux fines échelles. Cependant, il n'est pas garanti que pour toute ondelette ψ , toute suite de maxima d'ondelettes converge lorsque l'échelle diminue. En particulier, si $W_s(f)(x)$ a un maxima d'ondelettes localisé aux points (x_1, s_1) , alors il se peut que $|W_s(f)(x)|$ n'est pas d'autres maxima d'ondelettes au voisinage de x_1 quand s tend vers zéro. La même remarque est également vraie pour une suite de zéro-crossings. Etant donné que nous voulons détecter les landmarks d'un signal en suivant la propagation des maxima d'ondelettes et des zero-crossings de sa transformée en ondelettes, les propriétés de convergence de ces derniers doivent être étudiées plus attentivement.

3.2.4 Quelques propriétés des lignes de maxima d'ondelettes et de zero-crossings

Nous appellerons *lignes de maxima d'ondelettes* (resp. *lignes de zero-crossings*) toute courbe connectée $m(s)$ (resp. $z(s)$) dans le plan temps-échelle (x, s) le long de laquelle tous les points sont des maxima d'ondelettes (resp. zero-crossings). Nous proposons d'étudier quelques propriétés de ces courbes. Dans ce but, nous allons combiner des résultats de la théorie des ondelettes (Mallat et Hwang [64]) avec des concepts de la littérature temps-échelle (Lindeberg [58]). Les résultats que nous présentons sont similaires à ceux obtenus par Chaudhuri et Marron [20] dans le contexte de l'estimation nonparamétrique de courbes à partir de lissage par des méthodes du noyau.

3.2.4.1 Les bonnes propriétés temps-échelles de la Gaussienne et de ses dérivées

D'après ce qui précède, l'on pourrait penser qu'il est suffisant de suivre n'importe quelle ligne de maxima d'ondelettes ou de zero-crossings pour détecter les landmarks d'une fonction. Cependant, il n'est pas garanti que, pour toute ondelette $\psi = (-1)^r \theta^{(r)}$, $r \geq 0$, un maxima d'ondelette localisé au point (x_1, s_1) appartiennent à une ligne de maxima d'ondelettes qui se propage vers les fines échelles. Une ligne de maxima d'ondelettes (resp. zero-crossings) peut être interrompue quand l'échelle diminue. La proposition suivante (voir Proposition 6.1 de Mallat [63]) prouve que ceci n'est jamais le cas quand θ est une Gaussienne i.e si :

$$\theta(x) = K \frac{1}{\sqrt{2\pi}\beta} e^{-\frac{x^2}{2\beta^2}},$$

pour des réels positifs K et β .

Proposition 3.1 *Soit $\psi = (-1)^r \theta^{(r)}$, $r \geq 0$ où θ est une Gaussienne. Pour tout $f \in L^2(\mathbb{R})$, les maxima en modules de $W_s^r(f)(x) = s^r f^{(r)} * \bar{\theta}_s(x)$ appartiennent à des lignes connectées qui ne sont jamais interrompues quand l'échelle diminue.*

Pour $r \geq 1$, la même proposition est également vraie pour la propagation des zero-crossings, puisque les zéros de $W_s^r(f)(x)$ correspondent à des extrema locaux de $W_s^{r-1}(f)(x)$. Rappelons que notre objectif est l'identification des landmarks d'un signal en suivant la propagation des maxima et des zero-crossings de sa transformée en ondelettes depuis les larges échelles vers des niveaux plus fins. Dans ce but, une condition essentielle est que lorsque l'on se déplace depuis une échelle fine vers une échelle plus grossière, le nombre de zero-crossings (ou de maxima d'ondelettes) doit diminuer, de sorte que de nouvelles caractéristiques ne soient pas créées lorsque le paramètre d'échelle augmente. Cette idée a été formulée par la notion de causalité dans la littérature temps-échelle (voir Lindeberg [58]). Le concept de causalité signifie que la construction d'une représentation temps-échelle doit être telle que la transition depuis une échelle fine vers une échelle plus grossière puisse réellement

être considérée comme une simplification. Lorsque l'on regarde les lignes de zero-crossings et de maxima d'ondelettes dans le plan temps-échelle, la causalité signifie que ces lignes forment des courbes fermées le long des échelles qui ne sont jamais fermées lorsque l'on se déplace depuis une échelle grossière vers une échelle plus fine. La Figure 3.2 donne un exemple où toutes les lignes de zero-crossings et de maxima d'ondelettes ne sont jamais interrompues quand l'échelle diminue et où le principe de causalité est vérifié. L'expression du signal f_1 donné dans la Figure 3.2(a) est :

$$f_1(x) = 2 \sin(6\pi x) + 15|x - 0.4|^{2/5}, \text{ pour } x \in [0, 1]. \quad (3.3.10)$$

Une description du comportement des lignes de zero-crossings et de maxima d'ondelettes de f_1 est donné dans la section 3.2.4.3.

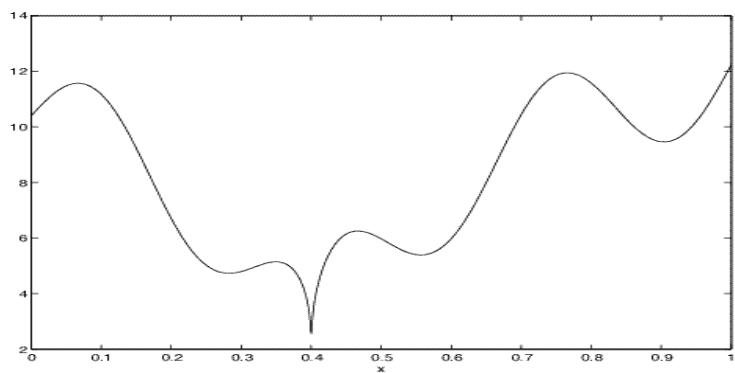
Remarque 3.1 *Dans toutes les figures qui représentent les lignes de zero-crossings ou de maxima d'ondelettes d'un signal, l'axe vertical représente $-\log_2(s)$ où s est le paramètre d'échelle de la transformée en ondelettes.*

Si la causalité est vérifiée pour une représentation temps-échelle, alors les zero-crossings et les maxima d'ondelettes peuvent être vus comme des fonctions du paramètre d'échelle s . De plus, si l'on peut montrer que ces courbes ne sont jamais interrompues quand l'échelle diminue, alors on peut les considérer comme des fonctions continues de s .

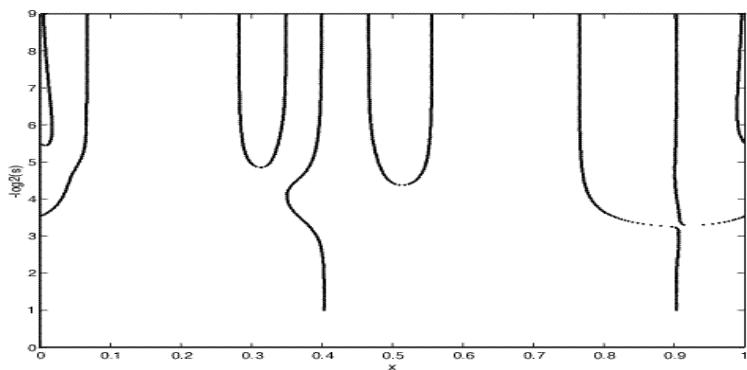
Nous allons maintenant étudier le concept de causalité pour la transformée en ondelettes d'un signal 1D. **Par définition, nous dirons qu'une ondelette est causale**, si pour tout $f \in L^2(\mathbb{R})$ le nombre de zero-crossings $N_z(s)$ et le nombre de maxima d'ondelettes $N_m(s)$ de $x \mapsto W_s(f)(x)$ sont des *fonctions décroissantes et continues à droite du paramètre s* . Les points de sauts (discontinuités) de $N_m(s)$ ou $N_z(s)$ sont appelés *des échelles critiques*. Le théorème suivant donne quelques propriétés des lignes de maxima d'ondelettes $m(s)$ et des lignes de zero-crossings $z(s)$ quand elles sont calculées avec une ondelette qui est causale. En particulier, il est montré que les courbes $z(s)$ et $m(s)$ ne sont jamais interrompues quand l'échelle diminue. Les maxima d'ondelettes (resp. zero-crossings) où $\frac{d^2}{dx^2}W_s(f)(m(s)) = 0$ (resp. $\frac{d}{dx}W_s(f)(z(s)) = 0$) sont appelés des maxima d'ondelettes (resp. zero-crossings) *dégénérés*, alors que les maxima d'ondelettes (resp. zero-crossings) où $\frac{d^2}{dx^2}W_s(f)(m(s)) \neq 0$ (resp. $\frac{d}{dx}W_s(f)(z(s)) \neq 0$) sont appelés des maxima d'ondelettes (resp. zero-crossings) *non-dégénérés*.

Théorème 3.4 *Soit $f \in L^2(\mathbb{R})$ et $\psi = (-1)^r\theta^{(r)}$, $r \geq 0$. Une ondelette causale. Alors,*

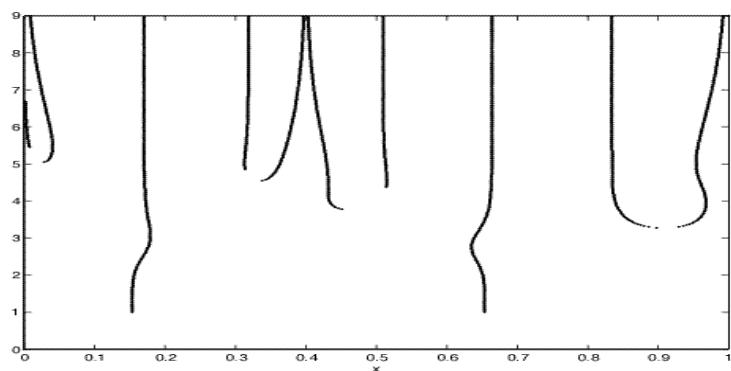
- *Les courbes $m(s)$ (maxima d'ondelettes) et $z(s)$ (zero-crossings) ne sont jamais interrompues quand l'échelle diminue.*
- *Si $(m(s), s)$ (resp. $(z(s), s)$) est un maxima d'ondelettes non-dégénéré (resp. zero-crossings), alors $m(\cdot)$ (resp. $z(\cdot)$) est continûment dérivable au voisinage de s .*



(a)



(b)



(c)

FIG. 3.2: (a) Signal f_1 (voir expression (3.3.10)). (b) Zero-crossings et (c) Maxima d'ondelettes de la transformée en ondelettes de f_1 calculés avec $\psi = -\theta'$ où θ est une Gaussienne.

- Un maxima d’ondelettes (resp. zero-crossing) correspond à une échelle critique seulement s’il s’agit d’un maxima d’ondelettes dégénéré (resp. zero-crossing). De plus, si s_0 est une échelle critique, elle correspond à un point de bifurcation de la trajectoire d’une courbe $m(s)$ ou d’une courbe $z(s)$ dans le plan temps-échelle.

Preuve. Tout d’abord, rappelons que $W_s(f)(x) = f \star \bar{\psi}_s$ où \star dénote le produit de convolution. Puisque ψ est à décroissance rapide et que $f \in L^2(\mathbb{R})$, $s \mapsto W_s(f)(x)$ est continûment dérivable sur $]0, \infty[$ et la propriété de régularisation du produit de convolution implique que $x \mapsto W_s(f)(x)$ est C^1 sur \mathbb{R} . Ainsi, $(x, s) \mapsto W_s(f)(x)$ est C^1 sur $\mathbb{R} \times]0, \infty[$.

Nous allons uniquement prouver le théorème pour les maxima d’ondelettes, car la preuve est identique pour les zero-crossings. Soit $m(s_0)$ un maxima d’ondelettes non-dégénéré i.e. $\frac{d}{dx} W_{s_0}(f)(m(s_0)) = 0$ et $\frac{d^2}{dx^2} W_{s_0}(f)(m(s_0)) \neq 0$. Alors, en utilisant les mêmes arguments que Chaudhuri et Marron [20] dans leur preuve du Théorème 2.2, nous obtenons par la continuité de $W_s(f)(x)$ qu’il existe un $\epsilon > 0$ et un $\delta > 0$ tels que pour tout $x \in [m(s_0) - \delta, m(s_0) + \delta]$ et tout $s \in [s_0 - \epsilon, s_0 + \epsilon]$, $\frac{d^2}{dx^2} W_s(f)(x)$ est non nul et de même signe. Alors, par le théorème des fonctions implicites, nous obtenons que ϵ et δ peuvent être choisis de sorte que pour tout $s \in [s_0 - \epsilon, s_0 + \epsilon]$ il existe un unique $m(s) \in [m(s_0) - \delta, m(s_0) + \delta]$ qui satisfasse $\frac{d}{dx} W_s(f)(m(s)) = 0$ et tel que $m(.)$ soit C^1 sur $[s_0 - \epsilon, s_0 + \epsilon]$. Ainsi, la courbe $m(s)$ n’est pas interrompue au point s_0 quand l’échelle diminue.

Maintenant, supposons que $m(s_0)$ est dégénéré et que la courbe $m(s)$ est interrompue au point s_0 i.e il existe un $\epsilon > 0$ et un $\delta > 0$ tels que pour tout $(x, s) \in]m(s_0) - \delta, m(s_0) + \delta[\times]s_0 - \epsilon, s_0[$ $W_s(f)(x)$ n’a pas de maxima en module. Cependant, puisque l’ondelette ψ est causale, le nombre de maxima locaux dans $]m(s_0) - \delta, m(s_0) + \delta[$ doit augmenter quand l’échelle diminue. Puisque pour $x \in]m(s_0) - \delta, m(s_0) + \delta[$, $x \mapsto W_s(f)(x)$ a un extrema local pour $s = s_0$ et pas d’extrema local pour $s \in]s_0 - \epsilon, s_0[$, ceci contredit la causalité de ψ et implique que la courbe $s \mapsto m(s)$ n’est pas interrompue au point s_0 quand l’échelle augmente, ce qui prouve la première partie du théorème.

Maintenant, supposons que s_0 est une échelle critique, i.e. s_0 est un point de discontinuité de $N_m(s)$. Alors, d’après la causalité de ψ , il existe nécessairement un maxima d’ondelettes au point $(m(s_0), s_0)$, un $\epsilon > 0$ et un $\delta > 0$ tels que pour tout $s \in]s_0 - \epsilon, s_0[$ il existe plus d’un point x dans $]m(s_0) - \delta, m(s_0) + \delta[$ tels que $\frac{d}{dx} W_s(f)(x) = 0$. Donc, $m(s_0)$ est nécessairement dégénéré sinon ceci contredirait l’unicité de la courbe $m(s)$ au voisinage de $(m(s_0), s_0)$ quand $(m(s_0), s_0)$ est non-dégénérée. Puisque les lignes de maxima d’ondelettes ne sont jamais interrompues quand l’échelle diminue, nous avons immédiatement que $(m(s_0), s_0)$ est un point de bifurcation de la courbe $m(.)$. \square

Le Théorème 3.4 prouve que si ψ est causale, alors pour toute courbe $m(s)$ (resp. $z(s)$), il existe une échelle s_m (resp. s_z) telles que $m(.)$ (resp. $z(.)$) peut être considérée comme une fonction continue de $]0, s_m[$ (resp. $]0, s_z[$) dans $[0, 1]$ (voir Figure 3.2). Remarquons que dans la Figure 3.2(b) une bifurcation dans la trajectoire d’une

ligne de zero-crossings a lieu autour du point (0.9, 3.4) et montre l'existence d'une échelle critique.

Si l'ondelette est la dérivée d'ordre r d'une Gaussienne, alors la proposition suivante montre que c'est une ondelette causale :

Proposition 3.2 *Soit $\psi = (-1)^r \theta^{(r)}$, $r \geq 0$ où θ est une Gaussienne. Soit $N_z(s)$ et $N_m(s)$ respectivement le nombre de zero-crossings et de maxima d'ondelettes de $W_s(f)(x)$ pour $s > 0$. Alors, $N_z(s)$ et $N_m(s)$ sont des fonctions décroissantes et continues à droite de s (i.e. l'ondelette ψ est causale).*

Preuve. Soit $\theta_s(x) = \frac{1}{s} \theta\left(\frac{-x}{s}\right)$, où $\theta(x) = K \frac{1}{\sqrt{2\pi\beta}} e^{-\frac{x^2}{2\beta^2}}$. Pour simplifier la preuve, nous supposerons que $K = 1$ et $\beta = 1$. Soit $g(x, s) = \frac{W_s(f)(x)}{s^{r+1/2}} = (f^{(r)} * \theta_s)(x)$ pour $f \in L^2(\mathbb{R})$ (où la dérivée r -ème de f est prise au sens des distributions et $*$ dénote le produit de convolution). Notons que pour $s > 0$, les zero-crossings et les maxima d'ondelettes de $g(x, s)$ et $W_s(f)(x)$ sont aux mêmes positions. Tout d'abord, nous pouvons remarquer que pour tout $s_1, s_2 > 0$:

$$(g(., s_1) * \theta_{s_2})(x) = g(x, \sqrt{s_1^2 + s_2^2}). \quad (3.3.11)$$

L'équation (3.3.11) signifie qu'une représentation à une échelle grossière $g(x, s_2)$ peut être calculée à partir d'une représentation à une échelle plus fine $g(x, s_1)$ par une convolution avec $\theta_{\sqrt{s_2^2 - s_1^2}}$. Maintenant, la positivité totale du noyau Gaussien et la propriété de variation-diminishing d'une transformation continue par convolution de la forme $g(x, s_2) = g(x, s_1) * \theta_{\sqrt{s_2^2 - s_1^2}}$ (voir Lindeberg section 3.5) impliquent que le nombre de zero-crossings de $g(x, s_2)$ est plus petit que le nombre de zero-crossings de $g(x, s_1)$ pour $s_2 \geq s_1 > 0$. Le même résultat est également vrai pour les maxima en module de $g(x, s)$ en considérant les zero-crossings de $f^{(r+1)} * \theta_s(x)$. Ainsi, $N_z(s)$ et $N_m(s)$ sont nécessairement des fonctions décroissantes de s .

Supposons ensuite que $g(x, s)$ a $k \geq 0$ changements de signes pour un $s_0 > 0$ fixé. Dans la preuve du Théorème 3.4, nous avons montré que $(x, s) \mapsto W_s(f)(x)$ est continue sur $\mathbb{R} \times]0, \infty[$ ce qui implique que $g(x, s)$ est une fonction continue de x et s . Donc, il existe $\epsilon > 0$ tel que pour tout $s \in [s_0, s_0 + \epsilon]$, $g(x, s)$ aura au moins k changements de signes. Alors, la décroissance monotone de $N_z(s)$ implique que le nombre de zero-crossings de $g(x, s)$ est exactement égal à k pour tout $s \in [s_0, s_0 + \epsilon]$ ce qui prouve la continuité à droite de $N_z(s)$. Des arguments similaires peuvent être utilisés pour prouver la continuité à droite de $N_m(s)$ ce qui complète la preuve. \square .

Donc, si ψ est la dérivée d'une Gaussienne, il s'agit d'une ondelette causale, et le Théorème 3.4 montre que pour tout $f \in L^2(\mathbb{R})$, les lignes de zero-crossings et de maxima d'ondelettes peuvent être vues comme des fonctions continues du paramètre d'échelle s .

Ainsi, les ondelettes qui sont les dérivées d'une Gaussienne ont de très bonnes propriétés. Cependant, ces ondelettes ne sont pas à support compact ce qui implique

que le Théorème 3.3 ne peut pas être utilisé pour garantir qu'il existe une suite de maxima d'ondelettes qui converge vers n'importe quelle singularité d'une fonction, puisque ce Théorème suppose que l'ondelette utilisée est à support compact. Puisque nous souhaitons détecter les singularités d'ordre α d'un signal en suivant la propagation de ses maxima d'ondelettes, nous devons nous demander s'il est possible de trouver d'autres ondelettes à décroissance rapide dont les lignes $z(s)$ et $m(s)$ associées se propagent aux fines échelles et peuvent être considérées comme des fonctions continues. De nombreux chercheurs ont étudié la propagation de structures dans une représentation temps-échelles (e.g. extrema locaux ou zero-crossing, voir Lindeberg [58], Yuille et Poggio [96]). L'idée principale sous-jacente d'une représentation temps-échelle est que l'information aux fines échelles doit être successivement supprimée quand l'échelle augmente. Cette condition a été formalisée de différentes manières qui ont toujours conduit à l'unicité du noyau Gaussien. Pour illustrer cette propriété d'unicité, nous pouvons donné un résultat prouvé par Lindeberg pour des signaux continus 1D. Lindeberg a considéré une représentation temps-échelles de la forme $L(x, s) = f * g_s(x)$ où $f : \mathbb{R} \rightarrow \mathbb{R}$ est continue et $g : \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}$ est une famille paramétrée de fonctions symétriques. Si l'on requiert que g satisfasse une propriété de semi-groupe : $g_{s_1}(\cdot) * g_{s_2}(\cdot) = g_{s_1+s_2}(\cdot)$, que la norme L^1 de g soit normalisée à un, et que pour $s_2 > s_1$ le nombre d'extrema locaux dans $L(x, s_2)$ ne doit pas excéder le nombre d'extrema locaux dans $L(x, s_1)$, alors nécessairement (et suffisamment) $g_s(x) = \frac{1}{\sqrt{2\pi\beta s}} e^{-\frac{x^2}{2\beta s}}$ pour un $\beta > 0$. Ainsi, il n'est pas certain qu'il soit possible de trouver une ondelette causale qui ne soit pas la dérivée d'une Gaussienne.

Lorsqu'un signal est observé avec du bruit, la détection des zero-crossings ou des maxima d'ondelettes, ainsi que l'étude des propriétés de convergence des estimateurs sont plus faciles avec des ondelettes à support compact du fait de leurs bonnes propriétés de localisation. Cependant, lors de l'utilisation d'ondelettes qui ne sont pas les dérivées d'une Gaussienne, nous ne savons pas comment les lignes $z(s)$ et $m(s)$ se comportent aux fines échelles. Une idée naturelle serait de trouver une ondelette à support compact qui soit proche de la dérivée d'une Gaussienne, de sorte que l'on puisse conjecturer que l'ondelette est causale. Dans ce but, nous proposons d'étudier maintenant quelques propriétés de convergence des ondelettes B-Spline.

3.2.4.2 Convergence asymptotique des ondelettes B-Spline

Définition 3.1 Soit q un entier positif. Nous appellerons ondelette B-Spline d'ordre q avec $r \geq 0$ moments nuls, une fonction ψ_q de la forme :

$$\psi_q(u) = K_{q,r}(-1)^r \theta_q^{(r)}(u), \text{ où } \hat{\theta}_q(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2} \right)^q,$$

où $\hat{\theta}_q(\omega)$ dénote la transformée de Fourier de θ_q et $K_{q,r}$ est une constante de normalisation telle que $\int_{-\infty}^{+\infty} \{\psi_q(u)\}^2 du = 1$

Une ondelette B-Spline ψ_q est donc à une ondelette à support compact dont le support est égal à $[-\frac{q}{2}, \frac{q}{2}]$. Une propriété très intéressante de ces ondelettes est leur

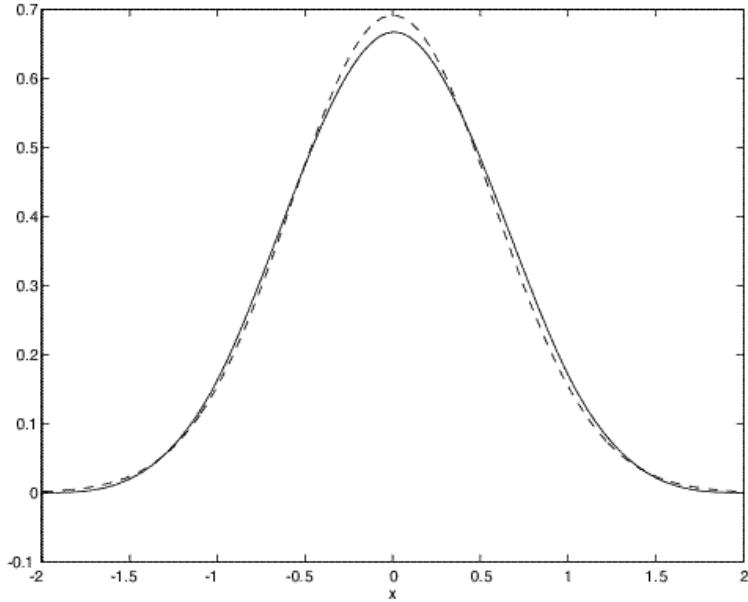


FIG. 3.3: B-Spline d'ordre 4 sans moment nul (trait plein) avec la Gaussienne correspondante (trait pointillé).

convergence vers la dérivée r -ème d'une Gaussienne quand l'ordre q des B-Splines tend vers l'infini. Pour prouver ce résultat, il suffit de montrer la convergence de $\theta_q(u)$ vers une Gaussienne quand q tend vers l'infini, ce qui est donné par le théorème suivant dont la preuve peut être trouvée dans Unser *et. al.* [86] :

Théorème 3.5 *Soit q un entier positif et $\hat{\theta}_q(\omega) = \left(\frac{\sin(\omega/2)}{\omega/2}\right)^q$. Alors, pour tout $u \in \mathbb{R}$:*

$$\sqrt{\frac{q}{3}}\theta_q\left(\sqrt{\frac{q}{3}}u\right) \rightarrow \frac{\sqrt{2}}{\sqrt{\pi}} \exp(-2u^2) \text{ quand } q \rightarrow \infty.$$

et cette convergence est également vraie dans $L^d(\mathbb{R})$ pour $d \geq 2$.

Pour $q = 4$, l'erreur relative d'approximation en norme L^2 entre θ_q et $u \mapsto \frac{\sqrt{2}}{\sqrt{\pi}} \exp(-2u^2)$ est déjà inférieure à 0.2 % (voir Figure 3.3).

De plus, Wang et Lee [90] ont proposé une théorie temps-échelles basée sur des noyaux B-Spline. Ils ont montré que les représentations temps-échelles obtenues à partir de B-Splines héritent des bonnes propriétés du noyau Gaussien. En particulier, ils ont montré que la propriété de causalité (dans le sens discret) est encore vraie pour un filtrage discret avec des B-Splines. Le nombre d'extrema locaux ou de zero-crossings d'un signal discret n'augmente pas lorsque l'on se déplace d'une échelle fine vers une échelle plus grossière, ce qui justifie l'utilisation des noyaux B-Splines discrets en pratique.

Ainsi, pour q suffisamment large, nous conjecturerons qu'une ondelette B-Spline ψ_q est causale, ce qui implique que les lignes de maxima d'ondelettes $m(s)$ et de zero-crossings $z(s)$ sont des fonctions continues de s qui ne sont jamais interrompues quand l'échelle diminue.

3.2.4.3 Intensité structurelle des maxima d'ondelettes et des zero-crossings

Dans cette section, nous supposerons que l'ondelette ψ est causale avec $r \geq 1$ moments nuls. Nous préciserons selon les différentes Propositions, si l'ondelette est supposée à support compact. Nous étudions tout d'abord le comportement des courbes $z(s)$ et $m(s)$ dans le plan temps-échelle lorsqu'une fonction $f \in L^2(\mathbb{R})$ a plusieurs landmarks de type I et de type II (voir section 3.1).

Proposition 3.3 Soit $f \in L^2(\mathbb{R})$ et supposons que ψ soit à support compact égal à $[-C, C]$. Supposons que f a $q \geq 1$ singularités d'ordre α_i localisées aux points x_1, \dots, x_q avec $\max_{1 \leq i \leq q} \{\alpha_i\} < r$, que f est C^{r+1} sur $\mathbb{R} \setminus \{x_1, \dots, x_q\}$ et que f a $p \geq 1$ zéros dans sa dérivée r -ème aux points y_1, \dots, y_p avec $f^{(r+1)}(y_i) \neq 0$ pour $i = 1, \dots, p$. Alors,

- Il existe p lignes de zero-crossings $z_i(s)$ qui convergent respectivement vers y_i pour $i = 1, \dots, p$ quand $s \rightarrow 0$.
- Toutes les lignes de zero-crossings convergent soit vers un point x_i , soit vers un point y_i quand l'échelle s tend vers zéro.
- Il existe q lignes de maxima d'ondelettes $m_i(s)$ qui convergent respectivement vers x_i pour $i = 1, \dots, q$ quand $s \rightarrow 0$.
- Toutes les lignes de maxima d'ondelettes convergent soit vers un point x_i , soit vers un zéro de $f^{(r+1)}$ quand l'échelle s tend vers zéro.

Preuve. Soit $f \in L^2(\mathbb{R})$ qui vérifie les hypothèses ci-dessus. Comme dans la preuve de la Proposition 3.2 définissons $g(x, s) = \frac{W_s(f)(x)}{s^{r+1/2}} = f^{(r)} * \theta_s(x)$ (la dérivée r -ème de f est prise au sens des distributions). Remarquons qu'étudier le comportement des zero-crossings et des maxima en module de $W_s(f)(x)$ ou $g(x, s)$ sont des procédures équivalentes. Nous avons déjà mentionné (voir la preuve du Théorème 3.4) que $(x, s) \mapsto g(x, s)$ est une fonction continue de $\mathbb{R} \times]0, \infty[$. Rappelons que si f est r fois continûment dérivable dans un voisinage de $x \in \mathbb{R}$, alors $g(x, s) \rightarrow Kf^{(r)}(x)$ quand $s \rightarrow 0$ (où $K = \int_{-\infty}^{+\infty} \theta(u) du \neq 0$). Donc, nous pouvons définir $\tilde{g}(x, s) = g(x, s)$ pour $s > 0$, $\tilde{g}(x, 0) = Kf^{(r)}(x)$ et $\frac{d}{dx} \tilde{g}(x, 0) = Kf^{(r+1)}(x)$ pour $x \in \mathbb{R} \setminus \{x_1, \dots, x_q\}$. Observons que $\tilde{g}(x, s)$ est une fonction continue de $\mathbb{R} \times [0, \infty[$ telle que $\tilde{g}(y_i, 0) = 0$ et $\frac{d}{dx} \tilde{g}(y_i, 0) \neq 0$ pour $i = 1, \dots, p$. Le théorème des fonctions implicites implique alors qu'il existe p lignes de zero-crossings $z_i(s), i = 1, \dots, p$ qui sont continues et telles que $z_i(s) \rightarrow y_i$ quand $s \rightarrow 0$.

Considérons maintenant une ligne de zero-crossings $z(s)$ dans le plan temps-échelle. Nous avons que $z(s)$ est une fonction continue qui n'est jamais interrompue quand l'échelle diminue et telle que pour s suffisamment petit $\tilde{g}(z(s), s) = 0$. Soit $x_0 \in \mathbb{R} \setminus \{x_1, \dots, x_q, y_1, \dots, y_p\}$ et V_{x_0} un voisinage de x_0 qui ne contient pas les points

$\{x_1, \dots, x_q, y_1, \dots, y_p\}$. Puisque $\tilde{g}(x, s)$ est continue sur $V_{x_0} \times [0, \infty[$ et telle que $\tilde{g}(x, 0) \neq 0$ pour $x \in V_{x_0}$, nous avons qu'il existe une échelle $s_0 > 0$ telle que pour tout $(x, s) \in V_{x_0} \times [0, s_0[, \tilde{g}(x, s) \neq 0$. Ainsi, pour s suffisamment petit nous avons nécessairement que $z(s) \notin V_{x_0}$ ce qui implique que $z(s)$ converge soit vers un point x_i soit vers un point y_i quand l'échelle s tend vers zéro.

Soit $m(s)$ une ligne de maxima d'ondelettes dans le plan temps-échelle. Puisqu'un extrema local de $f^{(r)} * \theta_s(x)$ est un zéro-crossing de $f^{(r+1)} * \theta_s(x)$, des arguments similaires peuvent être utilisés pour montrer que $m(s)$ converge soit vers un point x_i , soit vers un zéro de $f^{(r+1)}$ quand l'échelle tend vers zéro. Finalement, notons que $f \in L^1(\mathbb{R})$. Ainsi, puisque l'ondelette ψ est à support compact, les Théorèmes 3.3 et 3.4 impliquent qu'il existe q lignes de maxima d'ondelettes $m_i(s), i = 1, \dots, q$ qui convergent respectivement vers x_i quand $s \rightarrow 0$, ce qui termine la preuve. \square

La Proposition 3.3 montre que les landmarks d'un signal peuvent être détectés en suivant la propagation des lignes $z(s)$ et $m(s)$ de sa transformée en ondelettes. Notons que différentes lignes de maxima d'ondelettes et de zero-crossings peuvent converger vers les mêmes landmarks. Par exemple, Mallat et Hwang [64] (Théorème 5) ont montré que si $f \in L^2(\mathbb{R})$ a une singularité d'ordre α au point x_0 ($\alpha < r$) telle que la dérivée r -ème de f a un signe constant sur un voisinage gauche et droit de x_0 mais change de signe au point x_0 , alors il existe une ligne de zero-crossings et deux lignes de maxima d'ondelettes qui convergent vers x_0 . Généralement, le nombre de lignes de maxima d'ondelettes qui convergent vers une singularité dépend du nombre d'extrema locaux de l'ondelette ψ i.e. de son nombre d'oscillations. La Figure 3.2 donne un exemple de propagation de maxima d'ondelettes et de zero-crossings d'un signal qui vérifie les hypothèses de la Proposition 3.3. Il y a deux lignes de maxima d'ondelettes et une ligne de zero-crossings qui convergent vers la singularité au point $x = 0.4$. Puisque la transformée en ondelettes de f_1 est calculée avec une ondelette qui est la dérivée première d'une Gaussienne, les autres lignes de zero-crossings convergent vers les extrema locaux de f_1 où celle-ci est régulière, et les autres lignes de maxima d'ondelettes convergent vers les points d'inflexion de f_1 . Remarquons que les courbes qui convergent vers $x = 0$ et $x = 1$ sont dues à des transitions brutales ($f_1(0)$ et $f_1(1)$ sont très différents) lorsque l'on rend f_1 périodique pour calculer sa transformée en ondelettes.

La Proposition 3.3 montre également que les zero-crossings et les maxima d'ondelettes peuvent converger vers différents types de landmarks. Ainsi, pour décider si un landmark est de type I or de type II, nous avons besoin d'étudier à la fois les lignes $z(s)$ et $m(s)$ qui peuvent converger vers celui-ci. Les Théorèmes 3.1 et 3.2, qui relient la décroissance du module de la transformée en ondelettes à la régularité d'une fonction, sont donc des outils très utiles pour faire cette distinction, car ils permettent de décider si une ligne de maxima d'ondelettes converge vers un point de régularité r ou vers une singularité d'ordre $\alpha < r$, où r est le nombre de moments nuls d'une ondelette ψ . Avant de proposer une méthode pour estimer les zero-crossings et les maxima d'ondelettes d'un signal bruité, nous allons proposer

un nouvel outil pour automatiquement détecter les limites des courbes $z(s)$ et $m(s)$ quand s tend vers zéro.

En effet, jusqu'à maintenant, nous avons montré que les landmarks d'une fonction pouvaient être détectés en suivant les lignes $z(s)$ et $m(s)$ dans le plan temps-échelle. Cependant, nous ne connaissons pas d'expressions analytiques pour les fonctions $z(s)$ et $m(s)$ qui donneraient les landmarks d'un signal sous forme analytique. Nous disposons seulement d'une représentation visuelle (zero-crossings et maxima d'ondelettes dans le plan temps-échelle) qui indique "où" les landmarks sont localisés. Notons que dans le contexte de l'estimation de densité par des méthodes du noyau, Minotte et Scott [67] ont introduit le Mode Tree qui est un outil pour visualiser les caractéristiques d'une densité estimée de façon nonparamétrique, et ont proposé un algorithme empirique pour lier les zero-crossings d'une représentation temps-échelle à des niveaux voisins. Cependant, ils n'ont pas montré que leur algorithme garantie une détection correcte des lignes de zero-crossings. De plus, comme nous le montrerons dans les sections 3.4 et 3.5, l'estimation des zero-crossings ou des maxima d'ondelettes à différentes échelles peut conduire à des lignes qui ne "sont plus continues" (certains zéro ou maxima ne sont pas détectés à certaines échelles) et il n'est alors plus possible de lier ces points à des niveaux de résolution voisins.

Remarquons que si ψ est à support compact et que si $x_0 \in \mathbb{R}$ est un landmark d'une fonction $f \in L^2(\mathbb{R})$, alors toutes les courbes $m_{x_0}(s)$ et $z_{x_0}(s)$ qui peuvent converger vers celui-ci sont inclus dans le cône d'influence de x_0 . Ainsi, si nous pouvons calculer la "densité" des points $m_{x_0}(s)$ et $z_{x_0}(s)$ le long de différentes échelles, nous pouvons espérer que l'intensité ainsi obtenue possède exactement un mode localisé au point x_0 . Cette idée est similaire à la méthode proposée par Gasser et Kneip [39] pour identifier les caractéristiques qui sont communes à un ensemble de courbes. Par exemple, lorsque l'on cherche à identifier les maxima locaux communs à un ensemble de courbes, Gasser et Kneip [39] ont proposé de tout d'abord extraire tous les maxima locaux dans chaque courbe, de les trier dans un seul vecteur et de les soumettre ensuite à une procédure d'estimation de densité par noyau. Les maxima communs donneront alors lieu à des pics dans la densité ainsi estimée. Ici, les fonctions $x \mapsto W_s(f)(x)$ peuvent être vues comme un ensemble de courbes indexées par un paramètre d'échelle s . Ainsi, nous pouvons essayer d'adapter la méthodologie suggérée par Kneip et Gasser [39] pour calculer une "densité" dont les maxima locaux seront localisés au niveau des landmarks de f .

Proposition 3.4 *Soit $f \in L^2(\mathbb{R})$ et $\psi(u) = (-1)^r \frac{d^r \theta(u)}{dt^r}$ une ondelette causale à support compact égal à $[-C; C]$. Supposons qu'il existe q lignes de maxima d'ondelettes $m_i(s)$ qui convergent respectivement vers $x_i \in \mathbb{R}, i = 1, \dots, q$ et p lignes de zero-crossings $z_i(s)$ qui convergent vers $y_i \in \mathbb{R}, i = 1, \dots, p$ quand s tend vers zéro (notons qu'il peut y avoir $x_i = x_j$ ou $y_i = y_j$ pour $i \neq j$). De plus, supposons qu'il existe une constante C' telle que $|x_i - m_i(s)| \leq C's$ et $|y_j - z_j(s)| \leq C's$ avec $0 < C' < C$. Pour $x \in \mathbb{R}$, définissons l'intensité structurelle des maxima d'ondelettes*

$G_m(x)$ et l'intensité structurelle des zero-crossings $G_z(x)$ comme :

$$\begin{aligned} G_m(x) &= \frac{1}{M_m} \sum_{i=1}^q \int_0^{s_{m_i}} \frac{1}{s} \theta\left(\frac{x - m_i(s)}{s}\right) ds \\ G_z(x) &= \frac{1}{M_z} \sum_{i=1}^p \int_0^{s_{z_i}} \frac{1}{s} \theta\left(\frac{x - z_i(s)}{s}\right) ds, \end{aligned}$$

où $[0, s_{m_i}]$ est le support de la ligne de maxima d'ondelettes $m_i(\cdot)$, $[0, s_{z_i}]$ est le support de la ligne de zero-crossings $z_i(\cdot)$, M_m et M_z sont des constantes de normalisation choisies de sorte que G_m et G_z soient des densités de probabilités. Soit $\{x_1^*, \dots, x_{q^*}^*\}$ l'ensemble des points distincts $x_i, i = 1, \dots, q$ et $\{y_1^*, \dots, y_{p^*}^*\}$ l'ensemble des points distincts $y_i, i = 1, \dots, p$. Alors, G_m et G_z sont intégrables sur \mathbb{R} , continûment différentiables sur $\mathbb{R} \setminus \{x_1^*, \dots, x_{q^*}^*\}$ et $\mathbb{R} \setminus \{y_1^*, \dots, y_{p^*}^*\}$ respectivement, et telles que :

- $G_m(x) \rightarrow +\infty$ quand $x \rightarrow x_i^*, i = 1, \dots, q^*$.
- $G_z(x) \rightarrow +\infty$ quand $x \rightarrow y_i^*, i = 1, \dots, p^*$.

Preuve. Soit $f \in L^2(\mathbb{R})$ satisfaisant les hypothèses ci-dessus. Définissons :

$$g_{m_i}(x) = \int_0^{s_{m_i}} \frac{1}{s} \theta\left(\frac{x - m_i(s)}{s}\right) ds \text{ pour } x \in \mathbb{R} \text{ et } 1 \leq i \leq q.$$

Soit $x \in \mathbb{R} \setminus \{x_i\}$ et définissons $s_x = \frac{|x_i - x|}{2C}$. Etant donné que $|x_i - m_i(s)| \leq Cs$, nous avons que $\frac{|x - m_i(s)|}{s} \geq C$ pour tout $s \leq s_x$ ce qui implique que :

$$\begin{aligned} g_{m_i}(x) &= \int_{s_x}^{s_{m_i}} \frac{1}{s} \theta\left(\frac{x - m_i(s)}{s}\right) ds \\ g_{m_i}(x) &\leq \|\theta\|_\infty \int_{s_x}^{s_{m_i}} \frac{1}{s} ds \\ g_{m_i}(x) &\leq \|\theta\|_\infty \left(\log(s_{m_i}) - \log\left(\frac{|x_i - x|}{2C}\right) \right). \end{aligned}$$

Maintenant, définissons $x_{min} = \min_{s \in [0, s_{m_i}]} \{m_i(s) - Cs\}$ et $x_{max} = \max_{s \in [0, s_{m_i}]} \{m_i(s) + Cs\}$ et remarquons que pour $x \notin [x_{min}, x_{max}]$, $g_{m_i}(x) = 0$. Puisque $x \mapsto \log(|x - x_i|)$ est intégrable sur $[x_{min}, x_{max}]$, les inégalités ci-dessus montrent que g_{m_i} est intégrable sur \mathbb{R} .

Soient $a < x_i$ et $x \in]-\infty, a[$. Remarquons que $0 < s_a < s_x$ et que pour tout $s_a \leq s \leq s_x$, $\theta\left(\frac{x - m_i(s)}{s}\right) = 0$. Ainsi,

$$g_{m_i}(x) = \int_{s_a}^{s_{m_i}} \frac{1}{s} \theta\left(\frac{x - m_i(s)}{s}\right) ds.$$

Notons que pour tout $s \in [s_a, s_{m_i}]$, $x \mapsto \frac{1}{s} \theta\left(\frac{x - m_i(s)}{s}\right)$ est continûment dérivable sur $] -\infty, a[$. Puisque θ' est bornée sur \mathbb{R} et étant donné que $s \mapsto \frac{1}{s^2}$ est intégrable

sur $[s_a, s_{m_i}]$, le théorème de dérivation sous le signe somme implique que g_{m_i} est continûment dérivable sur $]-\infty, a[$ et que :

$$g'_{m_i}(x) = \int_{s_a}^{s_{m_i}} \frac{1}{s^2} \theta' \left(\frac{x - m_i(s)}{s} \right) ds.$$

Nous pouvons montrer de manière similaire que g_{m_i} est continûment dérivable sur $]a, +\infty[$ pour $a > x_i$. Maintenant, puisque $\frac{|x_i - m_i(s)|}{s} \leq C'$ avec $0 < C' < C$ pour tout $s \in]0, s_{m_i}]$, il existe une constante $M > 0$ telle que pour tout $s \in]0, s_{m_i}]$, $\theta \left(\frac{|x_i - m_i(s)|}{s} \right) \geq M$ ce qui implique que $s \mapsto \frac{1}{s} \theta \left(\frac{|x_i - m_i(s)|}{s} \right)$ n'est pas intégrable sur $[0, s_{m_i}]$. Puisque $s_x \rightarrow 0$ quand $x \rightarrow x_i$, nous obtenons finalement que $g_{m_i}(x) \rightarrow +\infty$ quand $x \rightarrow x_i$.

Puisque $G_m(x) = \sum_{i=1}^q g_{m_i}(x)$, le résultat suit immédiatement et la preuve pour $G_z(x)$ est identique. \square

La Proposition 3.4 montre qu'en pratique, les landmarks d'une fonction peuvent être obtenus à partir des positions des maxima locaux de $G_m(x)$ et $G_z(x)$ pour $x \in \mathbb{R}$. Les Figures 3.4-3.5 illustrent la détection des landmarks du signal f_1 (voir expression (3.3.10)) via les intensités structurelles de ces lignes de zero-crossings et de maxima d'ondelettes pour une ondelette B-Spline d'ordre 4 avec $r = 2$ moments nuls. Sur ces figures, il peut être constaté que la hauteur d'un maxima local de $G_m(x)$ (resp. $G_z(x)$) est proportionnel au nombre et à la longueur des supports des lignes de maxima d'ondelettes (resp. zero-crossings lines) qui convergent vers celui-ci. Lorsque l'on considère l'intensité structurelle des maxima d'ondelettes, il peut être intéressant de distinguer les pics qui correspondent à des singularités d'un signal f , de ceux qui sont dus à la présence d'un zéro dans la dérivée $(r+1)$ -ème de f . Notons que dans la Proposition 3.8, nous montrerons que si $m(s)$ est une ligne de maxima d'ondelettes qui converge vers une singularité d'ordre α , alors il existe une constante $A > 0$ telle que $\frac{W_s(f)(m(s))}{s^{\alpha+1/2}} \rightarrow A$ quand s tend vers zéro, et remarquons que si $m(s)$ est une ligne de maxima d'ondelettes qui converge vers un point y_i où f est localement C^r , alors il existe une constante $A > 0$ telle que $\frac{W_s(f)(m(s))}{s^{r+1/2}} \rightarrow A$. Maintenant, définissons l'intensité structurelle suivante qui est une version pondérée de celle suggérée dans la Proposition 3.4 :

$$G_m^*(x) = \frac{1}{M} \sum_{i=1}^q \int_0^{s_{m_i}} \frac{h_i(s)}{s} \theta \left(\frac{x - m_i(s)}{s} \right) ds,$$

où $h_i(s) = \frac{W_s(f)(m_i(s))}{s^{r+1/2}}$. D'après les remarques précédentes, il suit que si $m_i(s)$ converge vers une singularité de f alors, aux fines échelles, $h_i(s)$ se comporte comme $s^{\alpha-r}$ et si $m_i(s)$ converge vers un zéro de $f^{(r+1)}$ alors $h_i(s)$ est bornée. Dans la Figure 3.6b, les maxima d'ondelettes du signal :

$$f_3(x) = 2 \sin(6\pi x) + 15|x - 0.4|^{2/5} - 15|x - 0.7|^{3/5} + 15|x - 0.2|^{4/5}, \text{ pour } x \in [0, 1], \quad (3.3.12)$$

sont calculés pour une ondelette B-Spline d'ordre 4 avec $r = 2$ moments. Le signal tracé dans la Figure 3.6a a 3 singularités aux points $x = 0.2$, $x = 0.4$ et $x = 0.7$ d'ordre $4/5$, $2/5$ et $3/5$ respectivement. Alors, pour l'intensité structurelle pondérée $G_m^*(x)$, les pics correspondant aux singularités du signal dominent les autres comme ceci peut être constaté sur la Figure 3.6d. Comme attendu, les hauteurs des pics sont inversement proportionnelles à l'ordre des singularités. Les larges pics aux points $x = 0$ et $x = 1$ sont dus à la discontinuité du signal aux bornes de $[0, 1]$ lorsque celui-ci est rendu périodique pour calculer sa transformée en ondelettes. En pratique, lorsque f est inconnue, nous estimons les fonctions $h_i(s)$ en remplaçant $W_s(f)(m_i(s))$ par sa version bruitée (voir section 4.1.3.3 pour plus de détails).

Remarque 3.2 *Lorsque l'on regarde les intensités structurelles données dans les Figures 3.4-3.6, on pourrait se demander si les cusps qui correspondent aux limites des lignes de zero-crossings ou de maxima d'ondelettes sont les seuls maxima locaux de $G_z(x)$ et $G_m(x)$. Le nombre de maxima locaux de ces intensités structurelles dépend de la forme des lignes $m(s)$ et $z(s)$ dans le plan temps-échelle. Pour certains signaux, de petits maxima locaux de "faibles intensités" des fonctions $G_z(x)$ ou $G_m(x)$ peuvent ne pas correspondre à la présence d'un landmark. Dans la Figure 3.7b, les zero-crossings du signal f_3 (voir expression (3.3.12)) sont tracés pour une ondelette B-Spline d'ordre 4 avec $r = 2$ moments nuls. Dans la Figure 3.7c, il peut être alors constaté que l'intensité structurelle correspondante a un petit maxima, plat, autour de $x = 0.18$ qui ne correspond pas à la limite de l'une des lignes de zero-crossings de la Figure 3.7b.*

Remarque 3.3 *Dans ce chapitre, nous donnons uniquement des exemples de calcul des intensités structurelles à partir d'ondelettes B-Splines. Toutefois, il est également possible de calculer les lignes de maxima d'ondelettes et de zero-crossings à partir d'une ondelette Gaussienne, et d'utiliser un noyau Gaussien pour calculer les intensités structurelles correspondantes. Dans le chapitre 4, une étude numérique permet de comparer la qualité de l'estimation des landmarks d'un signal à partir d'ondelettes B-Splines ou Gaussiennes.*

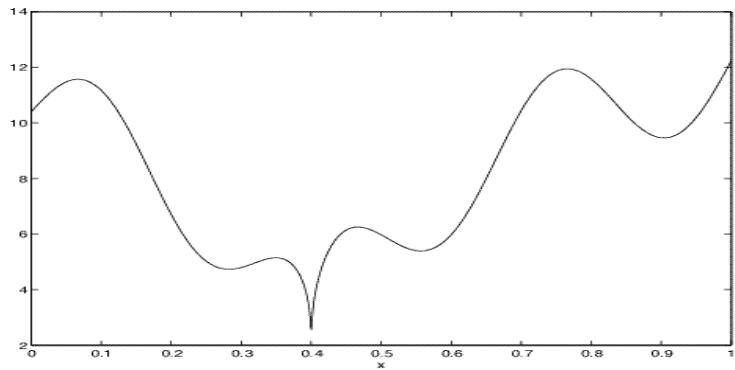
3.3 Modèle de régression nonparamétrique et modèle de bruit blanc

Supposons que f soit observée selon un modèle de bruit blanc :

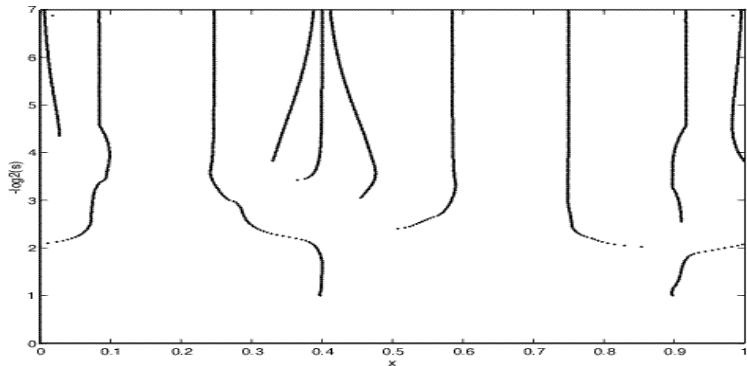
$$Y(dx) = f(x)dx + \tau B(dx), \quad x \in [0, 1], \quad (3.3.13)$$

où τ est un paramètre de bruit, f une fonction inconnue qui peut avoir des singularités, et B un mouvement Brownien standard. Le modèle de bruit blanc (3.3.13) peut être relié au modèle de régression nonparamétrique suivant (voir Brown et Low [11], Donoho et Johnstone ([32], [30])) :

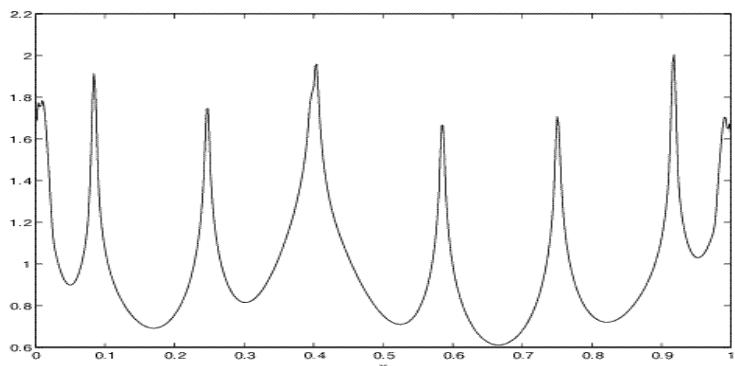
$$y_i = f(x_i) + \sigma \epsilon_i, \quad i = 1, \dots, n, \quad (3.3.14)$$



(a)

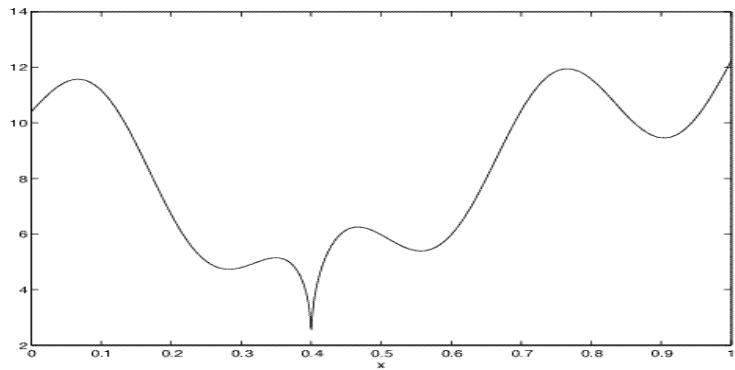


(b)

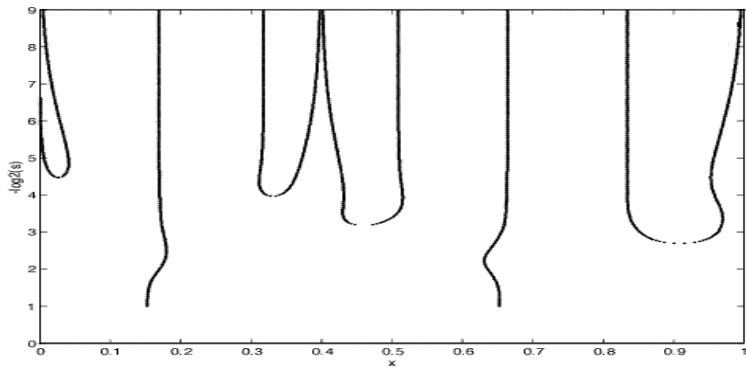


(c)

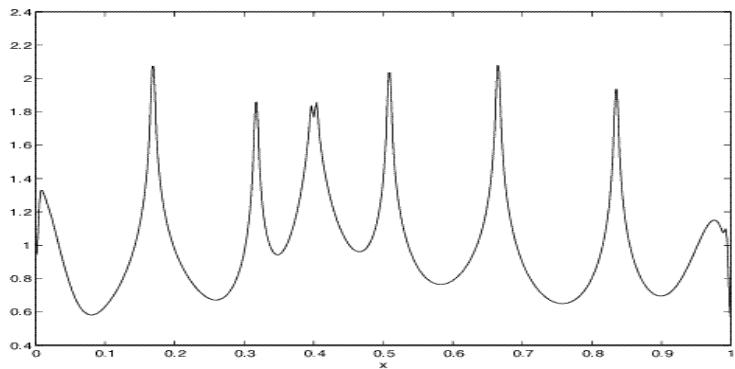
FIG. 3.4: (a) Signal f_1 (voir expression (3.3.10)). (b) Maxima d'ondelettes de la transformée en ondelettes de f_1 calculés avec une ondelette B-Spline d'ordre 4 avec deux moments nuls. (c) Intensité structurelle des maxima d'ondelettes.



(a)

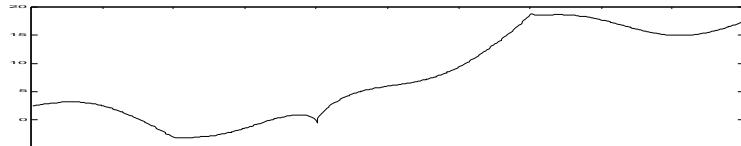


(b)

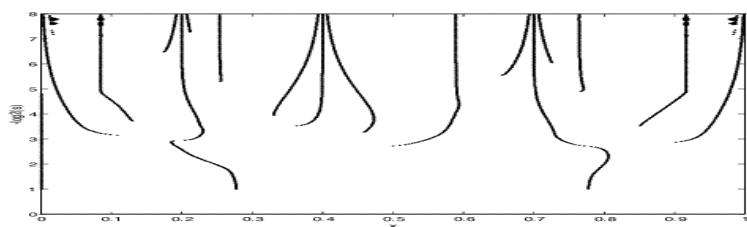


(c)

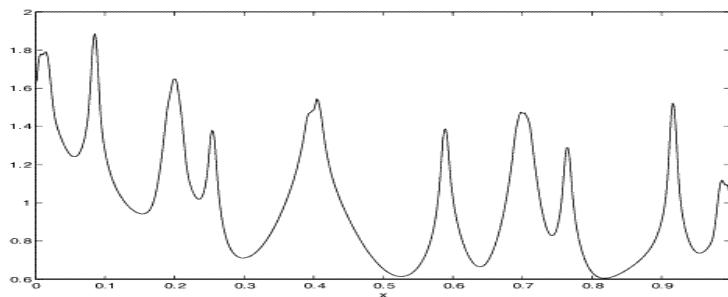
FIG. 3.5: (a) Signal f_1 (voir expression (3.3.10)). (b) Zeros-crossings de la transformée en ondelettes de f_1 calculés avec une ondelette B-Spline d'ordre 4 avec deux moments nuls. (c) Intensité structurelle des zero-crossings



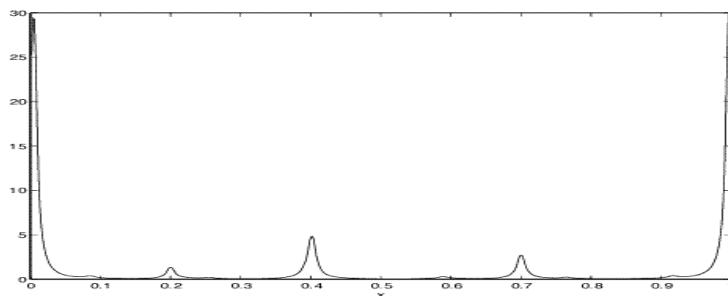
(a)



(b)

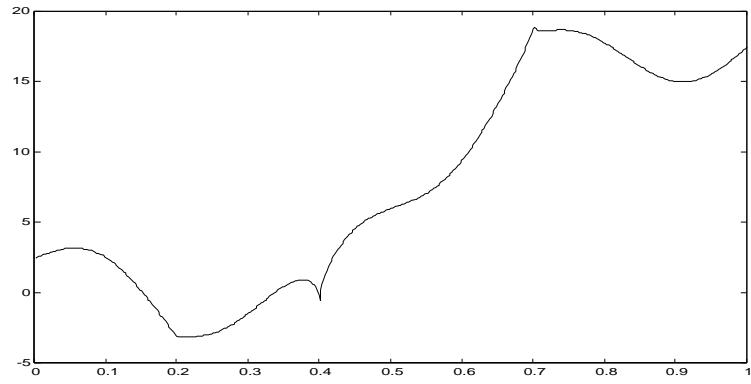


(c)

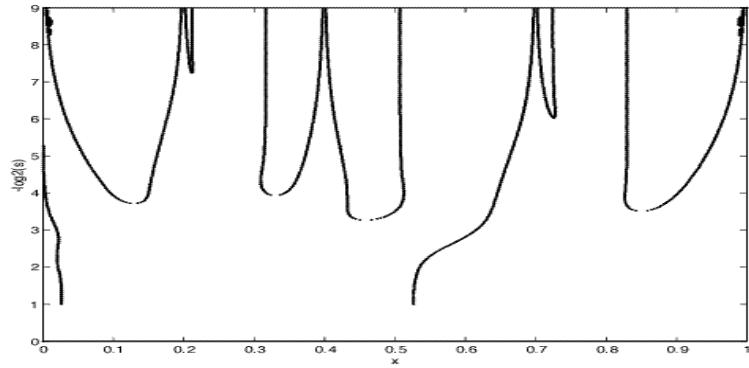


(d)

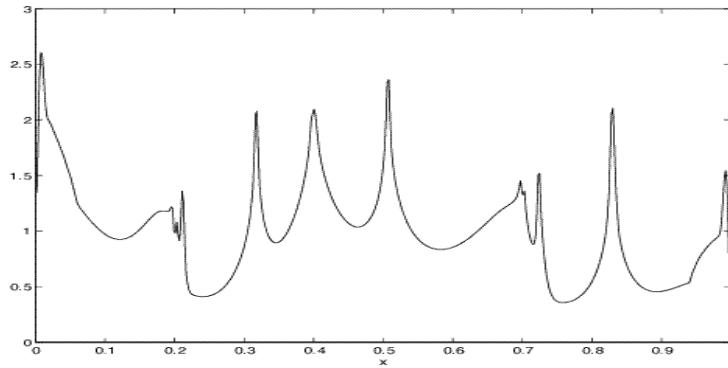
FIG. 3.6: (a) Signal f_3 (voir expression (3.3.12)). (b) Maxima d'ondelettes de la transformée en ondelettes de f_3 calculés avec une ondelette B-Spline d'ordre 4 avec deux moments nuls. (c) Intensité structurelle non pondérée des maxima d'ondelettes (d) Intensité structurelle des maxima d'ondelettes pondérée avec les fonctions $h_i(s) = \frac{W_s(f)(m_i(s))}{s^{r+1/2}}$



(a)



(b)



(c)

FIG. 3.7: (a) Signal f_3 (voir expression (3.3.12)). (b) Zeros-crossings de la transformée en ondelettes de f_3 calculés avec une ondelette B-Spline d'ordre 4 avec deux moments nuls. (c) Intensité structurelle des zero-crossings

où $x_i = \frac{i}{n}$, f est une fonction inconnue, σ est le niveau de bruit, et ϵ_i sont des variables aléatoires i.i.d. de loi normale, de moyenne nulle et de variance 1. En effet, définissons la fonction étagée f_n comme :

$$f_n(x) = \sum_{i=1}^n f(x_i) \mathbb{1}_{\{x_{i-1} \leq x < x_i\}}, \quad x \in [0, 1],$$

le processus de régression $\{Y_n(x) : x \in [0, 1]\}$:

$$Y_n(dx) = f_n(x)dx + \tau B(dx), \quad x \in [0, 1].$$

Quand $\tau = \frac{\sigma}{\sqrt{n}}$, Brown et Low [11] ont étudié le degré d'approximation des expériences (Y_n, Θ) par (Y, Θ) où Θ consiste en un ensemble (qui peut être large) de choix de fonctions sur $[0, 1]$. Pour une classe Θ de fonctions qui vérifient :

$$\sup_{\Theta} \|f - f_n\|_2^2 = o\left(\frac{1}{n}\right),$$

ils ont montré que les expériences (Y_n, Θ) et (Y, Θ) sont asymptotiquement indistinguables par n'importe quel test statistique, et que les résultats dans le modèle de bruit blanc (3.3.13) fournissent des résultats dans le modèle de régression nonparamétrique (3.3.14) et vice versa. Par exemple, dans le contexte de l'estimation nonparamétrique fonctionnelle, nous avons que si l est une fonction bornée et si \hat{f} est un estimateur mesurable, le risque dans le pire des cas, $\sup_{\Theta} El(\|\hat{f} - f\|_2^2)$ a la même limite asymptotique dans les deux expériences. En particulier, les problèmes (3.3.13) et (3.3.14) ont les mêmes risques asymptotiques minimax et un estimateur bon dans un modèle est encore bon dans l'autre. Donoho et Johnstone [30] ont établi des résultats similaires pour le risque non-bornée $E(\|\hat{f} - f\|_2^2)$ et plusieurs classes de fonctions Θ en utilisant des estimateurs à base d'ondelettes. Cependant, leur approche pour montrer ces résultats est différente de la méthodologie suivie par Brown et Low [11]. Elle est basée sur une définition précise d'une transformée en ondelettes empirique et sur des bornes de l'écart entre les coefficients d'ondelettes empiriques et les coefficients d'ondelettes théoriques. Nous pensons que pour le problème de la détection de landmarks, il serait possible de comparer des résultats de minimalité asymptotique entre les modèles de bruit blanc et de régression nonparamétrique en suivant les idées de Donoho et Johnstone [30].

Supposons que nous travaillons avec une ondelette ψ qui est la dérivée d'ordre r d'une fonction régulière i.e. $\psi = (-1)^r \theta^{(r)}$ avec $\int_{-\infty}^{+\infty} \theta(u)du \neq 0$ et que $f \in L^2([0, 1])$ est observée selon un modèle de bruit blanc (3.3.13). Si nous modifions une fonction qui appartient à $L^2(\mathbb{R})$ en la multipliant par la fonction indicatrice de $[0, 1]$, nous ne modifions pas sa régularité et ses landmarks sur $]0, 1[$. Nous supposerons donc que $f(x) = 0$ pour $x \notin [0, 1]$. La transformée en ondelettes de f à l'échelle $s > 0$ est alors égale à :

$$W_s(f)(x) = \int_{-\infty}^{+\infty} f(u) \psi_s(u - x) du = \int_0^1 f(u) \psi_s(u - x) du, \quad \text{pour } x \in [0, 1].$$

Notons qu'en prenant $f(x) = 0$ pour $x \notin [0, 1]$, nous introduisons généralement deux discontinuités aux points $x = 0$ et $x = 1$ ce qui génère des lignes de maxima d'ondelettes qui convergent vers 0 et 1. Cependant, ceci n'affecte pas la procédure d'estimation, puisque nous souhaitons uniquement détecter les landmarks de f qui sont inclus dans $]0, 1[$. Dans tout ce chapitre, nous supposerons que nous travaillons avec un espace probabilisé complet (Ω, A, P) , que B est un mouvement Brownien standard sur \mathbb{R} (i.e. avec $B(0) = 0$) et que nous avons choisi une version de B avec des réalisations continues. La transformée en ondelettes du bruit blanc $B(du)$ est alors définie comme :

$$W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x) B(du), \text{ pour } x \in [0, 1].$$

A une échelle donnée s , $\{W_s(B)(x); x \in [0, 1]\}$ est un processus Gaussien stationnaire de moyenne nulle et de fonction de covariance :

$$E(W_s(B)(x)W_s(B)(y)) = \int_{-\infty}^{+\infty} \psi_s(u - x)\psi_s(u - y)du, \quad (3.3.15)$$

$$Var(W_s(B)(x)) = \int_{-\infty}^{+\infty} \{\psi_s(u - x)\}^2 du = 1. \quad (3.3.16)$$

Dans ce qui suit, le paramètre de bruit τ (voir le modèle de bruit blanc (3.3.13)) sera remplacé par $\frac{\sigma}{\sqrt{n}}$. Les résultats asymptotiques seront établis pour $n \rightarrow \infty$. La transformée continue en ondelettes de Y est alors égale à :

$$W_{s,n}(Y)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x) Y(du) = W_s(f)(x) + \frac{\sigma}{\sqrt{n}} W_s(B)(x) \quad (3.3.17)$$

La proposition suivante étudie le comportement des lignes de zero-crossings et de maxima d'ondelettes du processus $W_{s,n}(Y)(x)$ quand f est continue par morceaux.

Proposition 3.5 *Supposons que $f \in L^2([0, 1])$ est continue par morceaux sur $[0, 1]$ et observée selon un modèle de bruit blanc (3.3.13). Supposons de plus que l'ondelette $\psi = (-1)^r \theta^{(r)}$ est causale. Alors, pour tout $n \in \mathbb{N}$, les lignes de maxima d'ondelettes et de zero-crossings du processus $W_{s,n}(Y)(x)$ sont presque sûrement jamais interrompues quand l'échelle diminue et peuvent être considérées comme des fonctions continues du paramètre d'échelle s .*

Preuve. Tout d'abord, supposons que ψ_s est à support compact, alors une intégration par partie donne :

$$W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u - x) B(du) = - \int_{-\infty}^{+\infty} \frac{1}{s} \psi'_s(u - x) B(u) du, \quad (3.3.18)$$

où $\psi'_s(u) = (-1)^r \frac{1}{\sqrt{s}} \theta^{(r+1)}(\frac{u}{s})$. Maintenant, si ψ n'est pas à support compact, nous pouvons montrer que l'équation (3.3.18) est toujours vraie en tronquant l'ondelette

ψ . Pour être plus précis, soit $\rho \in C^1(\mathbb{R})$ tel que $\rho(x) = 1$ si $|x| \leq 1$, $\rho(x) = 0$ si $|x| \geq 2$ et $0 \leq \rho \leq 1$. Pour $j \in \mathbb{N}$, soient $\rho_j = \rho(\frac{x}{j})$, $\psi_{s,j} = \rho_j \psi_s$ et :

$$W_{s,j}(B)(x) = \int_{-\infty}^{+\infty} \psi_{s,j}(u-x) B(du).$$

Notons que pour un $u \in \mathbb{R}$ fixé, $|\psi_{s,j}(u-x) - \psi_s(u-x)| \rightarrow 0$ quand $j \rightarrow \infty$ et que $|\psi_{s,j}(u-x) - \psi_s(u-x)| \leq 2\psi_s(u-x)$. Ainsi, puisque $\psi \in L^2(\mathbb{R})$, le théorème de convergence dominée de Lebesgue implique que pour $j \rightarrow \infty$:

$$E((W_{s,j}(B)(x) - W_s(B)(x))^2) = \int_{-\infty}^{+\infty} (\psi_{s,j}(u-x) - \psi_s(u-x))^2 du \rightarrow 0.$$

Puisque, $\psi_{s,j}$ est à support compact, $W_{s,j}(B)(x) = - \int_{-\infty}^{+\infty} \psi'_{s,j}(u-x) B(u) du$, où :

$$\psi'_{s,j}(u-x) = \frac{1}{s} \rho'_j(u-x) \psi'_s(u-x) + \rho'_j(u-x) \psi_s(u-x),$$

et $\rho'_j(u) = \frac{1}{j} \rho'(\frac{u}{j})$. Puisque $\rho'_j(u) = 0$ pour $|u| \leq j$ nous avons que pour un $u \in \mathbb{R}$ fixé, $|\psi'_{s,j}(u-x) - \frac{1}{s} \psi'_s(u-x)| \rightarrow 0$ quand $j \rightarrow \infty$. Notons que d'après nos hypothèses sur ρ , il existe une constante M telle que pour tout $j \in \mathbb{N}^*$, $\rho'_j \leq M$ et $\rho_j \leq M$. Ainsi, pour tout $j \in \mathbb{N}^*$, $|\psi'_{s,j}(u-x) - \frac{1}{s} \psi'_s(u-x)| \leq 2M(\frac{1}{s} \psi'_s(u-x) + \psi_s(u-x))$. Par la loi du logarithme itéré, nous avons qu'il existe un sous-ensemble $\Omega_0 \subset \Omega$ tel que $P(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$, $\lim_{u \rightarrow \infty} \frac{|B(\omega, u)|}{\sqrt{2u \log_2(u)}} = 1$. Puisque ψ et ψ' sont à décroissance rapide, nous obtenons que pour tout $\omega \in \Omega_0$, $u \mapsto (\frac{1}{s} \psi'_s(u-x) + \psi_s(u-x)) B(w, u)$ est intégrable sur \mathbb{R} . Donc, d'après le théorème de convergence dominée de Lebesgue, il suit que pour tout $\omega \in \Omega_0$:

$$W_{s,j}(B)(\omega, x) \rightarrow - \int_{-\infty}^{+\infty} \frac{1}{s} \psi'_s(u-x) B(\omega, u) du.$$

Finalement, nous avons montré que, $E((W_{s,j}(B)(x) - W_s(B)(x))^2) \rightarrow 0$ et que $W_{s,j}(B)(x)$ converge presque sûrement vers $- \int_{-\infty}^{+\infty} \frac{1}{s} \psi'_s(u-x) B(u) du$ quand $j \rightarrow \infty$, ce qui implique que $W_s(B)(x) = - \int_{-\infty}^{+\infty} \frac{1}{s} \psi'_s(u-x) B(u) du$ et prouve que l'équation (3.3.18) est également vraie quand ψ n'est pas à support compact.

Ainsi, pour tout $\omega \in \Omega_0$ nous avons que :

$$W_s(B)(\omega, x) = - \frac{1}{s} \tilde{\psi}'_s \star B(\omega, .)(x).$$

avec $\tilde{\psi}'_s(u) = s^{-1/2} \psi'(-\frac{u}{s})$. Soit $x_1, \dots, x_q \in]0, 1[$ les points de discontinuités de f . Soit F une primitive de f définie sur les points de $]0, 1[$ où f est continue et telle que $F(u) = 0$ pour $u \notin]0, 1[$. Notons que F est également discontinue aux points x_1, \dots, x_q . Soit $\tilde{f}(u) = F^{(1)}(u)$ une dérivée de F prise au sens des distributions.

Etant donné que \tilde{f} et f sont égales presque partout et que ψ est à décroissance rapide :

$$W_s(f)(x) = W_s(\tilde{f})(x) = \int_{-\infty}^{+\infty} F^{(1)}(u)\psi_s(u-x)du = -\frac{1}{s} \int_{-\infty}^{+\infty} F(u)\psi'_s(u-x)du.$$

Ainsi pour tout $\omega \in \Omega_0$,

$$W_{s,n}(Y)(\omega, x) = -\frac{1}{s} \left\{ \int_{-\infty}^{+\infty} (F(u) + \frac{\sigma}{\sqrt{n}}B(\omega, u))\psi'_s(u-x)du \right\}.$$

Notons que pour presque tout $\omega \in \Omega_0$, $u \mapsto F(u) + \frac{\sigma}{\sqrt{n}}B(\omega, u)$ est une distribution tempérée. Puisque $\psi = (-1)^r\theta^{(r)}$ est causale, ψ' est également causale. Notons que la transformée en ondelettes peut être facilement étendue aux cas des distributions tempérées (i.e. à des fonctions généralisées). Ainsi, en étendant la notion de causalité pour $f \in S(\mathbb{R})$ et en répétant les preuves du Théorème 3.4, nous obtenons que pour tout $\omega \in \Omega_0$, les lignes de maxima d'ondelettes et de zero-crossings de $W_{s,n}(Y)(\omega, x)$ sont des fonctions continues qui ne sont jamais interrompues quand l'échelle diminue. \square

Dans les sections suivantes, nous supposerons que l'ondelette ψ est causale et qu'elle est la dérivée d'ordre r d'une fonction régulière.

3.4 Estimation des maxima d'ondelettes

Nous considérons le problème de l'estimation des maxima d'ondelettes dans un modèle de bruit blanc. Nous supposons qu'un signal inconnu régulier, observé avec du bruit, a une singularité d'ordre α au point $x_0 \in \mathbb{R}$, et nous voulons estimer la position et l'amplitude des maxima d'ondelettes correspondants à différentes échelles.

Soit m^* un entier fixé strictement positif tel que $r \geq m^* + 1$ où r est le nombre de moments nuls de l'ondelette ψ . Soit $0 \leq \alpha < r$, $0 < a < b < 1$ et $x_0 \in [a, b]$. Nous supposons que f vérifie les hypothèses suivantes :

Hypothèse 3.1 f est r fois continûment dérivable pour tout $x \in [0, 1]$, $x \neq x_0$.

Hypothèse 3.2 f a une singularité d'ordre α au point x_0 .

Dans cette section, nous supposerons que l'ondelette ψ est à support compact égal à $[-C, C]$. Soit $V_s = [x_0 - sC, x_0 + sC]$ pour une échelle donné $s \geq 0$. Alors, les Hypothèses 3.1 et 3.2 impliquent des propriétés spécifiques de décroissance de l'amplitude de la transformée en ondelettes à l'extérieur et à l'intérieur du cône d'influence de x_0 :

Proposition 3.6 Supposons qu'une fonction $f \in L^2([0, 1])$ satisfasse les Hypothèses 3.1 et 3.2. Alors, il existe une constante $A_1 > 0$ telle que pour tout $s > 0$ et tout $x \notin V_s$:

$$|W_s(f)(x)| \leq A_1 s^{r+1/2}.$$

Preuve. Soit $\epsilon > 0$. Puisque f est r fois continûment dérivable sur $[0, 1] \setminus \{x_0\}$, elle est uniformément Lipschitzienne d'ordre r sur $[0, x_0 - \epsilon]$ avec une constante Lipschitzienne qui est indépendante de ϵ . Alors, d'après le Théorème 3.1 nous avons qu'il existe une constante $A_1 > 0$ (indépendante de ϵ) telle que pour tout $(x, s) \in [0, x_0 - \epsilon] \times \mathbb{R}^+$, $|W_s(f)(x)| \leq A_1 s^{r+1/2}$. Soit $s > 0$ et $x \notin V_s$. Supposons que $x < x_0$. Pour ϵ suffisamment petit, $x \in [0, x_0 - \epsilon]$ et alors $|W_s(f)(x)| \leq A_1 s^{r+1/2}$. La preuve est identique pour $x > x_0$. \square

La proposition suivante est similaire au Corollaire 1 de Jaffard [48].

Proposition 3.7 *Supposons qu'une fonction $f \in L^2([0, 1])$ satisfasse les Hypothèses 3.1 et 3.2. Alors, il existe une constante $A_2 > 0$ et une suite $(x_p, s_p) \rightarrow (x_0, 0)$, telles que $x_p \in V_{s_p}$ et :*

$$\lim_{p \rightarrow \infty} \frac{|W_{s_p}(f)(x_p)|}{s_p^{\alpha+1/2}} = A_2,$$

Nous appellerons une telle suite (x_p, s_p) une suite minimisante pour f au point x_0 .

Preuve. Puisque la régularité Lipchitzienne de f au point x_0 est α , il existe une constante $A_2 > 0$ telle que pour tout $s > 0$ et tout $x \in V_s$, $|W_s(f)(x)| \leq A_2 s^{\alpha+1/2}$ (voir Théorème 3.2). Soit $\epsilon > 0$ et supposons qu'il existe une échelle $s_\epsilon > 0$ telle que pour tout $s \leq s_\epsilon$ et tout $x \in V_s$, $|W_s(f)(x)| \leq A_2 s^{\alpha+\epsilon+1/2}$. La condition (3.3.9) du Théorème 3.2 est donc satisfaite à l'intérieur du cône d'influence de x_0 . D'après la Proposition 3.6, nous avons qu'il existe une constante $A_1 > 0$ telle que pour tout $s > 0$ et tout $x \notin V_s$, $|W_s(f)(x)| \leq A_1 s^{r+1/2}$, ce qui implique qu'il existe une constante $A' > 0$ telle que pour tout $s > 0$ et tout $x \notin V_s$, $|W_s(f)(x)| \leq A' s^{r-\alpha-\epsilon+1/2} C s^{\alpha+\epsilon} \leq A' s^{r-\alpha-\epsilon+1/2} |x - x_0|^{\alpha+\epsilon}$. Alors, si ϵ est suffisamment petit, la condition (3.3.9) du Théorème 3.2 est donc aussi satisfaite à l'extérieur du cône d'influence de x_0 , ce qui implique que f est ponctuellement Lipschitzienne d'ordre $\alpha + \epsilon > \alpha$ au point x_0 (notons que ϵ peut être choisi de sorte que $\alpha + \epsilon$ ne soit pas un entier) et contredit le fait que la régularité Lipschitzienne de f au point x_0 est α . Ainsi, pour tout $\epsilon > 0$ suffisamment petit, il existe nécessairement une suite $(s_p, x_p) \rightarrow (0, x_0)$ avec $x_p \in V_{s_p}$ telle que $|W_{s_p}(f)(x_p)| > A_2 s_p^{\alpha+\epsilon+1/2}$, et donc nous obtenons que $A_2 s_p^\epsilon < \frac{|W_{s_p}(f)(x_p)|}{s_p^{\alpha+1/2}} \leq A_2$. Notons que s_p peut être choisie de sorte que $s_p < \epsilon$ et donc en laissant ϵ converger vers zéro, nous obtenons que nous pouvons construire une suite $(x_p, s_p) \rightarrow (x_0, 0)$, telle que $x_p \in V_{s_p}$ et $\lim_{p \rightarrow \infty} \frac{|W_{s_p}(f)(x_p)|}{s_p^{\alpha+1/2}} = A_2$. \square

Proposition 3.8 *Supposons qu'une fonction $f \in L^2([0, 1])$ satisfasse les Hypothèses 3.1 et 3.2. Alors, il existe une constante $A_2 > 0$ et une ligne de maxima d'ondelettes $m(s)$ qui converge vers x_0 telle que :*

$$\lim_{s \rightarrow 0} \frac{|W_s(f)(m(s))|}{s^{\alpha+1/2}} = A_2.$$

Preuve. Etant donné que f vérifie les Hypothèses 3.1 et 3.2, la Proposition 3.3 implique qu'il existe une ligne de maxima d'ondelettes $m(s)$ qui converge vers x_0 et appartient à V_s . Soit $g(s) = \frac{|W_s(f)(m(s))|}{s^{\alpha+1/2}}$ pour $s > 0$. Puisque l'ondelette ψ est causale, le Théorème 3.4 montre que $m(\cdot)$ est C^1 pour tout $s > 0$ suffisamment petit. Notons que le long d'une ligne de maxima d'ondelettes qui converge vers une singularité isolée, $W_s(f)(m(s))$ a un signe constant pour tout s dans le support de $m(s)$. Donc, si l'on suppose que $W_s(f)(m(s)) > 0$, nous obtenons que $g(s) = \frac{W_s(f)(m(s))}{s^{\alpha+1/2}}$ est C^1 pour $s > 0$ suffisamment petit. Alors, puisque $W_s(f)(m(s))$ est un maximum local de $x \mapsto W_s(f)(x)$, nous obtenons que $g'(s) = -(\alpha + 1) \frac{W_s(f)(m(s))}{s^{\alpha+3/2}}$ est négatif pour tout $s > 0$ suffisamment petit. Ainsi, g est une fonction strictement décroissante pour $s > 0$ suffisamment petit. D'après la Proposition 3.7, nous avons immédiatement qu'il existe une suite $s_p \rightarrow 0$ et une constante $A_2 > 0$ telles que $\lim_{p \rightarrow \infty} g(s_p) = A_2$. Puisque g est une fonction continue et strictement décroissante pour $s > 0$ petit, nous avons nécessairement que $\lim_{s \rightarrow 0} g(s) = A_2$. La preuve est similaire si nous supposons que $W_s(f)(m(s)) < 0$. \square

La section suivante montre que les Propositions 3.6 et 3.8 peuvent être utilisées pour construire une procédure de seuillage pour détecter les maxima d'ondelettes créés par la singularité au point x_0 quand la fonction f est observée avec du bruit.

3.4.1 Détection des maxima d'ondelettes

Nous voulons estimer la position et l'amplitude des maxima d'ondelettes qui correspondent à la singularité d'ordre α de la fonction f au point x_0 . Dans ce but, nous allons montrer qu'il existe une échelle grossière s_r et une échelle plus fine s_{α^*} telles que :

- pour $s_r \geq s \geq s_{\alpha^*}$, les maxima d'ondelettes de $|W_s(f)(x)|$ créés par la singularité au point x_0 dominent les maxima d'ondelettes de $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$.
- pour $s \leq s_{\alpha^*}$, $|W_{s,n}(Y)(x)|$ est dominé par $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$
- l'échelle grossière s_r est choisie de sorte que les maxima d'ondelettes qui correspondent à la singularité au point x_0 dominent les maxima d'ondelettes qui peuvent être localisés dans les régions “régulières” du signal f .

Une fois que les niveaux s_r et s_{α^*} ont été choisis, la position et l'amplitude des maxima d'ondelettes dus aux singularités du signal f peuvent être détectés en examinant les maxima de $\{|W_{s,n}(Y)(x)|; x \in [0, 1]\}$ qui sont plus grands qu'un seuil bien choisi pour $s_r \geq s \geq s_{\alpha^*}$.

3.4.1.1 Choisir un seuil

La première étape de notre procédure d'estimation est le choix d'un seuil λ tel que les maxima d'ondelettes créés par la fonction f puissent être distingués de ceux créés par les fluctuations du bruit. Rappelons que l'ondelette ψ a un support compact égal à $[-C, C]$, et que $V_s = [x_0 - Cs, x_0 + Cs]$ pour $s > 0$ définit le cône

d'influence de la singularité au point x_0 .

Notre idée est de sélectionner une échelle grossière s_r où les amplitudes de $|W_s(f)(x)|$ et $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$ sont équilibrées si $x \notin V_s$, et une échelle plus fine s_{α^*} telle que pour $s_r \geq s \geq s_{\alpha^*}$, les maxima d'ondelettes de $|W_s(f)(x)|$ qui sont localisés dans V_s dominent $\frac{\sigma}{\sqrt{n}}|W_s(B)(x)|$. Puisque $\{W_s(B)(x); x \in \mathbb{R}\}$ est un processus Gaussien dont la variance est égale à un, l'ordre d'amplitude du bruit à une échelle donné s est approximativement égal à $\frac{\sigma}{\sqrt{n}}$. Rappelons que d'après la Proposition 3.6, si $x \notin V_s$ alors $|W_s(f)(x)| \leq A_1 s^{r+1/2}$. Ainsi, une idée naturelle est de définir s_r comme la solution de l'équation $s_r^{r+1/2} = \sigma n^{-\frac{1}{2}}$ pour un $n \in \mathbb{N}^*$ fixé. Dans cette section, nous supposerons que l'ordre de la singularité au point x_0 est connu, de sorte que nous puissions définir s_α et s_{α^*} comme les solutions des équations $s_\alpha^{\alpha+1/2} = \sigma n^{-\frac{1}{2}}$ et $s_{\alpha^*}^{\alpha+1/2} = s_\alpha^{\alpha+1/2} c_n \sqrt{\log(n)}$ pour $n \in \mathbb{N}^*$ donné, où $c_n \rightarrow \infty$ avec $c_n = o((\log(n))^{1/2})$ (i.e. $c_n \rightarrow \infty$ avec une vitesse arbitrairement petite). De plus, puisque nous voulons détecter les maxima en module pour $s_r \geq s \geq s_{\alpha^*}$ nous devons garantir que $s_r > s_{\alpha^*}$ ce qui sera supporté par l'hypothèse suivante, qui implique que $-\log(s_{\alpha^*}) + \log(s_r) \rightarrow \infty$ quand $n \rightarrow \infty$:

Hypothèse 3.3 $c_n^2 \log(n) = o(n^{1-\frac{2\alpha+1}{2r+1}})$.

Dans les propositions suivantes, nous allons montrer qu'il est possible de distinguer les maxima d'ondelettes créés par la singularité au point x_0 de celles créées par les fluctuations du bruit pour $s_r \geq s \geq s_{\alpha^*}$. Cependant, en pratique, α est inconnu, et une procédure pour estimer l'échelle s_{α^*} est donc proposée dans la section 3.4.1.2.

Proposition 3.9 Soit $f \in L^2([0, 1])$ satisfaisant les hypothèses 3.1 et 3.2, alors :

– il existe une constante $A_1 > 0$ telle que pour tout $s \leq s_r$ et tout $x \notin V_s$:

$$|W_s(f)(x)| \leq A_1 \frac{\sigma}{\sqrt{n}} \text{ pour } n \in \mathbb{N}^*.$$

– il existe une constante $A_3 > 0$ et une ligne de maxima d'ondelettes $m(s)$ qui converge vers x_0 telles que pour $s_r \geq s \geq s_{\alpha^*}$:

$$|W_s(f)(m(s))| \geq A_3 \frac{\sigma}{\sqrt{n}} c_n \sqrt{\log(n)} \text{ pour tout } n \in \mathbb{N}^* \text{ suffisamment large.}$$

Preuve. Pour $s \leq s_r$, nous avons que $s^{r+1/2} \leq s_r^{r+1/2}$. Alors, d'après la Proposition 3.6 et la définition de s_r , nous avons immédiatement qu'il existe une constante A_1 telle que la première partie de la proposition soit vraie. D'après la Proposition 3.8, nous avons qu'il existe une constante A_2 et une ligne de maxima d'ondelettes $m(s) \in V_s$ qui converge vers x_0 telle que :

$$\lim_{s \rightarrow 0} \frac{|W_s(f)(m(s))|}{s^{\alpha+1/2}} = A_2.$$

Notons que, pour $n \rightarrow \infty$, $s_{\alpha^*} \rightarrow 0$, $s_r \rightarrow 0$ et sous l'Hypothèse 3.3 $s_{\alpha^*} < s_r$. Ainsi, pour tout n suffisamment large, il existe une constante $A_3 > 0$ telle que pour $s_r \geq s \geq s_{\alpha^*}$, $|W_s(f)(m(s))| \geq A_3 s^{\alpha+1/2} \geq A_3 s_{\alpha^*}^{\alpha+1/2}$, et la seconde partie de la proposition dérive alors de la définition de s_{α^*} . \square

Considérons la relation (3.3.17), où $|W_s(f)(x)|$ est observée avec du bruit. La Proposition 3.9 peut être utilisée pour distinguer les maxima d'ondelettes créés par la singularité au point x_0 , de ceux créés par les fluctuations du bruit. Dans ce but, nous allons avoir besoin du lemme suivant qui donne la distribution asymptotique des maxima en module de la transformée en ondelettes continue d'un mouvement Brownien quand l'échelle s tend vers zéro :

Lemme 3.1 *Rappelons que $W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x) B(du)$ où B est un mouvement Brownien standard. Soit $\tilde{M}_s = \max_{x \in [0,1]} |W_s(B)(x)|$. Alors, quand $s \rightarrow 0$:*

$$P \left(\sqrt{2|\log(s)|} \tilde{M}_s - 2|\log(s)| - \log \left(\frac{\left[\int_{-\infty}^{+\infty} \{\psi'(u)\}^2 du \right]^{1/2}}{2\pi} \right) \leq t \right) \rightarrow \exp(-2e^{-t}). \quad (3.3.19)$$

Preuve. Ce lemme est basé sur les Théorème 8.2.7. et Théorème 1.8.3 du livre de Leadbetter *et. al.*. Le lemme suivant est une conséquence immédiate de ces deux théorèmes :

Lemme 3.2 *Soit $\{\xi(t); t \geq 0\}$ un processus stationnaire normal standardisé dont la fonction de covariance $r(\tau)$ vérifie :*

$$r(\tau) = 1 - \frac{\lambda_2 \tau^2}{2} + o(\tau^2), \text{ quand } \tau \rightarrow 0,$$

où λ_2 est un réel fini. Soit $\tilde{M}(T) = \max_{t \in [0,T]} |\xi(t)|$, alors quand $T \rightarrow \infty$:

$$P \left(\sqrt{2\log(T)} \tilde{M}(T) - 2\log(T) - \log \left(\frac{\lambda_2^{1/2}}{2\pi} \right) \leq t \right) \rightarrow \exp(-2e^{-t}).$$

Par la propriété d'auto-similarité du mouvement Brownien :

$$\text{Loi} \left(W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x) B(du) \right) = \text{Loi} \left(\int_{-\infty}^{+\infty} \psi(u-s^{-1}x) B(du) \right). \quad (3.3.20)$$

Soit $\{Z(y) = \int_{-\infty}^{+\infty} \psi(u-y) B(du); y \geq 0\}$. L'équation (3.3.20) prouve que :

$$\text{Loi}(\tilde{M}_s) = \text{Loi} \left(\max_{y \in [0,s^{-1}]} |Z(y)| \right).$$

Notons que Z est un processus stationnaire normal standardisé dont la fonction de covariance vérifie :

$$r_Z(x, y) = E(Z(x)Z(y)) = \int_{-\infty}^{+\infty} \psi(u - (x - y))\psi(u)du.$$

Par un développement de Taylor au voisinage de u et d'après les hypothèses de régularité sur ψ , nous obtenons que si x et y sont suffisamment proches :

$$\begin{aligned} r_Z(x, y) &= \int_{-\infty}^{+\infty} \{\psi(u)\}^2 du - (x - y) \int_{-\infty}^{+\infty} \psi'(u)\psi(u)du \\ &\quad + \frac{(x - y)^2}{2} \int_{-\infty}^{+\infty} \psi''(u)\psi(u)du + o((x - y)^2). \end{aligned}$$

Ensuite, une intégration par parties montre que :

$$\begin{aligned} \int_{-\infty}^{+\infty} \psi'(u)\psi(u)du &= 0, \\ \int_{-\infty}^{+\infty} \psi''(u)\psi(u)du &= - \int_{-\infty}^{+\infty} \{\psi'(u)\}^2 du. \end{aligned}$$

Ainsi,

$$r_Z(x, y) = 1 - \frac{\int_{-\infty}^{+\infty} \{\psi'(u)\}^2 du}{2}(x - y)^2 + o((x - y)^2).$$

Alors, l'équation (3.3.19) est une conséquence directe du Lemme 3.2 et l'expression de $r_Z(x, y)$ donnée ci-dessus. \square

Définissons le seuil : $\lambda = \frac{\sigma\sqrt{2\log(n\log(n))}}{\sqrt{n}}$. La proposition suivante prouve que ce seuil peut être utilisé pour contrôler l'amplitude de $|W_{s,n}(Y)(x)|$ à l'extérieur et l'intérieur du cône d'influence de la singularité au point x_0 :

Proposition 3.10 Soit $V_s = [x_0 - Cs, x_0 + Cs]$, alors pour tout $s_r \geq s > \frac{1}{n}$:

$$P(\max_{x \notin V_s} \{|W_{s,n}(Y)(x)|\} \leq \lambda) \rightarrow 1, \text{ quand } n \rightarrow \infty.$$

Soit k_s le nombre de maxima d'ondelettes localisés dans V_s qui sont tels que : $|W_s(f)(m_{s,i})| \geq A_3 s^{\alpha+1/2}$, $i = 1, \dots, k_s$, où A_3 est la constante positive définie dans la Proposition 3.9, et où les $m_{s,i}$ dénotent les positions des maxima d'ondelettes à l'échelle $s > 0$. Alors, pour $s_r \geq s \geq s_{\alpha^*}$:

$$P(|W_{s,n}(Y)(m_{s,i})| \geq \lambda) \rightarrow 1, \text{ quand } n \rightarrow \infty, \text{ pour } i = 1, \dots, k_s.$$

Preuve. Par inégalité triangulaire, nous avons $|W_{s,n}(Y)(x)| \geq \lambda \Rightarrow \sigma n^{-1/2} |W_s(B)(x)| \geq \lambda - |W_s(f)(x)|$. D'après la Proposition 3.9, nous avons qu'il existe une constante

$A_1 > 0$ telle que pour $s \leq s_r$, $\max_{x \notin V_s} |W_s(f)(x)| \leq A_1 \frac{\sigma}{\sqrt{n}}$. Ainsi, il suit que pour $s \leq s_r$:

$$\begin{aligned} P(\max_{x \notin V_s} |W_{s,n}(Y)(x)| \geq \lambda) &\leq P(\max_{x \notin V_s} |W_s(B)(x)| \geq \sqrt{2 \log(n \log(n))} - A_1) \\ &\leq P(\tilde{M}_n \geq T_n), \end{aligned}$$

où $\tilde{M}_n = \sqrt{2|\log(s)|} \max_{x \notin V_s} |W_s(B)(x)| - 2|\log(s)| - K$, $T_n = \sqrt{2|\log(s)|} \sqrt{2 \log(n \log(n))} - A_1 \sqrt{2|\log(s)|} - 2|\log(s)| - K$, et $K = \log\left(\frac{[\int_{-\infty}^{+\infty} \{\psi'(u)\}^2 du]^{1/2}}{2\pi}\right)$.

Notons que $T_n = 2|\log(s)| \left(\sqrt{\frac{\log(\log(n))+\log(n)}{|\log(s)|}} - 1 - \frac{A_1}{\sqrt{2|\log(s)|}} - \frac{K}{2|\log(s)|} \right)$. Puisque

$$\frac{1}{n} < s \leq s_r = (\sigma^2/n)^{\frac{1}{2r+1}}, \text{ ceci implique que } \frac{1}{\log(n)} < \frac{1}{|\log(s)|} \leq \frac{1}{(2r+1)^{-1}(\log(n)-2\log(\sigma))}.$$

Donc, $\sqrt{\frac{\log(\log(n))}{\log(n)} + 1} < \sqrt{\frac{\log(\log(n))+\log(n)}{|\log(s)|}} \leq \sqrt{\frac{\frac{\log(\log(n))}{\log(n)} + 1}{(2r+1)^{-1}(1-2\frac{\log(\sigma)}{\log(n)})}}$. Ensuite, puisque

$$(2r+1)^{-1}(1-2\frac{\log(\sigma)}{\log(n)}) < 1, \text{ nous avons } \sqrt{\frac{\log(\log(n))+\log(n)}{|\log(s)|}} > 1 \text{ ce qui implique que } T_n \rightarrow \infty.$$

Ainsi, pour n suffisamment large, le lemme 3.1 implique que $P(\tilde{M}_n \geq T_n) = 1 - \exp(-2e^{-T_n}) \rightarrow 0$ ce qui prouve la première partie de la proposition.

Maintenant, d'après la Proposition 3.9, nous avons qu'il existe une constante $A_3 > 0$ telle que pour $s_r \geq s \geq s_{\alpha^*}$ il y a au moins un maxima d'ondelette $m(s) \in V_s$ tel que $|W_s(f)(m(s))| \geq A_3 s^{\alpha+1/2} \geq A_3 \frac{\sigma}{\sqrt{n}} c_n \sqrt{\log(n)}$ pour tout $n \in \mathbb{N}^*$ suffisamment large. Par inégalités triangulaires, nous avons que $|W_{s,n}(Y)(m_s)| \leq \lambda \Rightarrow \sigma n^{-1/2} |W_s(B)(m_s)| \geq |W_s(f)(m_s)| - \lambda$. Ainsi, pour $s_r \geq s \geq s_{\alpha^*}$:

$$\begin{aligned} P(|W_{s,n}(Y)(m_s)| \leq \lambda) &\leq P(\sigma n^{-1/2} |W_s(B)(m_s)| \geq |W_s(f)(m_s)| - \lambda) \\ &\leq P(|W_s(B)(m_s)| \geq A_3 c_n \sqrt{\log(n)} - K \sqrt{2 \log(n \log(n))}) \\ &\leq 2 - 2\Phi(T_n), \text{ où} \end{aligned}$$

$T_n = A_3 c_n \sqrt{\log(n)} - K \sqrt{2 \log(n \log(n))}$ et Φ est la fonction de répartition de la loi $N(0, 1)$ (loi normale standard). Puisque $T_n = \sqrt{\log(n)} \left(A_3 c_n - K \sqrt{2} \sqrt{\frac{\log(\log(n))}{\log(n)}} + 1 \right)$, nous avons $T_n \rightarrow \infty$ et donc $2 - 2\phi(T_n) \rightarrow 0$. La preuve est similaire pour $m_{s,i}, i = 2, \dots, k_s$. \square

3.4.1.2 Estimation des maxima d'ondelettes quand α est inconnu

La Proposition 3.9 conduit à une procédure très simple d'estimation de la position des maxima d'ondelettes qui correspondent à une singularité aux échelles $s_r \geq s \geq s_{\alpha^*}$. Cependant, en pratique, le paramètre α est inconnu, et donc le niveau s_{α^*} doit être estimé. Soit \hat{s}_{α^*} le niveau de résolution le plus fin où $\max_{x \in [0,1]} |W_{s,n}(Y)(x)|$ dépasse le seuil λ . La proposition suivante montre que \hat{s}_{α^*} se trouve entre s_{α^*} et s_α :

Proposition 3.11 *Quand $n \rightarrow \infty$,*

$$P(s_{\alpha^*} \geq \hat{s}_{\alpha^*} > s_{\alpha}) \rightarrow 1,$$

$$\text{i.e. } P\left(|\hat{s}_{\alpha^*} - s_{\alpha^*}| \leq \left(\frac{\sigma}{\sqrt{n}} c_n \sqrt{\log(n)}\right)^{\frac{1}{1+2\alpha}}\right) \rightarrow 1.$$

Preuve. D'après la définition de s_{α^*} et la Proposition 3.10, nous avons immédiatement que $P(s_{\alpha^*} \geq \hat{s}_{\alpha^*}) \rightarrow 1$. Si nous répétons la preuve des Proposition 3.9 et 3.10 pour α au lieu de r , nous obtenons que :

$$P\left(\max_{s_{\alpha} \geq s > \frac{1}{n}} \max_{x \in V_s} |W_{s,n}(Y)(x)| \leq \lambda\right) \rightarrow 1,$$

ce qui termine la preuve. \square

Une fois que le niveau s_{α^*} a été estimé, une procédure très simple peut être utilisée pour estimer la position des maxima en module. Remarquons que l'Hypothèse 3.3 implique que $s_{\alpha^*} < s_r$. Ainsi, pour $s_r \geq s \geq \hat{s}_{\alpha^*}$, nous pouvons définir $\hat{m}_{s,1}, \dots, \hat{m}_{s,\hat{k}_s}$ comme les positions des plus grands maxima de $|W_{s,n}(Y)(x)|$ qui dépassent le seuil λ à un niveau de résolution donné s (avec $|W_{s,n}(Y)(\hat{m}_{s,1})| \geq \dots \geq |W_{s,n}(Y)(\hat{m}_{s,\hat{k}_s})|$).

3.4.2 Plusieurs singularités

Rappelons que l'ondelette ψ a r moments nuls avec $r \geq m^* + 1$ où m^* est un entier fixé. Supposons que la fonction f a q singularités d'ordre α_i aux points $x_1 < \dots < x_q$, où q est un entier fini. En d'autres termes, f est maintenant supposée vérifier les hypothèses suivantes :

Hypothèse 3.4 *Il existe q points $x_1 < \dots < x_q \in [a, b]$, avec $0 < a < b < 1$, où la régularité Lipschitzienne de f est respectivement $\alpha_1, \dots, \alpha_q$. Nous supposerons que $x_{i+1} - x_i > \kappa$, $i = 1, \dots, q-1$ où $\kappa > 0$ est une constante inconnue. Nous imposerons également que $\alpha_{\max} < m^* + 1 \leq r$, où $\alpha_{\max} = \max\{\alpha_i; i = 1, \dots, q\}$.*

Hypothèse 3.5 *f est r fois continûment dérivable aux points $x \in [0, 1] \setminus \{x_1, \dots, x_q\}$.*

Comme dans la section 3.4.1.1, nous définissons s_r et s_{α_i} comme les solutions des équations $s_r^{r+1/2} = \sigma n^{-\frac{1}{2}}$ et $s_{\alpha_i}^{\alpha_i+1/2} = \sigma n^{-\frac{1}{2}}$ pour $n \in \mathbb{N}$ et $i = 1, \dots, q$. Nous définissons également $s_{\alpha_i^*}$ comme la solution de l'équation $s_{\alpha_i^*}^{\alpha_i+1/2} = s_{\alpha_i}^{\alpha_i+1/2} c_n \sqrt{\log(n)}$, où $c_n \rightarrow \infty$ et $c_n = o((\log(n))^{1/2})$. Rappelons que l'ondelette ψ est à support compact égal à $[-C, C]$ et soit $V_{s_i} = [x_i - Cs, x_i + Cs]$, $i = 1, \dots, q$ pour un niveau de résolution donné $s > 0$. Nous imposerons également que $s_{\alpha_i} < s_r$ ce qui sera supporté par l'hypothèse suivante :

Hypothèse 3.6 $c_n^2 \log(n) = o(n^{1-\frac{2\alpha_{\max}+1}{2r+1}})$.

Soit s_0 l'échelle la moins petite telle que $\kappa > 4Cs_0$. Alors, pour $s \leq s_0$ et tout $(x, y) \in V_{s_i} \times V_{s_{i+1}} ; i = 1, \dots, q$, les supports de $\psi_s(-x)$ et $\psi_s(-y)$ sont disjoints, ce qui implique que $W_{s,n}(Y)(x)$ et $W_{s,n}(Y)(y)$ sont des variables indépendantes. Notons que pour n suffisamment large, $s_r \leq s_0$. Ainsi, pour $s \leq s_r$ et n suffisamment large, les maxima d'ondelettes localisés dans V_{s_i} et les maxima d'ondelettes localisés dans $V_{s_{i'}}$ (pour $i \neq i'$) sont indépendants. Donc, nous pouvons directement utiliser les résultats de la section 3.4.1 (quand la fonction f a seulement une singularité d'ordre α) pour estimer les positions des maxima d'ondelettes dus aux q singularités d'ordre $\alpha_i ; i = 1, \dots, q$. En particulier, d'après les Propositions 3.9 et 3.10, nous pouvons dériver la proposition suivante :

Proposition 3.12 *Soit $V_{s_i} = [x_i - Cs, x_i + Cs]$, alors pour tout $s_r \geq s > \frac{1}{n}$:*

$$P\left(\max_{x \notin V_{s_i}, i=1, \dots, q} \{|W_{s,n}(Y)(x)|\} \leq \lambda\right) \rightarrow 1, \text{ quand } n \rightarrow \infty.$$

Soit k_{s_i} le nombre de maxima d'ondelettes localisés dans V_{s_i} qui sont tels qu'il existe une constante positive A_4 , telle que $|W_s(f)(m_{s,j,i})| \geq A_4 s^{\alpha_i + 1/2}$, $i = 1, \dots, q$, où les $m_{s,j,i}$ dénotent les positions des maxima d'ondelettes dans V_{s_i} à l'échelle $s > 0$ pour $j = 1, \dots, k_{s_i}$. Alors, pour $s_r \geq s \geq s_{\alpha_i^*}, i = 1, \dots, q$:

$$P(|W_{s,n}(Y)(m_{s,j,i})| \geq \lambda) \rightarrow 1, \text{ quand } n \rightarrow \infty, \text{ pour } j = 1, \dots, k_{s_i}.$$

En pratique, les ordres des singularités de la fonction f sont inconnus et les niveaux $s_{\alpha_i^*}$ doivent être estimés. Soient $\alpha_{\min} = \min\{\alpha_i ; i = 1, \dots, q\}$. Alors, de façon similaire au cas où f a seulement une singularité, la procédure d'estimation suivante peut être utilisée pour estimer les niveaux $s_{\alpha_{\min}}$ et les maxima d'ondelettes quand la fonction f a plusieurs singularités :

- définissons $\hat{s}_{\alpha_{\min}^*}$ comme le niveau de résolution le plus fin où $\max_{x \in [0,1]} |W_{s,n}(Y)(x)|$ dépasse le seuil λ
- pour $s_r \geq s \geq \hat{s}_{\alpha_{\min}^*}$, nous définissons les estimateurs des maxima d'ondelettes de f comme les positions des maxima de $|W_{s,n}(Y)(x)|$ qui dépassent le seuil λ à un niveau de résolution donné s .

3.5 Estimation des zero-crossings

Nous proposons d'utiliser les zero-crossings de la transformée en ondelettes d'un signal f pour localiser ses landmarks de type II (i.e. les points qui correspondent par exemple à la position des extrema ou des points d'inflexion au voisinage desquels une fonction est suffisamment régulière, voir section 3.1). Dans la section 3.2.2, nous avons expliqué que les zero-crossings de la transformée en ondelettes de f convergent vers les positions des zéros de la dérivée d'ordre r de f , où r est le nombre de moments nuls de l'ondelette ψ . Les positions des extrema de f peuvent donc être estimées en suivant la propagation le long des échelles des zéros de $W_s(f)(x)$ lorsque l'on utilise une ondelette avec $r = 1$ moment nul. Dans les sections suivantes, une

procédure de seuillage est proposée pour détecter les zero-crossings de la transformée en ondelettes d'un signal observé avec du bruit, et les propriétés de convergence des estimateurs sont étudiées.

Nous supposerons que nous travaillons avec une ondelette qui est causale, à support compact avec r moments nuls, et qu'une fonction inconnue f est observée selon un modèle de bruit blanc (3.3.13). La fonction f est supposée vérifier les hypothèses suivantes :

Hypothèse 3.7 f est $r+1$ fois continûment dérivable sur $[0, 1]$.

Hypothèse 3.8 $f^{(r)}$ a p zéros dans $]0, 1[$ aux points y_1, \dots, y_p qui sont tels que $f^{(r+1)}(y_i) \neq 0$, $i = 1, \dots, p$, où $p \geq 1$ est un entier positif.

Alors, d'après la Proposition 3.3, nous avons qu'il existe p lignes de zero-crossings $z_i(s), i = 1, \dots, p$ qui convergent respectivement vers les points y_1, \dots, y_p quand l'échelle $s \rightarrow 0$.

3.5.1 Détection des zero-crossing

L'approche que nous décrivons ici, pour détecter la présence d'un zéro dans $W_{s,n}(Y)(x)$, est basée sur les idées de Chaudhuri et Marron [20] qui ont développé un outil très simple appelé SiZer pour explorer les structures significatives d'un signal. Soit $H_0^{s,x}$ l'hypothèse nulle : $W_s(f)(x) = 0$ pour un $x \in [0, 1]$ fixé et une échelle donnée s . Supposons qu'un test statistique basé sur les observations $W_{s,n}(Y)(x)$ puisse être effectué pour cette hypothèse. Alors, si $H_0^{s,x}$ est rejetée, nous avons que $W_s(f)(x)$ est significativement positif ou négatif selon le signe de $W_{s,n}(Y)(x)$. Chaque zero-crossing significatif de $W_s(f)(x)$ à une échelle donnée s sera alors localisé entre une paire de points $x_1, x_2 \in [0, 1]$ tels que H_0^{s,x_1} et H_0^{s,x_2} seront rejetées, et $W_s(f)(x_1)$ et $W_s(f)(x_2)$ auront des signes opposés. Puisque $\{W_{s,n}(Y)(x) - W_s(f)(x), x \in [0, 1]\}$ est un processus Gaussien de moyenne nulle de variance σ^2/n , des tests asymptotiques pour la famille d'hypothèses $\{H_0^{s,x} | x \in [0, 1]\}$, pour une échelle fixé $s > 0$, peuvent facilement être construits. La proposition suivante, qui se focalise sur les performances de ce type de tests, est inspirée du Théorème 4.1 de Chaudhuri et Marron [20] :

Proposition 3.13 Soit $q_{1-\alpha}^n = \frac{1}{\sqrt{2 \log(n)}}(-\log(-\frac{1}{2} \log(1-\alpha)) + 2 \log(n) + \log(\frac{\int_{-\infty}^{+\infty} \{\psi'(u)\}^2}{2\pi}))$ pour $0 < \alpha < 1$. Considérons le test statistique qui accepte l'hypothèse nulle $H_0^{s,x}$: $W_s(f)(x) = 0$ si $|W_{s,n}(Y)(x)| \leq \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ et conclue que $W_s(f)(x)$ est significativement positif ou négatif si $W_{s,n}(Y)(x) > \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ ou $W_{s,n}(Y)(x) < -\frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ respectivement. Alors :

- Soit $s \geq \frac{1}{n}$ une échelle fixée. Alors, le niveau asymptotique simultané de ce test pour l'ensemble de la famille d'hypothèses $\{H_0^{s,x} | x \in [0, 1]\}$ est α . En d'autres termes, si nous supposons que $W_s(f)(x) = 0$ pour tout $x \in [0, 1]$, alors toutes

les hypothèses nulles $H_0^{s,x}$ seront acceptées par le test avec une distribution asymptotique d'au moins $(1 - \alpha)$ quand $n \rightarrow \infty$.

- Définissons s_* comme la solution de l'équation $s_*^{r+1/2} = \frac{\sigma}{\sqrt{n}}b(n)$ avec $b(n) = o((\log(n))^{-1/2})$. Alors, pour $s_* \geq s \geq \frac{1}{n}$, $P(\max_{x \in [0,1]} |W_{s,n}(Y)(x)| > \frac{\sigma}{\sqrt{n}}q_{1-\alpha}^n) \leq \alpha$ quand $n \rightarrow \infty$.
- Soit s une échelle fixé telle que $s > s_*$. Alors, si $W_s(f)(x)$ a p_s zero-crossings (i.e. p_s changements de signes) le test détectera la présence de ces p_s changements de signes avec probabilité tendant vers 1 quand $n \rightarrow \infty$.

Preuve. Notons que d'après le Lemme 3.1, $P(\max_{x \in [0,1]} |W_{n-1}(B)(x)| \leq q_{1-\alpha}^n) \rightarrow 1 - \alpha$ quand $n \rightarrow \infty$, où $W_{n-1}(B)(x) = \int_{-\infty}^{+\infty} \psi_{n-1}(u-x)B(du)$ et B est un mouvement Brownien standard. Supposons que pour une échelle fixée $s \geq \frac{1}{n}$ et tout $x \in [0, 1]$, $W_s(f)(x) = 0$. Notons que dans ce cas $P(|W_{s,n}(Y)(x)| \leq \frac{\sigma}{\sqrt{n}}q_{1-\alpha}^n) = P(|W_s(B)(x)| \leq q_{1-\alpha}^n)$. Rappelons que par la propriété d'auto-similarité du mouvement Brownien :

$$\text{Loi} \left(W_s(B)(x) = \int_{-\infty}^{+\infty} \psi_s(u-x)B(du) \right) = \text{Loi} \left(\int_{-\infty}^{+\infty} \psi(u - \frac{x}{s})B(du) \right). \quad (3.3.21)$$

Donc, puisque $\frac{1}{s} \leq n$,

$$\begin{aligned} P(|W_s(B)(x)| \leq q_{1-\alpha}^n) &\geq P(\max_{x \in [0,1]} |W_s(B)(x)| \leq q_{1-\alpha}^n) \\ &\geq P(\max_{x \in [0,1]} |W_{n-1}(B)(x)| \leq q_{1-\alpha}^n) \rightarrow 1 - \alpha, \text{ quand } n \rightarrow \infty, \end{aligned}$$

ce qui prouve la première assertion de la proposition.

Soit $s_* \geq s \geq \frac{1}{n}$, alors :

$$P(\max_{x \in [0,1]} |W_{s,n}(Y)(x)| \geq \frac{\sigma}{\sqrt{n}}q_{1-\alpha}^n) \leq P(\max_{x \in [0,1]} \frac{\sigma}{\sqrt{n}}|W_s(B)(x)| \geq \frac{\sigma}{\sqrt{n}}q_{1-\alpha}^n - |W_s(f)(x)|).$$

Puisque f est r fois continûment dérivable sur $[0, 1]$, le Théorème 3.1 implique qu'il existe une constante A_1 telle que pour tout $x \in [0, 1]$ et tout $s > 0$, $|W_s(f)(x)| \leq A_1 s^{r+1/2}$. Ainsi, pour $s_* \geq s \geq \frac{1}{n}$, nous obtenons :

$$\begin{aligned} P(\max_{x \in [0,1]} |W_{s,n}(Y)(x)| \geq \frac{\sigma}{\sqrt{n}}q_{1-\alpha}^n) &\leq P(\max_{x \in [0,1]} |W_s(B)(x)| \geq q_{1-\alpha}^n - A_1 b(n)) \\ &\leq P(\max_{x \in [0,1]} |W_{n-1}(B)(x)| \geq q_{1-\alpha}^n - A_1 b(n)). \end{aligned}$$

D'après le Lemme 3.1, nous avons que pour n suffisamment large :

$$\begin{aligned} P(\max_{x \in [0,1]} |W_{n-1}(B)(x)| \geq q_{1-\alpha}^n - A_1 b(n)) &= 1 - \exp\{-2 \exp(\log(-\frac{1}{2} \log(1 - \alpha))) \\ &\quad \exp(A_1 \sqrt{2 \log(n)} b(n))\}. \end{aligned}$$

Puisque $b(n) = o((\log(n))^{-1/2})$, $\exp(A_1 \sqrt{2 \log(n)} b(n)) \rightarrow 1$, et donc $1 - \exp(-2 \exp(\log(-\frac{1}{2} \log(1 - \alpha))) \exp(A_1 \sqrt{2 \log(n)} b(n))) \rightarrow \alpha$ quand $n \rightarrow \infty$, ce qui implique que pour $s_* \geq s \geq \frac{1}{n}$:

$$P(\max_{x \in [0,1]} |W_{s,n}(Y)(x)| \geq \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n) \leq \alpha, \text{ quand } n \rightarrow \infty.$$

Prouvons maintenant la troisième partie de la proposition. Soit $s > s_*$ une échelle fixée. Supposons que $W_s(f)(x)$ a p_s zero-crossings aux points $0 < z_{s,1} < \dots < z_{s,p_s} < 1$. Alors, il existe $p_s + 1$ points $0 < x_{s,1} < \dots < x_{s,p_s+1} < 1$ tels que $z_{s,i} \in [x_{s,i}, x_{s,i+1}]$ et $\operatorname{sgn}(W_s(f)(x_{s,i})) = -\operatorname{sgn}(W_s(f)(x_{s,i+1})) \neq 0$ pour $i = 1, \dots, p_s$. Puisque $|W_{s,n}(Y)(x) - W_s(f)(x)| \leq K \frac{\sqrt{|\log(s)|}}{\sqrt{n}}$ p.s. (voir Lemme 3.4), on obtient que $\operatorname{sgn}(W_{s,n}(Y)(x_{s,i})) = \operatorname{sgn}(W_s(f)(x_{s,i}))$ p.s. pour n suffisamment large. Notons que $\frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n \rightarrow 0$ quand $n \rightarrow 0$, ce qui implique pour n suffisamment large, $|W_{s,n}(Y)(x_{s,i})| > \frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ p.s. Ainsi, pour n suffisamment large, les hypothèses $H_0^{s,x_{s,i}}$ sont rejetées avec :

$$\operatorname{sgn}(W_{s,n}(Y)(x_{s,i})) = -\operatorname{sgn}(W_{s,n}(Y)(x_{s,i+1})) \text{ p.s.},$$

ce qui termine la preuve. \square

Soit s_{min} l'échelle la plus grossière à partir de laquelle on souhaite estimer les zero-crossings de $W_s(f)(x)$. Une simple procédure de seuillage pour estimer les zero-crossings de $W_s(f)(x)$ peut maintenant être dérivée de la Proposition 3.13. Pour $s_{min} \geq s > s_*$, soit $\{(\hat{x}_{s,i_1}, \hat{x}_{s,i_2}); i_1, i_2 \in \mathbb{N}\}$ la suite de points de $]0, 1[\times]0, 1[$ tels que :

- $H_0^{s,\hat{x}_{s,i_1}}$ et $H_0^{s,\hat{x}_{s,i_2}}$ sont rejetées.
- $W_{s,n}(Y)(\hat{x}_{s,i_1})$ et $W_{s,n}(Y)(\hat{x}_{s,i_2})$ ont des signes opposés.
- toutes les hypothèses $H_0^{s,x}$ pour $\hat{x}_{s,i_1} < x < \hat{x}_{s,i_2}$ sont acceptées.

Définissons $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ comme les zero-crossings de $W_{s,n}(Y)(x)$ qui sont positionnés dans les intervalles $\] \hat{x}_{s,i_1}, \hat{x}_{s,i_2} [$. Dans les sections suivantes, la convergence en probabilité et les distributions asymptotiques de ces estimateurs sont étudiées. Cependant, le niveau s_* est généralement inconnu et une procédure pour estimer les zero-crossings en pratique est donc suggérée dans la section 3.7.1.

Remarque 3.4 Comme mentionné dans Chaudhuri et Marron [20], il serait raisonnable de considérer différentes valeurs pour le seuil $q_{1-\alpha}^n$ adaptées à l'échelle s et d'effectuer des tests séparés pour différents niveaux de lissage. Dans ce cas, l'on doit utiliser les quantiles d'ordre $(1 - \alpha)$ de $\max_{x \in [0,1]} |W_s(B)(x)|$ pour différents $s > 0$. Nous pensons que les nombreux tests statistiques suggérés par Chaudhuri et Marron [19] pour détecter les zero-crossings significatifs de la dérivée de leur surface temps-échelle, pourraient être adaptés à notre problème.

3.5.2 Convergence en probabilité

Pour prouver la convergence en probabilité des estimateurs $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ pour $s_{\min} \geq s > s_*$, nous allons montrer que les positions des zéros du processus stochastique $W_{s,n}(Y)(x)$ convergent p.s. vers les positions des zéros de la fonction $W_s(f)(x)$ quand $n \rightarrow \infty$ et l'échelle s est fixée. Alors, la convergence en probabilité des estimateurs $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ se déduit de la Proposition 3.13.

Pour prouver ce résultat, nous allons utiliser quelques propriétés des réalisations du processus stochastique $W_{s,n}(Y)(x)$. Soit $C^1([0, 1])$ l'ensemble de toutes les fonctions continûment dérivables de $[0, 1]$ dans \mathbb{R} . Dans ce qui suit, $C^1([0, 1])$ sera muni de la norme $\|g\|_{\infty,1} = \sum_{p=0}^1 \sup_{t \in [0,1]} |g^{(p)}(t)|$ pour $g \in C^1([0, 1])$. Rappelons que nous travaillons avec un espace probabilisé (Ω, A, P) et que les réalisations du processus B sont des fonctions continues sur \mathbb{R} .

Lemme 3.3 *Supposons que l'ondelette $\psi = (-1)^r \theta^{(r)}$ est à décroissance rapide. Soit s une échelle fixée et n un entier positif fixé. Alors, pour presque tout $\omega \in \Omega$, $x \mapsto W_{s,n}(Y)(w, x) = \int_{-\infty}^{+\infty} \psi_s(u - x) Y(du) = W_s(f)(x) + \frac{\sigma}{\sqrt{n}} W_s(B)(w, x)$ appartient à $C^1([0, 1])$.*

Preuve. Dans la preuve de la Proposition 3.5, nous avons montré que si ψ est à décroissance rapide, alors il existe un sous-ensemble $\Omega_0 \subset \Omega$ tel que $P(\Omega_0) = 1$ et tel que pour tout $\omega \in \Omega_0$:

$$W_s(B)(\omega, x) = -\frac{1}{s} \tilde{\psi}'_s \star B(\omega, .)(x).$$

où $\tilde{\psi}'_s(u) = s^{-1/2} \psi'(-\frac{u}{s})$ et \star dénote le produit de convolution. Notons que d'après la Proposition 3.5, nous avons également que pour tout $\omega \in \Omega_0$, $B(., w)$ est une distribution tempérée. Etant donné que la fonction ψ'_s est à décroissance rapide, la propriété de régularisation du produit de convolution implique que, pour tout $\omega \in \Omega_0$, $W_s(B)(w, .)$ est C^∞ sur $[0, 1]$. De manière similaire, puisque $f \in L^2(\mathbb{R})$ est une distribution tempérée (avec $f(x) = 0$ pour $x \notin [0, 1]$), $W_s(f)(x) = f \star \tilde{\psi}'_s(x)$ est également C^∞ sur $[0, 1]$, ce qui termine la preuve. \square

Nous allons maintenant utiliser le Lemme suivant pour prouver que le processus $W_{s,n}(Y)(x)$ converge p.s. vers $W_s(f)(x)$ (où la norme réfère à $\|\cdot\|_{\infty,1}$) quand $n \rightarrow \infty$.

Lemme 3.4 *Supposons que l'ondelette $\psi = (-1)^r \theta^{(r)}$ soit à support compact, alors il existe une constante K telle que pour tout $x \in [0, 1]$ et les petites échelles s :*

$$|W_{s,n}(Y)(x) - W_s(f)(x)| \leq K \frac{\sqrt{|\log(s)|}}{\sqrt{n}}, \text{ p.s.} \quad (3.3.22)$$

Preuve. Ce résultat est similaire au Lemme 1 de Wang [89]. Un résultat prouvé par P.Levy (Modulus of Continuity, voir Théorème 2.7, p.31 de Revuz et Yor [79]) montre que pour un mouvement Brownien standard B_t sur $[0, 1]$:

$$P \left(\limsup_{\delta \rightarrow 0} (2\delta \log(1/\delta))^{-1/2} \max_{0 \leq v < t \leq 1; t-v \leq \delta} |B_t - B_v| = 1 \right) = 1.$$

Notons que :

$$|W_s(B)(x)| = s^{-1/2} \left| \int_{-\infty}^{+\infty} \psi(u) dB(x + su) du \right|.$$

et rappelons que dans la preuve du Théorème 3.5, nous avons montré que

$$|W_s(B)(x)| = s^{-1/2} \left| \int_{-\infty}^{+\infty} \psi'(u) B(x + su) du \right|$$

Puisque $\psi'(u) = (-1)^r \theta^{(r+1)}$ est une ondelette admissible, $\int_{-\infty}^{+\infty} \psi'(u) du = 0$ ce qui implique que :

$$|W_s(B)(x)| = s^{-1/2} \left| \int_{-\infty}^{+\infty} \psi'(u) (B(x + su) - B(x)) du \right|, \text{ pour } x \in [0, 1].$$

Soit $x \in]0, 1[$. Rappelons que l'ondelette ψ' est à support compact égal à $[-C, C]$. Donc, il existe une échelle s_0 telle que pour tout $s \leq s_0$ et tout $u \in [-C, C]$, $0 \leq x + su \leq 1$. Puisque $|x + su - x| \leq Cs$, la propriété de modulus continuity du mouvement Brownien prouve qu'il existe une constante K telle que pour tout $x \in [0, 1]$, pour tout s suffisamment petit et presque tout $\omega \in \Omega$:

$$\begin{aligned} |W_s(B)(w, x)| &\leq K s^{-1/2} \left| \int_{-\infty}^{+\infty} \psi'(u) (2su |\log(su)|)^{1/2} du \right| \\ &\leq 2K |\log(s)|^{1/2} \int_{-\infty}^{+\infty} |\psi'(u)| |u|^{1/2} du + K \int_{-\infty}^{+\infty} |\psi'(u)| |u \log(u)|^{1/2} du, \end{aligned}$$

Alors, le résultat suit immédiatement. \square

D'après le Lemme 3.3, nous avons que les réalisations du processus aléatoire $W_{s,n}(Y)(x)$ sont presque sûrement dérivables. Notons que :

$$W'_{s,n}(Y)(x) = -\frac{1}{s} \int_{-\infty}^{+\infty} \psi'_s(u - x) f(u) du - \frac{1}{s} \frac{\sigma}{\sqrt{n}} \int_{-\infty}^{+\infty} \psi'_s(u - x) B(du),$$

où $\psi'_s(u) = \frac{1}{\sqrt{s}} \psi'(\frac{u}{s})$. Puisque $\psi'(u) = (-1)^r \frac{d^{r+1}}{du^{r+1}} \theta(u)$ est également une ondelette dérivable à décroissance rapide, nous avons que $W'_{s,n}(Y)(x)$ est un processus Gaussien avec :

$$E(W'_{s,n}(Y)(x)) = W'_s(f)(x), \quad (3.3.23)$$

$$Var(W'_{s,n}(Y)(x)) = \frac{\sigma^2 \int_{-\infty}^{+\infty} \left\{ \frac{d^{r+1}}{du^{r+1}} \theta(u) \right\}^2 du}{ns^2}, \quad (3.3.24)$$

D'après le Lemme 3.4, nous obtenons également les vitesses de convergence du processus $W'_{s,n}(Y)(x)$ (notons que l'inégalité suivante est vraie uniformément pour x) :

$$|W'_{s,n}(Y)(x) - W'_s(f)(x)| \leq O\left(\frac{1}{s} \frac{\sqrt{|\log(s)|}}{\sqrt{n}}\right), \text{ p.s..} \quad (3.3.25)$$

D'après le Lemme 3.4 et la relation (3.3.25), nous avons que pour s suffisamment petit et presque tout $\omega \in \Omega$, $\|W_{s,n}(Y)(\omega, \cdot) - W_s(f)\|_{\infty,1} \rightarrow 0$ quand $n \rightarrow \infty$.

Pour étudier la convergence des positions des zéros du processus aléatoire $W_{s,n}(Y)(x)$, nous allons utiliser le concept de fonctionnelles structurelles développé par Kneip et Gasser [51]. Une fonctionnelle structurelle T est une application qui fait correspondre des éléments de l'espace des fonctions de régression avec l'axe temporel. Dans cette section, nous allons considérer des fonctionnelles structurelles qui identifient les positions des zéros d'une fonction $g \in C^1([0, 1])$. Les fonctionnelles T doivent respecter les principes suivants : tout d'abord, si l'identification unique n'est pas possible pour une fonction g , l'on doit avoir $T(g) = \mu$ pour une valeur pré-spécifiée $\mu \notin [0, 1]$. Deuxièmement, une fonctionnelle structurelle T doit être continue dans le sens où si deux fonctions f et g sont proches l'une de l'autre, alors $T(f)$ et $T(g)$ doivent également être proches. Dans la discussion qui suit, nous focaliserons sur des fonctionnelles T qui vérifient la définition suivante :

Définition 3.2 Soient μ_1 et μ_2 deux réels pré-spécifiés qui ne sont pas dans $]0, 1[$. Soit $[a, b]$ un sous-intervalle de $]0, 1[$. Une fonctionnelle $T : C^1([0, 1]) \rightarrow]0, 1[\cup \{\mu_1, \mu_2\}$ sera appelée une fonctionnelle structurelle de première espèce associée à $[a, b]$, si pour $g \in C^1([0, 1])$:

$$\begin{aligned} T(g) &= \zeta, \text{ si } g \text{ a exactement un zéro }]a, b[\text{ au point } \zeta \text{ tel que } g'(\zeta) \neq 0, \\ T(g) &= \mu_1, \text{ si } g \text{ n'a pas de zéro }]a, b[, \\ T(g) &= \mu_2, \text{ si } g \text{ a soit plusieurs zéros dans }]a, b[\text{ ou exactement un zéro} \\ &\quad \text{au point } \zeta \in]a, b[\text{ qui est dégénéré (i.e } g'(\zeta) = 0) . \end{aligned}$$

Alors, la proposition suivante peut être obtenue :

Proposition 3.14 Soit T une fonctionnelle structurelle de première espèce associée à $[a, b]$ avec $0 < a < b < 1$. Soit $f \in C^1([0, 1])$. Si f a soit aucun zéro ou exactement un zéro dans $]a, b[$ au point ζ avec $f'(\zeta) \neq 0$, alors T est continue en f .

Preuve : Supposons tout d'abord que $T(f) = \zeta$ pour un $\zeta \in]a, b[$, i.e. f a exactement un zéro localisé dans $]a, b[$ avec $f'(\zeta) \neq 0$. Soit $\epsilon > 0$. Puisque f et f' sont continues et si ϵ est suffisamment petit, il existe un $\gamma > 0$ tel que $|f(\zeta - \epsilon)| \geq \gamma$, $|f(\zeta + \epsilon)| \geq \gamma$, $\text{sgn}(f(\zeta - \epsilon)) = -\text{sgn}(f(\zeta + \epsilon))$ et $|f'(t)| \geq \gamma$ pour tout $t \in [\zeta - \epsilon, \zeta + \epsilon]$. Soit $U_\epsilon = \{g \in C^1([0, 1]); \|g - f\|_{\infty,1} < \gamma\}$. Alors, pour $g \in U_\epsilon$, $\sup_{t \in [0,1]} \{|f(t) - g(t)| < \gamma\}$ implique que $\text{sgn}(g(\zeta - \epsilon)) = -\text{sgn}(g(\zeta + \epsilon)) \neq 0$. Ainsi, par la continuité de g , il existe un $z_g \in [\zeta - \epsilon, \zeta + \epsilon]$ tel que $g(z_g) = 0$. Puisque $\sup_{t \in [0,1]} |f'(t) - g'(t)| < \gamma$, $\text{sgn}(|g'(t)|) > 0$ pour tout $t \in [\zeta - \epsilon, \zeta + \epsilon]$ ce qui montre l'unicité de z_g . Finalement, nous avons que pour tout $g \in U_\epsilon$, il existe exactement un $z_g \in [\zeta - \epsilon, \zeta + \epsilon]$ tel que $g(z_g) = 0$ et $g'(z_g) \neq 0$. Soit g_n une suite de fonctions de $C^1([0, 1])$ qui converge vers f . Puisque $|f(t)| > 0$ pour $t \in [a, T(f) - \epsilon \cup T(f) + \epsilon, b]$, nous avons que pour n suffisamment large, $|g_n(t)| > 0$ pour tout $t \in [a, T(f) - \epsilon \cup T(f) + \epsilon, b]$. Donc, pour n suffisamment large, g_n n'a pas de zéro $[a, T(f) - \epsilon \cup T(f) + \epsilon, b]$ et $g_n \in U_\epsilon$, ce qui implique que

$T(g_n) \in [T(f) - \epsilon, T(f) + \epsilon]$. Puisque ϵ peut être arbitrairement petit, nous avons $|T(g_n) - T(f)| \rightarrow 0$ quand $n \rightarrow \infty$.

Maintenant, supposons que $T(f) = \mu_1$ i.e. la fonction f n'a pas de zéro dans $]a, b[$. Soit g_n une suite fonctions de $C^1([0, 1])$ qui converge vers f . Étant donné que f est continue sur $]a, b[$, il existe une constante $\gamma > 0$ telle que $f(t) \geq \gamma$ ou $f(t) \leq -\gamma$ pour tout $t \in]a, b[$. Puisque $\|g_n - f\|_{\infty, 1} \rightarrow 0$, nous avons que $\sup_{t \in]a, b[} |g_n(t) - f(t)| \rightarrow 0$ et donc que pour n suffisamment large n , $g_n(t) > 0$ ou $g_n(t) < 0$ pour tout $t \in]a, b[$. Ainsi, quand $n \rightarrow \infty$, g_n n'a pas de zéros dans $]a, b[$, et donc $T(g_n) = \mu_1$ ce qui montre que $|T(g_n) - T(f)| \rightarrow 0$. \square

La convergence en probabilité des estimateurs $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ peut maintenant être facilement dérivée des Propositions 3.13 et 3.14 :

Théorème 3.6 *Soit $s_{min} \geq s > s_*$ et $z_{s,1} < \dots < z_{s,p_s}$ les zero-crossings de $W_s(f)$ qui sont localisés dans $]0, 1[$. Supposons que $\frac{d}{dx} W_s(f)(z_{s,i}) \neq 0$ pour $i = 1, \dots, p_s$ i.e les zero-crossings $z_{s,i}$ sont non-dégénérés. Alors, pour s suffisamment petit, $\hat{p}_s \rightarrow_p p_s$ et $\hat{z}_{s,i} \rightarrow_p z_{s,i}$ pour $i = 1, \dots, p_s$ quand $n \rightarrow \infty$.*

Preuve. Puisque $W'_s(f)(z_{s,i}) \neq 0$, les zéros de $W_s(f)(x)$ sont distincts. Ainsi, il existe des réels $0 < a_{s,1} < b_{s,1} < \dots < a_{s,p_s} < b_{s,p_s} < 1$ tels que $z_{s,i} \in [a_{s,i}, b_{s,i}]$ et $sgn(W_s(f)(a_{s,i})) = -sgn(W_s(f)(b_{s,i})) \neq 0$ pour $i = 1, \dots, p_s$. Soit $T_i, i = 1, \dots, p_s$ des fonctionnelles structurelles de première espèce respectivement associées à $[a_{s,i}, b_{s,i}]$, et $\tilde{T}_i, i = 0, \dots, p_s$ des fonctionnelles structurelles de première espèce respectivement associées à $[b_{s,i}, a_{s,i+1}]$ (avec $b_{s,0} = 0$ et $a_{s,p_s+1} = 1$). Notons que $T_i(W_s(f)) = z_{s,i}$ et $\tilde{T}_i(W_s(f)) = \mu_1$. Par le Lemme 3.3, $W_s(f) \in C^1([0, 1])$, et donc, T_i et \tilde{T}_i sont continues en $W_s(f)$.

Rappelons que pour s suffisamment petit, l'on a $\|W_{s,n}(Y) - W_s(f)\|_{\infty, 1} \rightarrow 0$ p.s. Ainsi, il suit de la Proposition 3.14, que $T_i(W_{s,n}(Y)) \rightarrow z_{s,i}$ p.s. et que $\tilde{T}_i(W_{s,n}(Y)) \rightarrow \mu_1$ p.s. quand $n \rightarrow \infty$. Puisque $sgn(W_s(f)(a_{s,i})) = -sgn(W_s(f)(b_{s,i})) \neq 0$, la Proposition 3.13 implique que, pour une échelle fixée $s_{min} \geq s > s_*$, les estimateurs $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ sont localisés dans les intervalles $[a_{s,i}, b_{s,i}]$ avec probabilité tendant vers 1 quand $n \rightarrow \infty$. Ainsi, $\hat{p}_s \rightarrow_p p_s$ et pour $i = 1, \dots, p_s$, $P(\hat{z}_{s,i} \in [a_{s,i}, b_{s,i}]) \rightarrow 1$ quand $n \rightarrow \infty$. Ceci complète la preuve, puisque nous avons montré que les zero-crossings du processus $W_{s,n}(Y)$ localisés dans $[a_{s,i}, b_{s,i}]$ convergent p.s vers les positions des zero-crossings de $W_s(f)$. \square

3.5.3 Distribution asymptotique des estimateurs

Nous allons maintenant étudier la distribution asymptotique des estimateurs $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$. Dans ce but, nous allons combiner la méthodologie utilisée par Müller [70] pour étudier les propriétés asymptotiques des estimateurs à noyau des positions des zéros et des extrema locaux de fonctions de régression, avec un théorème des valeurs intermédiaires pour les processus stochastiques (voir Jennrich [49]).

Théorème 3.7 Soit $s_{min} \geq s > s_*$ et $z_{s,1} < \dots < z_{s,p_s}$ les zero-crossings de $W_s(f)$ qui sont localisés dans $]0, 1[$, et supposons les conditions du Théorème 3.6. Alors, quand $n \rightarrow \infty$:

$$\sqrt{n}(\hat{z}_{s,i} - z_{s,i}) \rightarrow N\left(0, \frac{\sigma^2}{(\frac{d}{dx}W_s(f)(z_{s,i}))^2}\right), \quad i = 1, \dots, p_s.$$

Preuve. Pour prouver ce théorème, nous aurons besoin d'utiliser le Lemme suivant (voir Jennrich [49]) qui conduit à un théorème des valeurs intermédiaires pour les processus stochastiques :

Lemme 3.5 Soit f une fonction de $Y \times \Theta$ dans \mathbb{R} où Y est un espace mesurable, et Θ est un sous-ensemble de \mathbb{R}^n convexe et compact, dont l'intérieur est non-vide. Supposons que pour tout $\theta \in \Theta$, $f(., \theta)$ est mesurable et que pour tout $y \in Y$, $f(y, .)$ est continûment dérivable. Soit θ_1 et θ_2 deux fonctions mesurables de Y dans Θ . Alors, il existe une fonction mesurable θ^* de Y dans Θ telle que pour tout $y \in Y$:

$$f(y, \theta_1(y)) - f(y, \theta_2(y)) = \frac{df}{d\theta}(y, \theta^*(y))(\theta_1(y) - \theta_2(y)),$$

et $\theta^*(y)$ se trouve entre $\theta_1(y)$ et $\theta_2(y)$.

Maintenant, prouvons le résultat pour $z_s = z_{s,1}$. D'après le Lemme 3.3, nous avons qu'il existe un sous-ensemble $\Omega_0 \subset \Omega$ tel que $P(\Omega_0) = 1$ et pour tout $\omega \in \Omega_0$, $x \mapsto W_{s,n}(Y)(w, x)$ est continûment dérivable. Ainsi, nous pouvons utiliser le Lemme 3.5 avec $Y = \Omega_0$, $\Theta = [0, 1]$ et $f(w, .) = W_{s,n}(Y)(w, .)$, pour montrer qu'il existe une variable aléatoire ζ_s^* se trouvant entre \hat{z}_s et z_s telle que :

$$W_{s,n}(Y)(\hat{z}_s) = W_{s,n}(Y)(z_s) + (\hat{z}_s - z_s)W'_{s,n}(Y)(\zeta_s^*) \text{ p.s.},$$

Notons que :

$$|W'_{s,n}(Y)(\zeta_s^*) - W'_s(f)(z_s)| \leq |W'_{s,n}(Y)(\zeta_s^*) - W'_s(f)(\zeta_s^*)| + |W'_s(f)(\zeta_s^*) - W'_s(f)(z_s)|.$$

Par le Lemme 3.4, $|W'_{s,n}(Y)(\zeta_s^*) - W'_s(f)(\zeta_s^*)| \rightarrow 0$ p.s., et par la continuité de $W'_s(f)$, $|W'_s(f)(\zeta_s^*) - W'_s(f)(z_s)| \rightarrow_p 0$ (notons que $\zeta_s^* \rightarrow_p z_s$). Ainsi, $(W'_s(f)(z_s) - W'_{s,n}(Y)(\zeta_s^*)) \rightarrow_p 0$, et donc, pour n suffisamment large, $W'_{s,n}(Y)(\zeta_s^*) \neq 0$ (en probabilité). De plus, puisque $W_{s,n}(Y)(\hat{z}_s) = 0 = W_s(f)(z_s)$ p.s., nous obtenons :

$$\hat{z}_s - z_s = \frac{W_s(f)(z_s) - W_{s,n}(Y)(z_s)}{W'_{s,n}(Y)(\zeta_s^*)}.$$

Puisque $W_s(f)(z_s) - W_{s,n}(Y)(z_s) \sim N(0, \frac{\sigma^2}{n})$, l'assertion du théorème suit. \square

Par conservation du produit scalaire, nous avons :

$$E(W_{s_1}(B)(x)W_{s_2}(B)(y)) = \int_{-\infty}^{+\infty} \psi_{s_1}(u-x)\psi_{s_2}(u-y)du. \quad (3.3.26)$$

Les corrélations entre les estimateurs $\hat{z}_{s,i}$, $s_{min} \geq s > s_*$, $i = 1, \dots, p_s$ sont alors données par le théorème suivant :

Théorème 3.8 Soit $s_{min} \geq s_1 \geq s_2 > s_*$. Supposons que $W'_s(f)(z_{s_1,i_1}) \neq 0$ et $W'_s(f)(z_{s_2,i_2}) \neq 0$, pour $1 \leq i_1 \leq p_{s_1}$ et $1 \leq i_2 \leq p_{s_2}$. Supposons les conditions du Théorème 3.6. Alors, quand $n \rightarrow \infty$:

$$E(\sqrt{n}(\hat{z}_{s_1,i_1} - z_{s_1,i_1})\sqrt{n}(\hat{z}_{s_2,i_2} - z_{s_2,i_2})) \rightarrow \sigma^2 \frac{\int_{-\infty}^{+\infty} \psi_{s_1}(u-z_{s_1,i_1})\psi_{s_2}(u-z_{s_2,i_2})du}{W'_{s_1}(f)(z_{s_1,i_1})W'_{s_2}(f)(z_{s_2,i_2})}.$$

Preuve. De façon similaire à la preuve du Théorème 3.7, nous pouvons montrer qu'il existe deux valeurs moyennes ζ_{s_1,i_1}^* et ζ_{s_2,i_2}^* respectivement situées entre \hat{z}_{s_1,i_1} et z_{s_1,i_1} , et entre \hat{z}_{s_2,i_2} et z_{s_2,i_2} telles que pour $k = 1, 2$:

$$\hat{z}_{s_k,i_k} - z_{s_k,i_k} = \frac{W_{s_k}(f)(z_{s_k,i_k}) - W_{s_k,n}(Y)(z_{s_k,i_k})}{W'_{s_k,n}(Y)(\zeta_{s_k,i_k}^*)}.$$

Alors, comme dans la preuve du Théorème 3.7, nous pouvons montrer que $(W'_{s_k,n}(Y)(\zeta_{s_k,i_k}^*) - W'_{s_k}(f)(z_{s_k,i_k})) \rightarrow_p 0$ et donc que $W'_{s_k,n}(Y)(\zeta_{s_k,i_k}^*) \neq 0$ (en probabilité) pour n suffisamment grande. Puisque $W_{s_k}(f)(z_{s_k,i_k}) - W_{s_k,n}(Y)(x)$, $k = 1, 2$ sont deux processus Gaussiens dont les covariances sont données par la relation (3.3.26), nous avons finalement le résultat. \square

3.6 Estimation des landmarks

Dans les sections précédentes, nous avons donné une procédure pour estimer les lignes de maxima d'ondelettes et de zero-crossings à différentes échelles d'une fonction observée avec du bruit. Nous proposons maintenant de calculer les intensités structurelles correspondantes (voir Proposition 3.4) pour déterminer les landmarks de la fonction inconnue f . La proposition suivante prouve que les maxima locaux de l'intensité structurelle des lignes de maxima d'ondelettes convergent en probabilité vers les singularités de f :

Proposition 3.15 Soit f une fonction inconnue observée selon un modèle de bruit blanc (3.3.13). Supposons que $\psi(u) = (-1)^r \frac{d^r \theta(u)}{dt^r}$ est une ondelette causale avec r moments nuls dont le support est égal à $[-C, C]$ et supposons que :

- il existe q lignes de maxima d'ondelettes $m_i(s)$ qui convergent respectivement vers q points $x_i \in]0, 1[$ dont la régularité Lipchitzienne est $0 \leq \alpha_i < r$ quand $s \rightarrow 0$.
- f est r fois continûment dérivable pour tout $x \in [0, 1] \setminus \{x_1, \dots, x_q\}$.

- pour $s_{\alpha_i^*} \leq s \leq s_r$, q lignes de maxima d'ondelettes $\hat{m}_i(s)$ ont été estimées et sont telles que $P(\hat{m}_i(s) \in V_{s_i}) \rightarrow 1$ quand $n \rightarrow \infty$ pour $i = 1, \dots, q$, où $V_{s_i} = [x_i - Cs, x_i + Cs]$ et $s_{\alpha_i^*}$ et s_r sont les échelles définies dans la section 3.4.2. De plus, nous supposerons que pour $i = 1, \dots, q$, les lignes $\hat{m}_i(\cdot)$ sont des fonctions p.s. continues de l'échelle s .

Définissons :

$$\hat{G}_m(x) = \sum_{i=1}^q \int_{s_{\alpha_i^*}}^{s_r} \frac{1}{s} \theta\left(\frac{x - \hat{m}_i(s)}{s}\right) ds.$$

Soit $\hat{x}_1, \dots, \hat{x}_{\hat{q}}$ les maxima locaux de $\hat{G}_m(x)$, alors quand $n \rightarrow \infty$:

$$\hat{q} \rightarrow_p q \text{ et } \hat{x}_i \rightarrow_p x_i \text{ pour } i = 1, \dots, q.$$

Preuve. D'après nos hypothèses nous avons qu'il existe un sous-ensemble $\Omega_0 \subset \Omega$ tel que $P(\Omega_0) = 1$ et tel que pour tout $\omega \in \Omega_0$, $\hat{m}_i(\omega, s)$ est une fonction continue de l'échelle s . Alors, pour tout $\omega \in \Omega_0$, définissons :

$$\hat{g}_i(\omega, x) = \int_{s_{\alpha_i^*}}^{s_r} \frac{1}{s} \theta\left(\frac{x - \hat{m}_i(\omega, s)}{s}\right) ds.$$

Par le théorème de dérivation sous le signe intégral, nous avons que $x \mapsto \hat{g}_i(\omega, x)$ est continûment dérivable sur $[0, 1]$, puisque θ est par définition continûment dérivable sur $[0, 1]$ et puisque $s_{\alpha_i^*} > 0$. Définissons $\hat{x}_{i,\min}(\omega) = \inf\{\hat{m}_i(\omega, s) - Cs; s \in [s_{\alpha_i^*}, s_r]\}$ et $\hat{x}_{i,\max}(\omega) = \sup\{\hat{m}_i(\omega, s) + Cs; s \in [s_{\alpha_i^*}, s_r]\}$. Notons que si s_r est suffisamment petit, alors pour $x \leq \hat{x}_{i,\min}(\omega)$ ou $x \geq \hat{x}_{i,\max}(\omega)$, $\hat{g}_i(\omega, x) = 0$. Ainsi, étant donné que $\hat{g}_i(\omega, x) \geq 0$, le théorème des valeurs intermédiaires pour les processus stochastiques (voir Lemme 3.5) implique qu'il existe une variable aléatoire \hat{x}_i se trouvant entre $\hat{x}_{i,\min}$ et $\hat{x}_{i,\max}$ telle que $\hat{g}_i(\omega, \hat{x}_i(\omega))$ est un maxima local de $\hat{g}_i(\omega, \cdot)$. D'après nos hypothèses, nous avons que pour tout $s \in [s_{\alpha_i^*}, s_r]$ $P(\hat{m}_i(s) \in V_{s_i}) \rightarrow 1$. Puisque $s_{\alpha_i^*}$ et s_r tendent vers zéro quand $n \rightarrow \infty$, nous obtenons que $\hat{m}_i(s) \rightarrow_p x_i$ quand $n \rightarrow \infty$, ce qui implique que $\hat{x}_{i,\min} \rightarrow_p x_i$ et $\hat{x}_{i,\max} \rightarrow_p x_i$ et finalement que $\hat{x}_i \rightarrow_p x_i$ quand $n \rightarrow \infty$. Maintenant, notons que $\hat{G}_m(\omega, x) = \sum_{i=1}^q \hat{g}_i(\omega, x)$. Puisque s_{α_i} et s_r tendent vers zéro quand $n \rightarrow \infty$, les supports de $\hat{g}_i(\omega, x)$ et $\hat{g}_j(\omega, x)$ ($i \neq j$) seront disjoints pour n suffisamment large ce qui prouve finalement le résultat. \square

La proposition suivante montre que l'intensité structurelle des zero-crossings estimés définis dans la section précédente converge en probabilité vers l'intensité structurelle des vraies lignes de zero-crossings du signal f . Ainsi, si nous définissons les landmarks d'une fonction comme les maxima locaux de l'intensité structurelle des lignes de zero-crossings et les points où celle-ci tend vers l'infini, alors la Proposition 3.16 montre que les maxima locaux de $\hat{G}_z(x)$ convergent en probabilité vers les landmarks du signal inconnu.

Proposition 3.16 Soit f une fonction inconnue observée selon un modèle de bruit blanc (3.3.13). Supposons que $\psi(u) = (-1)^r \frac{d^r \theta(u)}{dt^r}$ est une ondelette causale avec r moments nuls dont le support est égal à $[-C, C]$ et supposons que :

- f est $r + 1$ fois continûment dérivable sur $[0, 1]$.
- il existe p lignes de zero-crossings $z_i(s)$ qui convergent respectivement vers $y_i \in]0, 1[, i = 1, \dots, p$ quand $s \rightarrow 0$, avec $f^{(r)}(y_i) = 0$ et $f^{(r+1)}(y_i) \neq 0$.
- pour $s_{min} \geq s > s_*$, p lignes de zero-crossings $\hat{z}_i(s)$ ont été estimées et convergent en probabilité vers $z_i(s), i = 1, \dots, p$ quand $n \rightarrow \infty$, où s_{min} et s_* sont les échelles définies dans la section 3.5.1. De plus, supposons que pour $i = 1, \dots, p$, les $\hat{z}_i(\cdot)$ sont des fonctions presque sûrement continues de l'échelle s .

Définissons :

$$\begin{aligned} G_z(x) &= \sum_{i=1}^p \int_0^{s_{min}} \frac{1}{s} \theta\left(\frac{x - z_i(s)}{s}\right) ds. \\ \hat{G}_z(x) &= \sum_{i=1}^p \int_{s_*}^{s_{min}} \frac{1}{s} \theta\left(\frac{x - \hat{z}_i(s)}{s}\right) ds. \end{aligned}$$

Supposons que G_z n'est pas constante sur aucun sous-intervalle de $[0, 1]$ et que les extrema locaux de G_z sont non-générés (i.e. $G''_z(y) \neq 0$ si y est un extrema local de G_z). De plus supposons les conditions de la Proposition 3.4 pour les lignes de zero-crossings $z_i(s), i = 1, \dots, p$. Soit $y_1^*, \dots, y_{p^*}^*$ les points y_i (i.e tels que $G_z(x) \rightarrow +\infty$ quand $x \rightarrow y_i, i = 1, \dots, p$), ou les positions des maxima de G_z où celle-ci est dérivable. Soit $\hat{y}_1^*, \dots, \hat{y}_{p^*}^*$ les maxima locaux de \hat{G}_z . Alors, quand $n \rightarrow \infty$:

$$\hat{p}^* \rightarrow_p p^* \text{ et } \hat{y}_i^* \rightarrow_p y_i^*, i = 1, \dots, p^*.$$

Preuve. Définissons :

$$g_{z_i}(x) = \int_0^{s_{min}} \frac{1}{s} \theta\left(\frac{x - z_i(s)}{s}\right) ds \text{ pour } x \in [0, 1] \text{ et } 1 \leq i \leq p.$$

Soit $x \neq y_i$ et définissons $s_{i,x} = \frac{|x_i - x|}{2C}$. Etant donné que $|x_i - z_i(s)| \leq Cs$, nous avons que $\frac{|x - z_i(s)|}{s} \geq C$ pour tout $s \leq s_{i,x}$. Alors,

$$g_{z_i}(x) = \int_{s_{i,x}}^{s_{min}} \frac{1}{s} \theta\left(\frac{x - z_i(s)}{s}\right) ds.$$

Etant données nos hypothèses sur θ , il s'agit d'une fonction bornée et deux fois continûment dérivable sur $[0, 1]$ avec des dérivées bornées. Puisque $s \mapsto \frac{1}{s^3}$ est intégrable sur $[s_{i,x}, s_{min}]$, nous pouvons montrer en utilisant le théorème de dérivation sous le signe somme que g_{z_i} est deux fois continûment dérivable sur $[0, 1] \setminus \{y_i\}$ comme dans la preuve de la Proposition 3.4. Soit $s_n > 0$ une suite d'échelles qui converge vers zéro et $z_{i,n}(s)$ une suite de fonctions continues sur $[0, 1]$ telles que $|z_{i,n}(s) - z_i(s)| \rightarrow 0$ quand $n \rightarrow \infty$, pour $s_n \leq s \leq s_{min}$ et $i = 1, \dots, p$. Définissons :

$$g_{z_{i,n}}(x) = \int_{s_n}^{s_{min}} \frac{1}{s} \theta\left(\frac{x - z_{i,n}(s)}{s}\right) ds, i = 1, \dots, p.$$

Alors, le théorème de dérivation sous le signe intégral implique que $g_{z_{i,n}}$ est deux fois continûment dérivable sur $[0, 1]$. Soit $[a, b]$ un sous-intervalle de $[0, 1]$ tel que $y_i < a < b < y_{i+1}, i = 0, \dots, p$ (avec $y_0 = 0$ et $y_{p+1} = 1$).

Soit \tilde{s} l'échelle la plus large telle que pour tout $s \leq \tilde{s}$ et tout $x \in [a, b]$, $\theta(\frac{x-z_i(s)}{s}) = 0$ et \tilde{s}_n l'échelle la plus large telle que pour tout $s_n \leq s \leq \tilde{s}_n$ et tout $x \in [a, b]$, $\theta(\frac{x-z_{i,n}(s)}{s}) = 0$. Notons que d'après la continuité de $z_i(s)$ et $z_{i,n}(s)$, et le fait que $|z_{i,n}(s) - z_i(s)| \rightarrow 0$ pour $s_n \leq s \leq s_{min}$, nous avons que $\tilde{s}_n \rightarrow \tilde{s}$ quand $n \rightarrow \infty$. Alors, pour $x \in [a, b]$, $k = 0, 1, 2$ et pour n suffisamment large, nous avons :

$$\begin{aligned} |g_{z_{i,n}}^{(k)}(x) - g_{z_i}^{(k)}(x)| &= \left| \int_{\tilde{s}}^{s_{min}} \frac{1}{s^{k+1}} \theta^{(k)}\left(\frac{x-z_i(s)}{s}\right) ds - \int_{\tilde{s}_n}^{s_{min}} \frac{1}{s^{k+1}} \theta^{(k)}\left(\frac{x-z_{i,n}(s)}{s}\right) ds \right| \\ &\leq \|\theta^{(k)}\|_\infty \left| \int_{\tilde{s}}^{s_{min}} \frac{1}{s^{k+1}} ds - \int_{\tilde{s}_n}^{s_{min}} \frac{1}{s^{k+1}} ds \right|, \end{aligned}$$

ce qui implique que $\sup_{x \in [a, b]} |g_{z_{i,n}}^{(k)}(x) - g_{z_i}^{(k)}(x)| \rightarrow 0$ quand $n \rightarrow \infty$. Soit $G_{z,n}(x) = \sum_{i=1}^p g_{z_{i,n}}(x)$. D'après les remarques précédentes, nous avons que pour tout intervalle $[a, b]$ avec $y_i < a < b < y_{i+1}, i = 0, \dots, p$, $\sum_{k=0}^2 \sup_{x \in [a, b]} |G_{z,n}^{(k)}(x) - G_z^{(k)}(x)| \rightarrow 0$ quand $n \rightarrow \infty$. Soit $C^2([a, b])$ l'espace des fonctions deux fois continûment dérivables de $[a, b]$ dans \mathbb{R} , muni de la norme $\|g\|_{\infty,2} = \sum_{p=0}^2 \sup_{t \in [a, b]} |g^{(p)}(t)|$ pour $g \in C^2([a, b])$. Ainsi, nous avons montré que $\|G_{z,n} - G_z\|_{\infty,2} \rightarrow 0$ quand $n \rightarrow \infty$. Nous allons maintenant suivre la méthodologie utilisée dans la section 3.5.2 pour prouver que les positions des extrema de $G_{z,n}$ dans $[a, b]$ convergent vers les positions des extrema de G_z dans $[a, b]$. Définissons les fonctionnelles structurelles suivantes :

Définition 3.3 Soit μ_1 et μ_2 deux réels pré-spécifiés qui ne sont pas dans $]0, 1[$. Soit $[a, b]$ un sous-intervalle de $]0, 1[$. Une fonctionnelle $T : C^2([0, 1]) \rightarrow]0, 1[\cup \{\mu_1, \mu_2\}$ sera appelée une fonctionnelle structurelle de seconde espèce associée à $[a, b]$, si pour $g \in C^2([0, 1])$:

- $T(g) = z$, si g a exactement un extremum local dans $]a, b[$ au point z qui est non-dégénéré i.e. $g''(z) \neq 0$,
- $T(g) = \mu_1$, si g n'a pas d'extremum local dans $]a, b[$,
- $T(g) = \mu_2$, si g a soit plusieurs extrema locaux dans $]a, b[$ ou exactement un extremum au point $z \in]a, b[$ qui est dégénéré (i.e $g''(z) = 0$) .

Alors le Lemme suivant peut être établi :

Lemme 3.6 Soit T une fonctionnelle structurelle de seconde espèce associée à $[a, b]$. Soit $f \in C^2([0, 1])$ et supposons que f est soit strictement monotone sur $[a, b]$ ou a exactement un extremum local dans $]a, b[$ localisé au point z avec $f''(z) \neq 0$, alors T est continue en f .

Pour prouver ce Lemme, nous avons besoin d'utiliser le lemme suivant qui est similaire à la Proposition 1 de Kneip et Gasser [51] :

Lemme 3.7 Soit $f \in C^2([0, 1])$ et soit z la position d'un extremum de f avec $f''(z) \neq 0$. Alors, pour tout $\epsilon > 0$ suffisamment petit, il existe un voisinage $U \subset C^2([0, 1])$ de f tel que pour tout $g \in U$ il existe exactement un $m_g \in [z - \epsilon, z + \epsilon]$ avec $g'(m_g) = 0$ et $\text{sgn}(g''(m_g)) = \text{sgn}(f''(z))$.

Preuve du Lemme 3.7. Soit $\epsilon > 0$. Puisque $f''(z) \neq 0$, f' et f'' sont continues, il existe $\gamma > 0$ tel que pour ϵ suffisamment petit, $|f'(z - \epsilon)| \geq \gamma$, $|f'(z + \epsilon)| \geq \gamma$, $\text{sgn}(f'(z - \epsilon)) = -\text{sgn}(f'(z + \epsilon))$ et $|f''(t)| \geq \gamma$ pour tout $t \in [z - \epsilon, z + \epsilon]$. Soit $U = \{g \in C^2([0, 1]); \|g - f\|_{\infty, 2} < \gamma\}$. Alors, pour tout $g \in U$, nous avons $\sup_{t \in [z - \epsilon, z + \epsilon]} |g'(t) - f'(t)| < \gamma$ ce qui implique que $\text{sgn}(g'(z - \epsilon)) = -\text{sgn}(g'(z + \epsilon)) \neq 0$. Puisque g' est continue, il existe un $m_g \in [z - \epsilon, z + \epsilon]$ avec $g'(m_g) = 0$ et $\sup_{t \in [z - \epsilon, z + \epsilon]} |g''(t) - f''(t)| < \gamma$ ce qui implique que $\text{sgn}(g''(m_g)) = \text{sgn}(f''(z))$. Alors, l'unicité de m_g se déduit de la continuité de g'' . \square

Preuve du Lemme 3.6. Soit $f \in C^2([0, 1])$. Supposons tout d'abord que $T(f) = z$ pour un $z \in]a, b[$ i.e. f a exactement un extremum local dans $]a, b[$ avec $f''(z) \neq 0$. Soit g_n une suite de fonctions de $C^2([0, 1])$ qui converge vers f (par rapport à la norme $\|\cdot\|_{\infty, 2}$). Soit U_ϵ un voisinage de f défini dans le Lemme 3.7. Puisque $|f'(t)| > 0$ pour $t \in [a, T(f) - \epsilon \cup T(f) + \epsilon, b]$, pour n suffisamment grande, $|g'_n(t)| > 0$ pour tout $t \in [a, T(f) - \epsilon \cup T(f) + \epsilon, b]$. Donc, pour n suffisamment grande, $g_n \in U_\epsilon$ et g_n n'a pas d'extremum local dans $[a, T(f) - \epsilon \cup T(f) + \epsilon, b]$, ce qui implique que $T(g_n) \in [T(f) - \epsilon, T(f) + \epsilon]$. Puisque ϵ peut être arbitrairement petit, nous avons $|T(g_n) - T(f)| \rightarrow 0$ quand $n \rightarrow \infty$.

Maintenant, supposons que f est strictement monotone sur $[a, b]$ ce qui implique que $T(f) = \mu_1$. Soit g_n une suite de fonctions de $C^2([0, 1])$ qui converge vers f . Puisque $f'(t) \neq 0$ pour tout $t \in]a, b[$ et étant donné que f' est continue sur $]a, b[$, il existe une constante $\gamma > 0$ telle que $f'(t) \geq \gamma$ ou $f'(t) \leq -\gamma$ pour tout $t \in]a, b[$. Puisque $\|g_n - f\|_{\infty, 2} \rightarrow 0$, nous avons $\sup_{t \in]a, b[} |g'_n(t) - f'(t)| \rightarrow 0$ et donc pour n suffisamment grande, $g'_n(t) > 0$ ou $g'_n(t) < 0$ pour tout $t \in]a, b[$. Alors, pour n suffisamment grande, g_n n'a pas d'extremum local dans $]a, b[$, et donc $T(g_n) = \mu_1$ ce qui implique que $|T(g_n) - T(f)| \rightarrow 0$ quand $n \rightarrow \infty$. \square

Maintenant, soit m_1, \dots, m_k les positions des extrema de G_z dans $[a, b]$ (si G_z n'a d'extremum dans $[a, b]$, poser $k = 0$). D'après nos hypothèses, $G_z(m_j) \neq 0, j = 1, \dots, k$. Ainsi, ces extrema locaux sont distincts les uns des autres ce qui implique qu'il existe des réels $a < a_1 < b_1 < \dots < a_k < b_k < b$ tels que $m_j \in [a_j, b_j], j = 1, \dots, k$ et G_z est strictement monotone sur $[b_j, a_{j+1}]$ (puisque G_z n'est pas constante sur tout sous-intervalle de $[0, 1]$) $j = 0, \dots, k$ (avec $b_0 = a$ et $a_{k+1} = b$). Soient $T_j, j = 1, \dots, k$ des *fonctionnelles structurelles de seconde espèce* respectivement associées à $[a_j, b_j]$, et $\tilde{T}_j, j = 0, \dots, k$ des *fonctionnelles structurelles de seconde espèce* respectivement associées à $[b_j, a_{j+1}]$. Notons que $T_j(G_z) = m_j$ et que $\tilde{T}_j(G_z)$ est égale à μ_1 . De plus, d'après le Lemme 3.6, nous avons que T_i et \tilde{T}_i sont continues en G_z . Puisque $\|G_{z,n} - G_z\|_{\infty, 2} \rightarrow 0$, nous obtenons finalement que $\text{argmax}_{x \in]a, b[} G_{z,n}(x) \rightarrow \text{argmax}_{x \in]a, b[} G_z(x)$ (notons que $\text{argmax}_{x \in]a, b[} G_z(x)$ peut être vide).

Rappelons que d'après la Proposition 3.4, nous avons que $G_z(x) \rightarrow +\infty$ quand $x \rightarrow y_i, i = 1, \dots, p$. De plus, d'après la Proposition 3.4, nous avons également qu'il existe un voisinage V_i de y_i tel que G_z est continûment dérivable sur $V_i \setminus \{y_i\}$, ce qui implique qu'il existe nécessairement deux suites $y_{i,n}^- \rightarrow y_i^-$ et $y_{i,n}^+ \rightarrow y_i^+$ telles que $G'_z(y_{i,n}^-) > 0$ et $G'_z(y_{i,n}^+) < 0$. Puisque $G'_{z,n}(x) \rightarrow G'_z(x)$ pour $x \neq y_i$, il existe nécessairement deux suites $y_{i,n}^- \rightarrow y_i^-$ et $y_{i,n}^+ \rightarrow y_i^+$ telles que pour tout n suffisamment large, $G'_{z,n}(y_{i,n}^-) > 0$ et $G'_{z,n}(y_{i,n}^+) < 0$ pour $i = 1, \dots, p$. Alors, la continuité de $G'_{z,n}$ implique que pour n suffisamment large, $G_{z,n}$ a un maxima local $y_{i,n} \in [y_{i,n}^-, y_{i,n}^+]$ pour $i = 1, \dots, p$, ce qui implique que $y_{i,n} \rightarrow y_i$ quand $n \rightarrow \infty$. Ainsi, nous avons finalement prouvé que les maxima locaux de $G_{z,n}$ convergent soit vers $y_i, i = 1, \dots, p$, soit vers la position d'un maxima G_z où celle-ci est dérivable. D'après nos hypothèses, nous avons que les lignes de zero-crossings $\hat{z}_i(s)$ sont des fonctions presque sûrement continues sur $[0, 1]$, et puisque $s_* \rightarrow 0$ et $\hat{z}_i(s) \rightarrow_p z_i(s)$ pour $s_* \leq s \leq s_{min}$, le résultat de déduit du théorème de l'application continue (continuous mapping theorem, voir Billingsley [7], Corollary 2 p.31). \square

Remarque 3.5 *En pratique, les intensités structurelles des maxima d'ondelettes et des zero-crossings estimés seront normalisées pour être des densités de probabilité.*

3.7 Simulations

Nous proposons maintenant d'effectuer quelques simulations pour tester les méthodes proposées dans les sections précédentes. Les données simulées que nous allons utiliser sont obtenues à partir du modèle :

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n \quad (3.3.27)$$

où $x_i = \frac{i}{n}$, f est une fonction inconnue qui peut avoir plusieurs types de landmarks sur $]0, 1[$ et ϵ_i sont des variables i.i.d de loi normale de moyenne nulle et de variance σ^2 .

L'ondelette utilisée est la dérivée d'ordre r d'une B-Spline de degré $q = 4$ (voir section 3.2.4.2). La valeur de σ est choisie pour correspondre à différentes valeurs de la racine du rapport signal-sur-bruit ($RSNR$) :

$$RSNR(f, \sigma) = \frac{\sqrt{\int_0^1 (f(x) - \bar{f})^2 dx}}{\sigma}, \quad \text{où } \bar{f} = \int_0^1 f(x) dx.$$

Pour différents niveaux de $RSNR$ et différents signaux f , les zero-crossings ou les maxima d'ondelettes de $W_s(f)(x)$ sont estimés pour $r = 1$ ou pour $r = 2$, selon les landmarks que l'on souhaite détecter.

La transformée continue en ondelettes du signal discret $y_i, i = 1, \dots, n$ est calculée pour les échelles dyadiques $s = 2^{-j}$ avec 20 niveaux par octave pour $1 \leq j \leq (\log_2(n) - 1)$. Les n coefficients d'ondelettes empiriques à l'échelle $s = 2^{-(j+k/\nu)}$

seront dénotés par $\hat{w}_{j+k/\nu,i}$, $i = 1, \dots, n$ (où ν est le nombre de niveaux par octave, et $0 \leq k \leq \nu - 1$). Si n est une puissance de deux, la Transformée de Fourier Rapide (FFT) est utilisée pour l'implémentation. Cependant, puisque nous utilisons des ondelettes B-Splines, nous pouvons remarquer que des algorithmes plus rapides pourraient être utilisés comme celui proposé par exemple par Wang et Lee [90], bien qu'une implémentation plus rapide est principalement nécessaire pour le cas 2D. Dans cette section, les axes horizontaux et verticaux des figures représentant des lignes de zero-crossings ou de maxima d'ondelettes, donnent respectivement x et $-\log_2(s)$.

Estimation de la variance : Jusqu'à maintenant, nous avons supposé que le niveau du bruit σ était connu, ce qui n'est pas le cas en pratique. Pour estimer le niveau de bruit σ , nous proposons d'utiliser l'estimateur robuste suggéré par Donoho et Johnstone [31] basé sur la *déviation absolue de la médiane* (MAD) des coefficients d'ondelettes empiriques associés à une base orthonormée de $L^2([0, 1])$. Soit $\mathbf{y} = (y_1, \dots, y_n)$ un vecteur de données bruitées. La transformée en ondelettes discrète de y est donné par :

$$\hat{\mathbf{d}} = W\mathbf{y},$$

où $\hat{\mathbf{d}}$ est un vecteur de dimension n comprenant à la fois les coefficients discrets d'échelle et d'ondelettes empiriques, et W est une matrice orthogonale $n \times n$ associée à la base orthonormée choisie. Si $n = 2^J$, alors la transformée en ondelettes discrète peut être calculée au moyen d'un algorithme rapide développé par Mallat [61] en seulement $O(n)$ opérations. Au niveau de résolution le plus fin, les coefficients d'ondelettes empiriques tendent à consister principalement de bruit. Puisque qu'il existe tout de même "un peu de signal" présent à ce niveau, Donoho et Johnstone [31] ont proposé un estimateur robuste pour estimer le niveau de bruit σ donné par la déviation de la médiane absolue aux niveaux de résolution le plus fin :

$$\hat{\sigma} = \sqrt{n} \frac{\text{median}(\{|\hat{d}_{J-1,k}| : k = 0, \dots, 2^{J-1} - 1\})}{0.6745}.$$

Dans nos simulations, nous avons pris la base d'ondelettes Symmlet 8 (comme décrit page 198 de Daubechies [28]) pour estimer σ .

Rappelons que l'estimation des landmarks de f sera effectuée via le calcul de l'intensité structurelle des zero-crossings et des maxima d'ondelettes estimés. Nous explorons donc tout d'abord la qualité de l'estimation de cette méthode sur quelques exemples simulés, et expliquons comment des estimateurs erronés induits par la présence de bruit peuvent être éliminés.

3.7.1 Estimation des lignes de zero-crossings

Tout d'abord, nous devons remarquer que la procédure de tests d'hypothèses décrite dans la section 3.5 pour estimer les zero-crossings de $W_s(f)(x)$ doit être

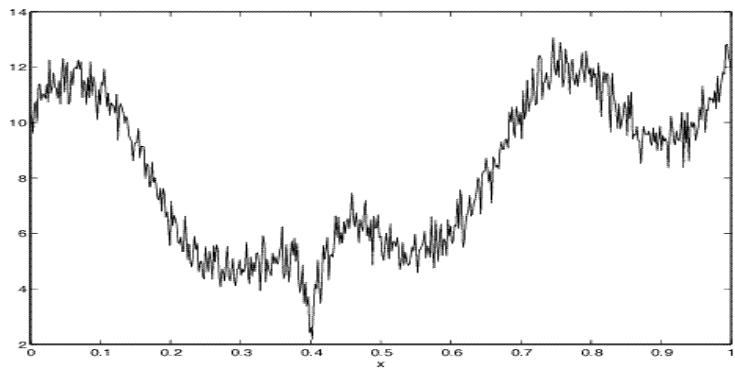
légèrement modifiée en pratique. Dans la section 3.5, pour $s_{min} \geq s > s_*$, nous avons défini $\hat{z}_{s,1}, \dots, \hat{z}_{s,\hat{p}_s}$ comme les zero-crossings de $W_{s,n}(Y)(x)$ qui sont localisés dans les intervalles $[\hat{x}_{s,i_1}, \hat{x}_{s,i_2}]$, où $\{(\hat{x}_{s,i_1}, \hat{x}_{s,i_2}); i_1, i_2 \in \mathbb{N}\}$ est la suite de points de $]0, 1[\times]0, 1[$ telle que :

- $H_0^{s,\hat{x}_{s,i_1}}$ et $H_0^{s,\hat{x}_{s,i_2}}$ sont rejetées.
- $W_{s,n}(Y)(\hat{x}_{s,i_1})$ et $W_{s,n}(Y)(\hat{x}_{s,i_2})$ ont des signes opposés.
- toutes les hypothèses $H_0^{s,x}$ pour $\hat{x}_{s,i_1} < x < \hat{x}_{s,i_2}$ sont acceptées.

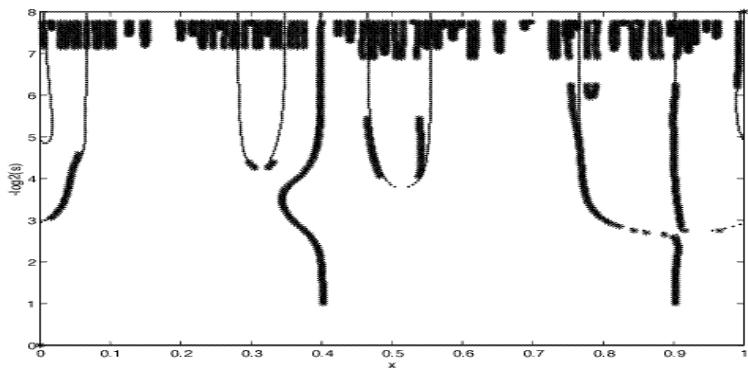
Cependant, le niveau s_* qui est l'échelle la plus fine à laquelle les zero-crossings sont estimés, est inconnu en pratique. Une idée naturelle pour l'estimer serait de définir \hat{s}_* comme l'échelle la plus fine où $\max_{x \in [0,1]} \{|W_{s,n}(Y)(x)|\}$ dépasse le seuil $\frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$, puisqu'un zero-crossing significatif à l'échelle s doit être localisé entre une paire de points pour lesquels les hypothèses nulles $H_0^{s..}$ sont rejetées. Cependant, le signal f peut avoir des singularités qui vont donner lieu à de grands coefficients d'ondelettes dont la valeur absolue est plus grande que le seuil $\frac{\sigma}{\sqrt{n}} q_{1-\alpha}^n$ et entre lesquels il peut y avoir de nombreux zero-crossings erronés dus aux fluctuations du bruit. En particulier, l'implémentation pratique de la CWT donnera généralement de larges coefficients aux bornes de $[0, 1]$ dus à des transitions brutales (si $f(0)$ et $f(1)$ sont très différents) lorsque l'on rend f périodique pour calculer sa transformée en ondelettes. Par exemple, supposons que l'on souhaite estimer les zero-crossings de $f_1(x) = 2 \sin(6\pi x) + 15|x - 0.4|^{2/5}$, $x \in [0, 1]$ avec $r = 1$, $RSNR = 5$ et $n = 512$ pour $\hat{s}_* \leq s \leq s_{min}$ (dans nos simulations nous avons pris $\alpha = 0.05$ pour la procédure de test, ce qui sera le cas pour toutes les fois où des zero-crossings seront détectés). Le résultat, donné dans la Figure 3.8b, est extrêmement mauvais car de nombreux zero-crossings erronés sont estimés aux fines échelles. Ainsi, nous suggérons de modifier légèrement la procédure d'estimation de la façon suivante : pour $s \leq s_{min}$, soit $\{(\hat{x}_{1,s,i}, \hat{x}_{2,s,i}); 1 \leq i \leq \hat{p}_s\}$ deux suites de points de $]0, 1[\times]0, 1[$ telles que :

- $H_0^{s,\hat{x}_{1,s,i}}$ et $H_0^{s,\hat{x}_{2,s,i}}$ sont rejetées.
- $W_{s,n}(Y)(\hat{x}_{1,s,i})$ et $W_{s,n}(Y)(\hat{x}_{2,s,i})$ ont des signes opposés.
- toutes les hypothèses $H_0^{s,x}$ pour $\hat{x}_{1,s,i} < x < \hat{x}_{2,s,i}$ sont acceptées.

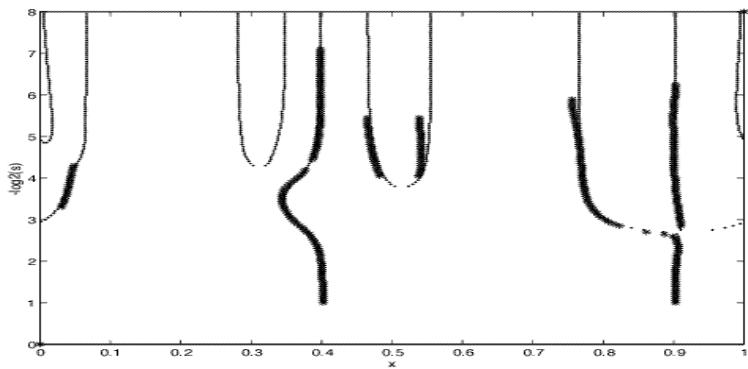
Pour $1 \leq i \leq \hat{p}_s$, si $W_{s,n}(Y)$ a **seulement un zero-crossings** $z_{1,s,i}$ dans $[\hat{x}_{1,s,i}; \hat{x}_{2,s,i}]$ nous posons $\hat{z}_i(s) = z_{1,s,i}$ sinon les zero-crossings localisés dans $[\hat{x}_{1,s,i}; \hat{x}_{2,s,i}]$ ne sont pas pris en compte puisque que la plupart d'entre eux sont probablement dus aux fluctuations du bruit. Dans ce qui suit, nous appellerons cette méthode l'*estimation empirique des zero-crossings*. Le résultat de cette approche empirique est donné dans la Figure 3.8c. Elle est très satisfaisante car tous les zero-crossings erronés détectés dans la Figure 3.8b ont été éliminés, et seules les lignes de zero-crossings les plus significatives ont été correctement estimées. Dans toutes les simulations suivantes, les zero-crossings sont donc estimés avec cette méthode. De plus, dans tout le reste de la thèse, nous utiliserons systématiquement cette approche pour estimer les zero-crossings d'un signal.



(a)



(b)



(c)

FIG. 3.8: (a) Signal f_1 . (b) Zero-crossings estimés (lignes épaisses) et vrais zero-crossings de f_1 (lignes fines) pour $r = 1$ et $\hat{s}_* \leq s \leq s_{min}$. (c) Zero-crossings estimés (lignes épaisses) et vrais zero-crossings de f_1 (lignes fines) pour $r = 1$ avec l'approche empirique.

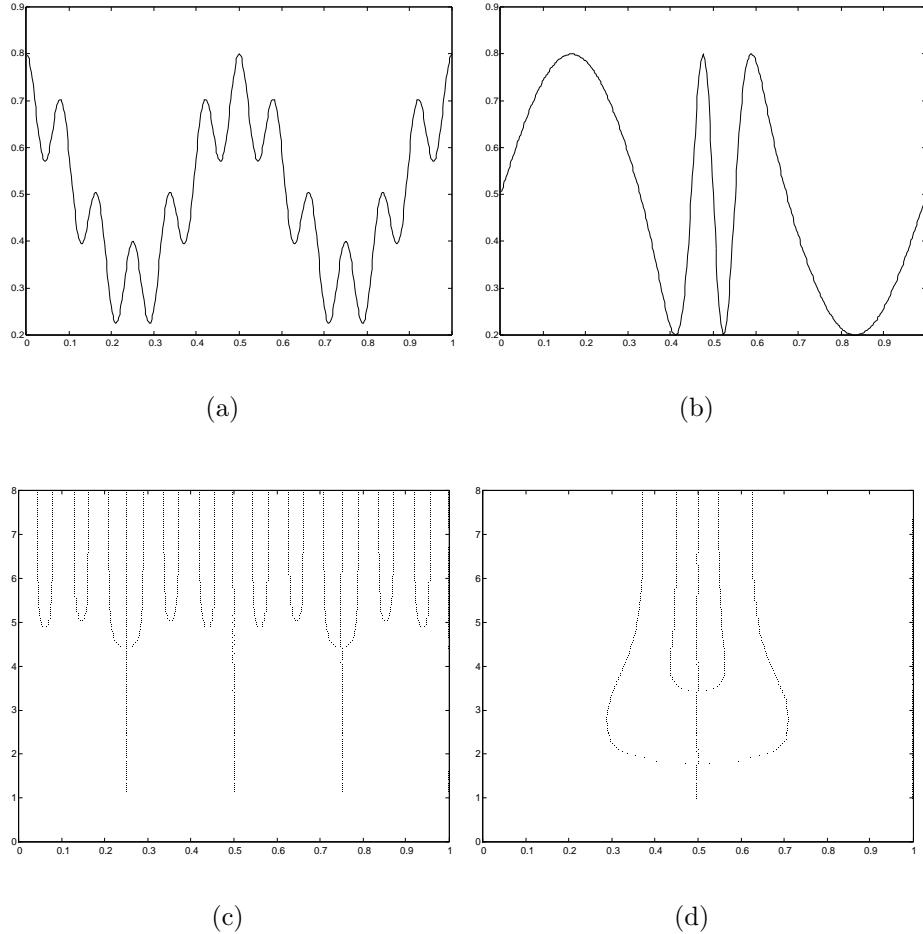


FIG. 3.9: (a) Signal Wave f_W . (b) Lignes de zero-crossings de f_W calculées avec une ondelette B-Spline d'ordre $q = 4$ avec $r = 1$ moment nul. (c) Time Shifted Sine f_T . (d) Lignes de zero-crossings de f_T calculées avec une ondelette B-Spline d'ordre $q = 4$ avec $r = 2$ moments nuls.

Dans cette section, les fonctions tests sont le signal Wave qui est une somme de deux sinusoïdes périodiques : $f_W(x) = 0.5 + 0.2 \cos(4\pi x) + 0.1 \cos(24\pi x)$, $x \in [0, 1]$ et le Time Shifted Sine $f_T = 0.3 \sin 3\pi[g(g(g(g(x)))) + x] + 0.5$, $x \in [0, 1]$, où $g(x) = (1 - \cos(\pi x))/2$. Dans les simulations suivantes, les zero-crossings de f_W pour $r = 1$ (i.e. positions des extrema) et des zero-crossings de f_T pour $r = 2$ (i.e. les positions des points d'inflexion) seront estimés. Ces deux signaux (sans bruit) et leurs lignes de zero-crossings sont tracés dans la Figure 3.9.

La Figure 3.10 donne un exemple d'estimation des lignes zero-crossings de f_W (avec $r = 1$) et f_T (avec $r = 2$) pour $RSNR = 5$ et $n = 512$. Le résultat est très satisfaisant puisqu'il peut être constaté dans les Figure 3.10c et Figure 3.10d que toutes les lignes sont correctement estimées. Les maxima locaux des intensités structurelles associées correspondent aux landmarks de chaque fonction, et la forme des

pics des Figures 3.10ef est reliée à la longueur et la position le long des échelles des lignes de zero-crossings. Un pic “aigu” correspond à une ligne qui se propage aux fines échelles, alors qu’un pic “plat” est lié à une ligne qui a été seulement estimée à des échelles grossières.

3.7.2 Estimation des lignes de maxima d’ondelettes

Pour illustrer la détection des singularités d’un signal, les fonctions de tests sont $f_3(x) = 2 \sin(6\pi x) + 15|x - 0.4|^{2/5} - 15|x - 0.7|^{3/5} + 15|x - 0.2|^{4/5}$, $x \in [0, 1]$ et la fonction Heavisine de Donoho et Johnstone [31] f_H (remise à l’échelle verticalement sur $[0.2, 0.8]$). Le signal f_3 a 3 singularités localisées aux points $x = 0.2$, $x = 0.4$ et $x = 0.7$ d’ordre $4/5$, $2/5$ et $3/5$ respectivement. En dehors de ces 3 points, f_3 a de nombreux extrema et points d’infexion où celle-ci est infiniment continûment dérivable (voir la Figure 3.11d). La fonction HeaviSine a deux points de discontinuités $x = 0.3$ et $x = 0.72$ et est infiniment continûment dérivable ailleurs (voir Figure 3.11a). Dans les simulations suivantes, les maxima d’ondelettes de f_3 pour $r = 2$, et les maxima d’ondelettes de f_H pour $r = 1$ et $r = 2$ sont estimés. Ces deux signaux (sans bruit) et leur lignes de maxima d’ondelettes sont représentés dans la 3.11.

Nous proposons de comparer deux types d’estimateurs des singularités d’un signal :

- les maxima locaux de l’intensité structurelle des maxima d’ondelettes *non-pondérée* (voir Proposition 3.15) :

$$\hat{G}_m(x) = \sum_{i=1}^q \int_{\hat{s}_{\alpha_{min}^*}}^{s_r} \frac{1}{s} \theta\left(\frac{x - \hat{m}_i(s)}{s}\right) ds,$$

- les maxima locaux de l’intensité structurelle des maxima d’ondelettes *pondérée* :

$$\hat{G}_{W,m}(x) = \sum_{i=1}^q \int_{\hat{s}_{\alpha_{min}^*}}^{s_r} \frac{\hat{h}_i(s)}{s} \theta\left(\frac{x - \hat{m}_i(s)}{s}\right) ds,$$

où θ est le noyau utilisé pour calculer l’ondelette ψ , $\hat{s}_{\alpha_{min}^*}$ est le niveau de résolution le plus fin où $\max_{x \in [0,1]} |W_{s,n}(Y)(x)|$ dépasse le seuil $\lambda = \frac{\hat{\sigma}\sqrt{2 \log(n \log(n))}}{\sqrt{n}}$, et $\hat{h}_i(s) = \frac{W_{s,n}(Y)(\hat{m}_i(s))}{s^{r+1/2}}$. Comme expliqué dans la section 3.2.4.3, les hauteurs des pics de $\hat{G}_{W,m}$ sont inversement proportionnelles à l’ordre des singularités d’un signal. Calculer $\hat{G}_{W,m}$ est également un moyen de décider si une ligne de maxima d’ondelettes converge vers une singularité de f ou vers un zéro de $f^{(r+1)}$. Les résultats de l’estimation des lignes de maxima d’ondelettes de f_3 et de f_H (pour $RSNR = 5$ et $n = 512$) sont donnés dans les Figures 3.12 et 3.13. Dans la Figure 3.12b, on peut constater que la procédure détecte des lignes qui converge vers les trois singularités de f_3 , la ligne de maxima d’ondelettes qui converge vers 0.59 qui est un zéro de $f^{(3)}$, et les

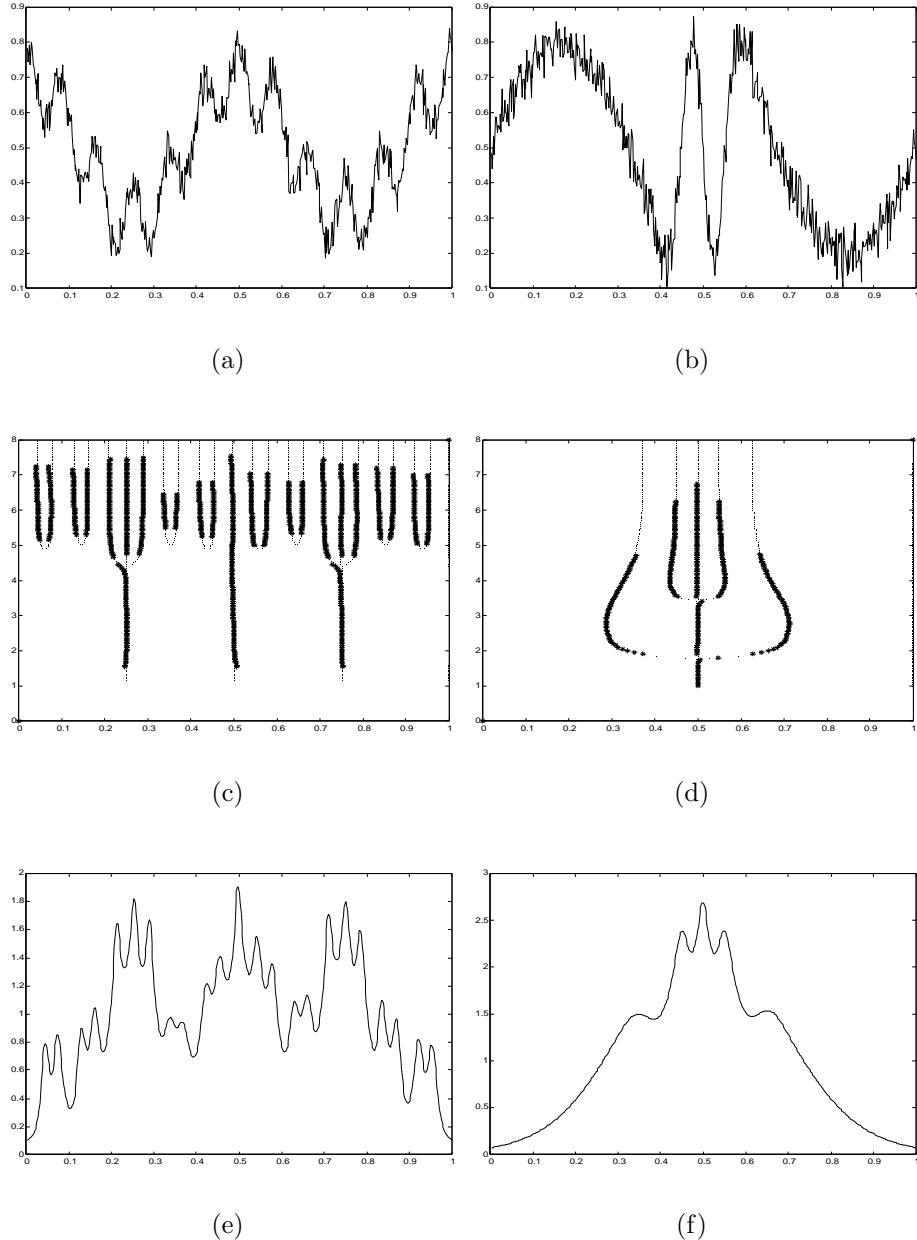


FIG. 3.10: (a) Signal bruité f_W avec $RSNR = 5, n = 512$. (c) Zero-crossings estimés (lignes épaisses) et vrais zero-crossings de f_W pour $r = 1$. (d) Intensité structurelle des zero-crossings estimés de f_W . (b) Signal bruité f_T avec $RSNR = 5, n = 512$. (d) Zero-crossings estimés (lignes épaisses) et vrais zero-crossings de f_T pour $r = 2$. (f) Intensité structurelle des zero-crossings estimés de f_T .

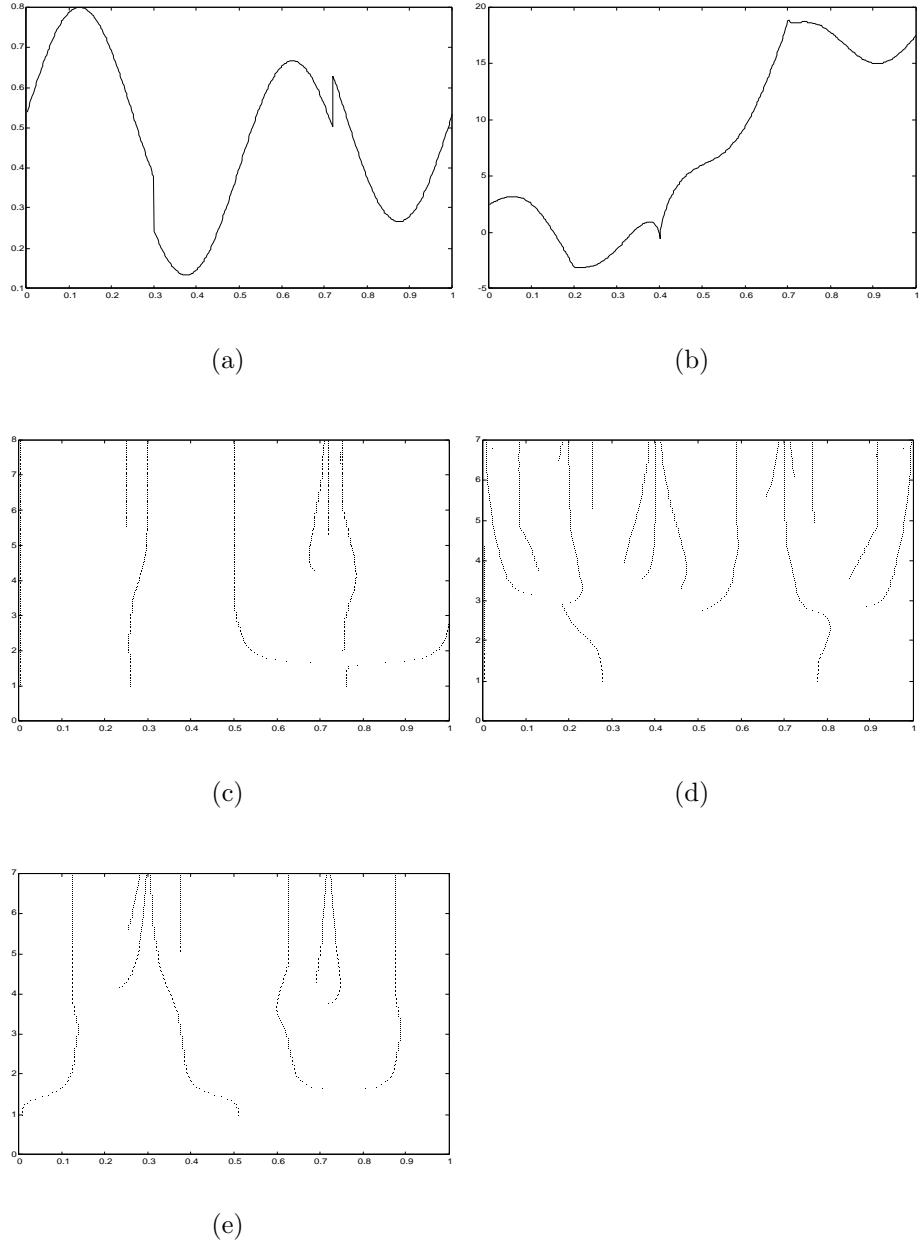


FIG. 3.11: (a) Fonction HeaviSine f_H . (c) Lignes de maxima d'ondelettes de f_H calculées avec une ondelette B-Spline d'ordre $q = 4$ avec $r = 1$ moment nul. (e) Lignes de maxima d'ondelettes de f_H calculées avec une ondelette B-Spline d'ordre $q = 4$ avec $r = 2$ moments nuls. (b) Signal f_3 . (d) Lignes de maxima d'ondelettes de f_3 calculées avec une ondelette B-Spline d'ordre $q = 4$ avec $r = 2$ moments nuls.

lignes dues aux discontinuités du signal aux bornes de $[0, 1]$. L'intensité structurelle non-pondérée représentée dans la Figure 3.12c a 7 maxima localisés aux voisinages des limites des lignes estimées. Cependant, d'après cette fonction il n'est pas possible de décider quels pics correspondent effectivement à une singularité. Ce problème peut être résolu en calculant $\hat{G}_{W,m}(x)$ (voir Figure 3.12d) où les pics principaux sont exactement localisés aux trois singularités de f_3 (remarquons que $\hat{G}_{W,m}(x)$ est seulement tracée sur $[0.04; 0.96]$ car l'intensité des deux pics aux points $x = 0$ et $x = 1$ est très élevée, et ceux-ci “masquent” les autres pics quand toute la fonction $\hat{G}_{W,m}(x)$ est représentée sur $[0, 1]$). Ainsi, les bons résultats obtenus avec $\hat{G}_{W,m}(x)$ pour f_3 suggèrent l'utilisation de l'intensité structurelle *pondérée* en pratique. Cependant, l'estimation des singularités de f_H n'est pas si satisfaisante. Dans les Figures 3.13e et 3.13f, les fonctions $\hat{G}_{W,m}(x)$ ont deux pics principaux localisés au niveau des discontinuités de f_H (i.e. aux points $x = 0.3$ et $x = 0.72$). Cependant, pour $r = 1$ et $r = 2$, il y a de nombreux maxima d'ondelettes erronés qui donnent de petits pics qui ne correspondent pas à des landmarks de f_H . Dans la section suivante, nous décrivons donc une approche empirique pour supprimer ce genre de pics erronés.

3.7.3 Bagging pour supprimer les estimations erronées

La prédiction par Bagging est une méthode pour générer des versions multiples d'un estimateur et utiliser celles-ci pour obtenir un estimateur qui en fait l'agrégation et qui conduit à un gain substantiel en précision. En général, l'agrégation consiste à moyenner les différentes versions lorsque l'on estime une quantité numérique (par exemple dans le cas de la régression). Les versions multiples sont formées en construisant par bootstrap des “copies” des données et en utilisant celles-ci comme nouvelles données. Pour une application du bagging pour des arbres de classification et pour la régression linéaire, nous renvoyons au papier de Breiman [9]. Généralement, le bagging réduit la variance sans augmenter de manière substantielle le biais. Dans notre problème, plutôt que de faire de la réplication par bootstrap des données séquence $Y = \{y_i, i = 1, \dots, n\}$, nous proposons de générer M nouvelles séquences de données $Y^{(k)} = \{y_i^{(k)}, i = 1, \dots, n\}$ en ajoutant un bruit Gaussien aux observations $\{y_i, i = 1, \dots, n\}$, i.e. :

$$y_i^{(k)} = y_i + \epsilon_i^{(k)}, \quad i = 1, \dots, n, \quad k = 1, \dots, M,$$

où $\epsilon_i^{(k)}$ sont des variables i.i.d. de loi normale, de moyenne nulle et de variance σ_*^2 . Pour chaque séquence $Y^{(k)}$, nous calculons soit l'intensité structurelle $\hat{G}_z^{(k)}(x)$ des lignes de zero-crossings estimées, soit l'intensité structurelle $\hat{G}_m^{(k)}(x)$ des lignes de maxima d'ondelettes estimées. Alors, nous définissons nos estimateurs par bagging

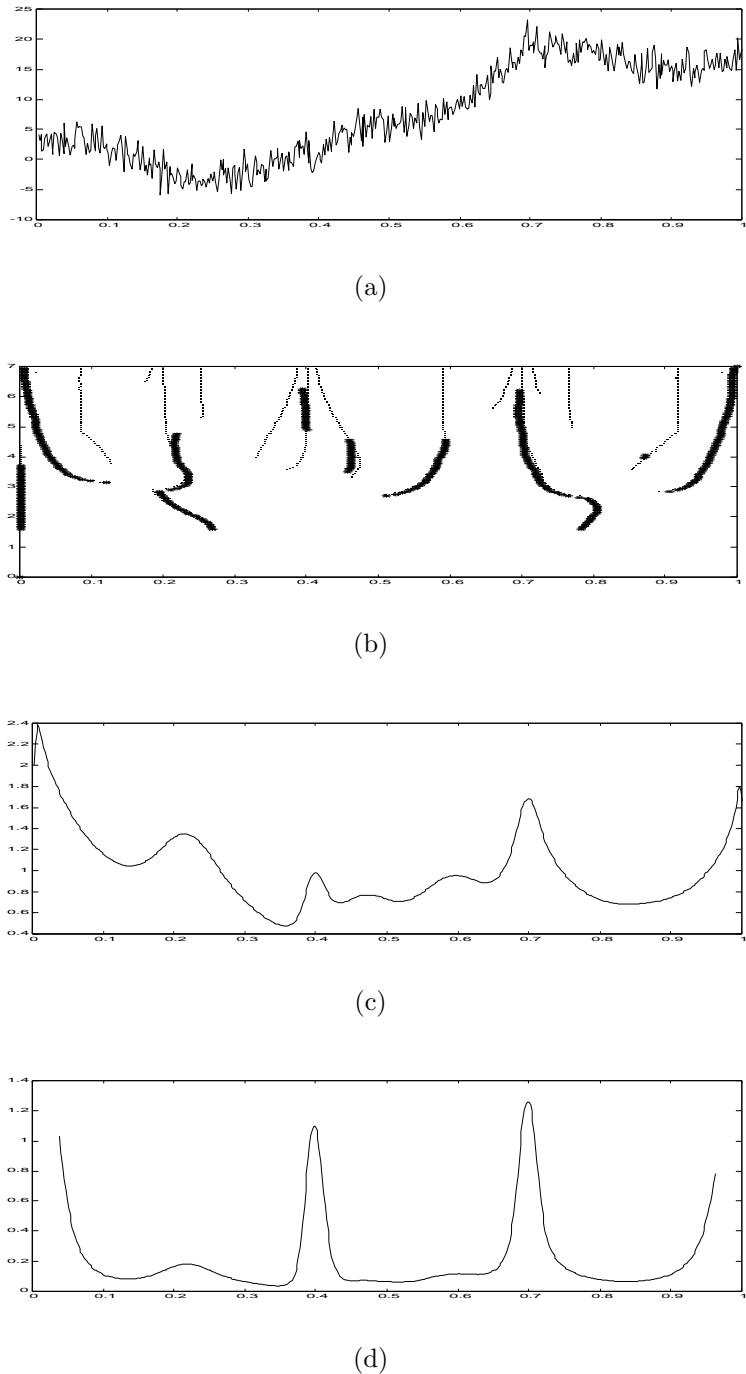


FIG. 3.12: (a) Signal bruité f_3 avec $RSNR = 5, n = 512$. (b) Maxima d'ondelettes estimés (lignes épaisses) et vrais maxima d'ondelettes (lignes fines) pour $r = 2$. (c) Intensité structurelle des maxima d'ondelettes estimés non-pondérée (d) Intensité structurelle des maxima d'ondelettes estimés pondérée.

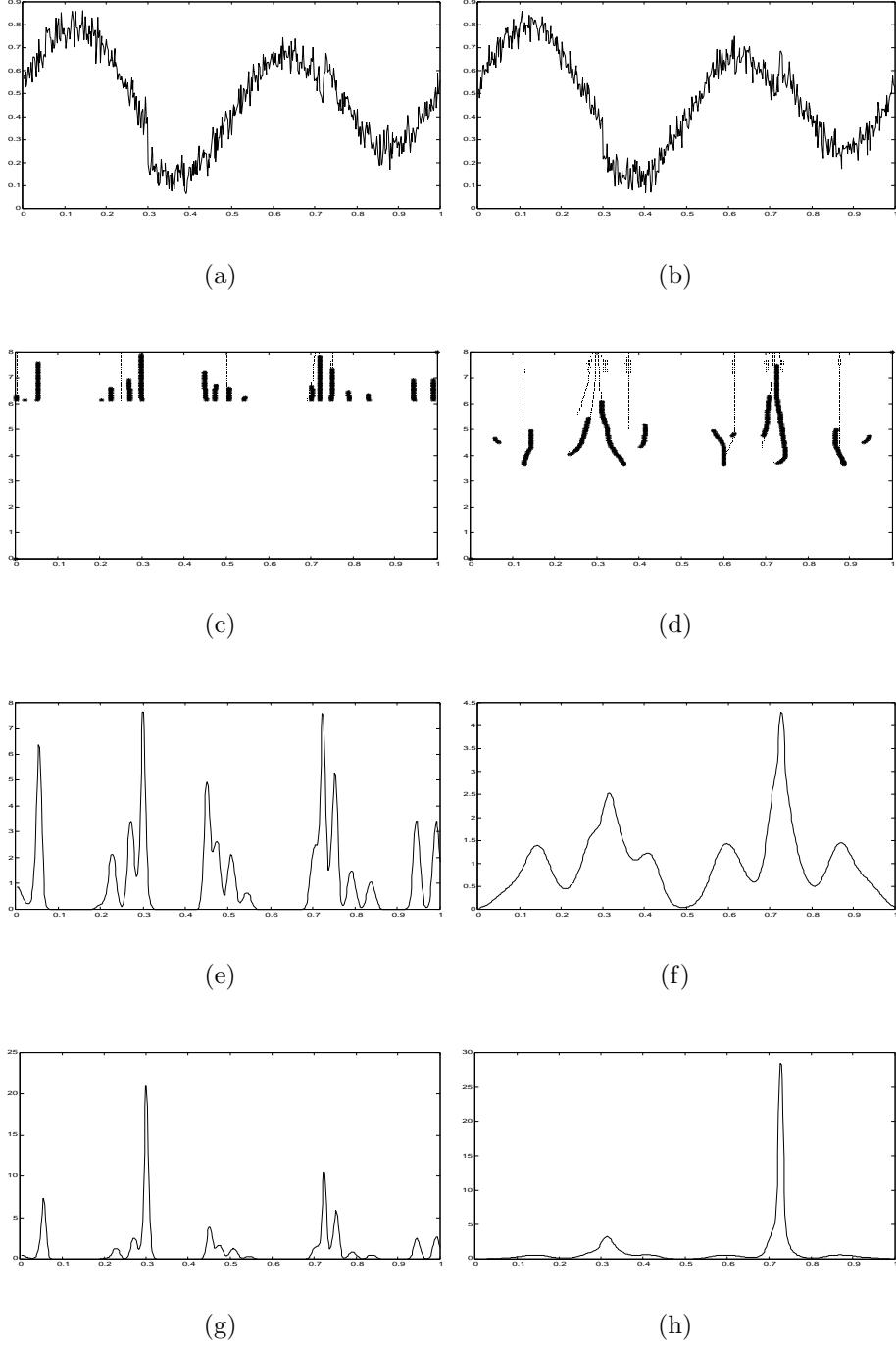


FIG. 3.13: (a) Signal bruité f_H avec $RSNR = 5, n = 512$. (c) Maxima d'ondelettes estimés (lignes épaisses) et vrais maxima d'ondelettes (lignes fines) pour $r = 1$. (e) Intensité structurelle des maxima d'ondelettes estimés non-pondérée pour $r = 1$ (g) Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 1$. (b) Signal bruité f_H avec $RSNR = 5, n = 512$. (d) Maxima d'ondelettes estimés (lignes épaisses) et vrais maxima d'ondelettes (lignes fines) pour $r = 2$. (f) Intensité structurelle des maxima d'ondelettes estimés non-pondérée pour $r = 2$ (h) Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 2$.

comme :

$$\begin{aligned}\hat{G}_z^{(Bag)}(x) &= \frac{1}{M} \sum_{k=1}^M \hat{G}_z^{(k)}(x), \quad x \in [0, 1], \\ \hat{G}_m^{(Bag)}(x) &= \frac{1}{M} \sum_{k=1}^M \hat{G}_m^{(k)}(x), \quad x \in [0, 1],\end{aligned}$$

Lorsque l'on perturbe les données Y , nous pensons que les pics des intensités structurelles qui correspondent aux vraies lignes de zero-crossings ou de maxima d'ondelettes resteront à la même position alors que celles qui sont dues au bruit seront instables. Ainsi, moyennant sur les intensités perturbées $\hat{G}_z^{(k)}(x)$, $k = 1, \dots, M$ va donner lieu à des pics importants au voisinage des landmarks de f et à des maxima plats et petits ailleurs. L'injection de bruit dans un ensemble d'apprentissage et le moyennage d'ensembles a été appliqué avec succès par Raviv et Intrator [78] avec des réseaux de neurones au problème des deux spirales. Leur approche est également utile pour les modèles additifs généralisés. Leurs simulations montrent que cette approche peut effectivement réduire la variance des estimateurs. Des méthodes de ré-échantillonnage, basée sur des variations des données originales, ont également été utilisées par Minotte *et al.* [68] pour la détection des modes d'une densité inconnue. Leur approche est similaire à notre méthode de bagging puisque celle-ci consiste à regarder simultanément une large collection de Mode Tree, qui sont des représentations temps-échelles des caractéristiques typiques d'une densité.

Afin d'illustrer les performances des estimateurs $\hat{G}_m^{(Bag)}(x)$, nous avons appliqué notre procédure de bagging aux exemples d'estimation des maxima d'ondelettes de f_H pour $r = 1$ et $r = 2$ (voir Figure 3.11) pour lesquels il existe de nombreux maxima d'ondelettes erronés. $M = 100$ nouvelles séquences de données sont générées en ajoutant un bruit Gaussien de moyenne nulle et de variance σ_*^2 égal à la variance estimée $\hat{\sigma}^2$. A ce point de notre explication, on peut se demander s'il serait possible de trouver une procédure pour automatiquement définir ce paramètre. Ce problème est envisagé à la fin de cette section dans une discussion sur les différents hyperparamètres qui ont été introduits. Notons que les intensités structurelles sont calculées *avec pondération*. Le résultat du bagging donné dans la Figure 3.14e montre qu'il semble s'agir d'une méthode très prometteuse. Comparée avec l'intensité structurelle de la Figure 3.14c, les pics erronés aux points $x \neq 0.3$ et $x \neq 0.72$ sont devenus des maxima plats et petits ou bien ont complètement disparus, alors que les "modes importants" de $\hat{G}_m^{(Bag)}(x)$ correspondent aux deux discontinuités de f_H . Pour l'estimation des maxima d'ondelettes avec $r = 1$, le résultat n'est pas si satisfaisant. Les pics autour de $x = 0.5$ et $x = 0.8$ dus aux estimations erronées dans la Figure 3.14d sont petits et plats dans la Figure 3.14f, mais l'intensité structurelle après bagging représentée dans la Figure 3.14f a seulement un "mode significatif" au point $x = 0.3$ et des "modes mineurs" ailleurs (la procédure a manqué la discontinuité au point $x = 0.72$).

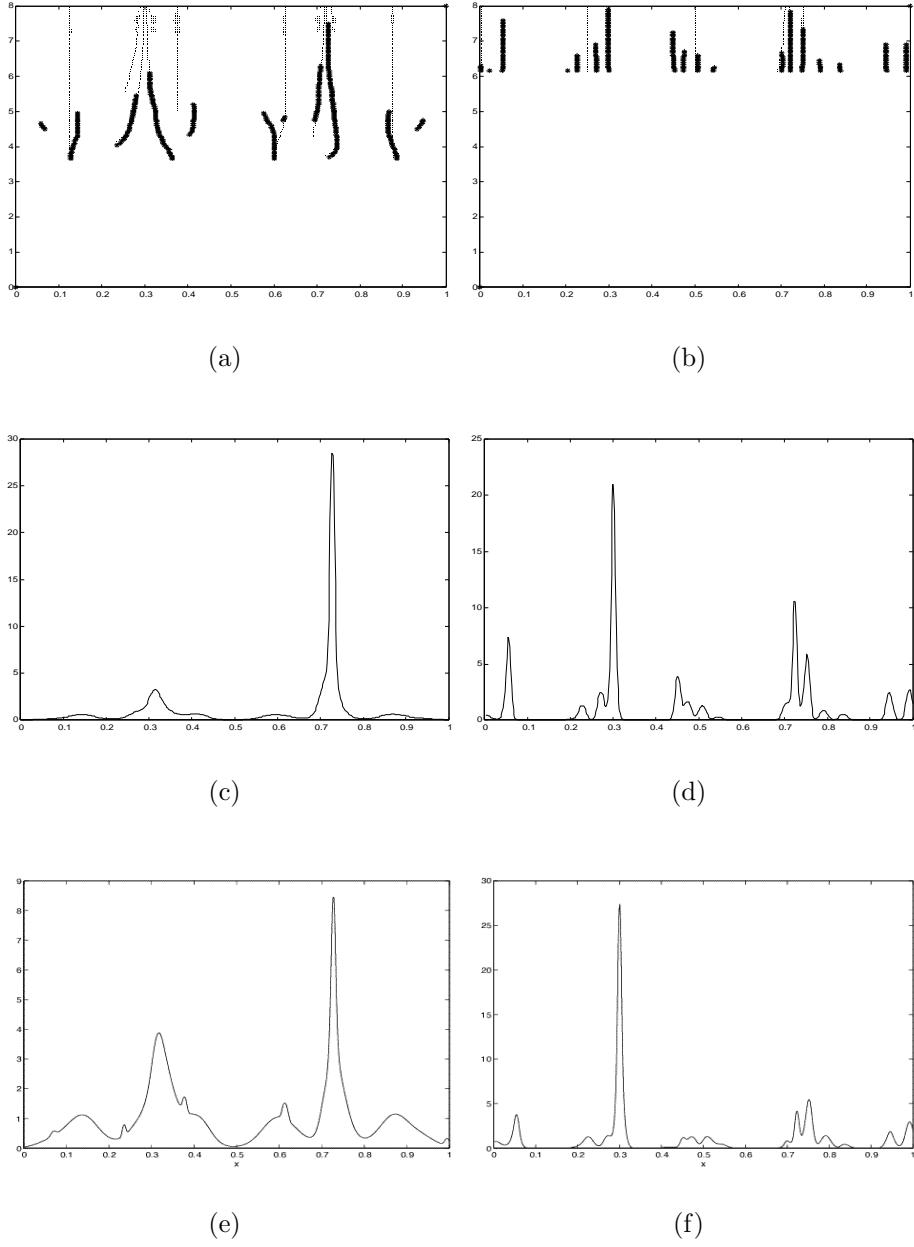


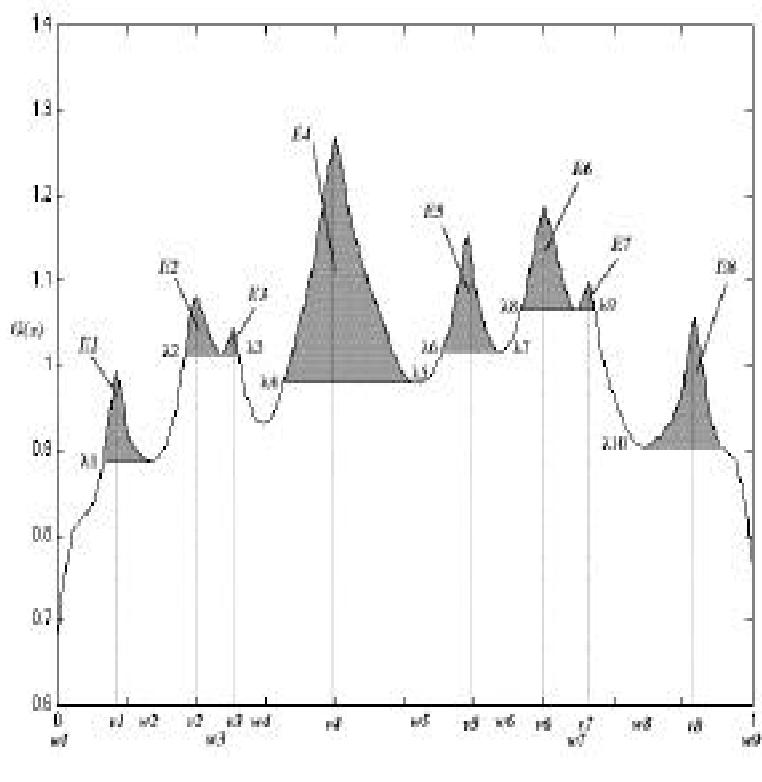
FIG. 3.14: (a) Maxima d'ondelettes estimés (lignes épaisses) et vrais maxima d'ondelettes (lignes fines) de f_H pour $r = 2$ ($RSNR = 5, n = 512$). (c) Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 2$ (e), Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 2$, $\hat{G}_m^{(Bag)}(x)$ (après bagging, $M = 100$). (b) Maxima d'ondelettes estimés (lignes épaisses) et vrais maxima d'ondelettes (lignes fines) de f_H pour $r = 1$ ($RSNR = 5, n = 512$). (d) Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 1$ (f), Intensité structurelle des maxima d'ondelettes estimés pondérée pour $r = 1$, $\hat{G}_m^{(Bag)}(x)$ (après bagging, $M = 100$).

Dans la discussion ci-dessus, une distinction a été faite entre les modes importants de $\hat{G}_m^{(Bag)}(x)$ qui révèlent la présence d'un landmark significatif, et d'autres modes qualifiés de mineurs qui ne contiennent pas d'information, et peuvent donc être négligés. Puisque les intensités structurelles ont été normalisées pour être des densités de probabilité, une masse peut être attribuée à chaque maxima local de $\hat{G}_m^{(Bag)}(x)$ pour déterminer son importance. Une telle mesure de modalité a été proposée par Fisher et Marron [38] pour le problème de l'identification du nombre de modalités dans échantillon de données. Leur mesure de modalité est illustrée dans la Figure 3.15 où la masse assignée à chaque mode est dénotée par chaque aire grisée. Assigner une masse à chaque maximum local de $\hat{G}_m^{(Bag)}(x)$ permet de négliger les modes mineurs quand leur masse est inférieure à un seuil pré-spécifié m_0 . Les détails du calcul des valeurs E_j et λ_j dans la Figure 3.15 peuvent être trouvés dans l'Appendix 1 de Fisher et Marron [38].

Si nous mesurons l'importance des modes de $\hat{G}_m^{(Bag)}(x)$ dans la Figure 3.14e ($r = 2$) et gardons uniquement ceux dont la masse excède $m_0 = 0.1$ (i.e. 10% de la masse totale), les modes mineurs erronés sont éliminés et nous gardons uniquement les modes qui correspondent à l'existence significative d'un landmark de f_H (i.e. ceux aux points 0.32 et 0.72). Les positions des modes des intensités structurelles après bagging (voir Figure 3.14e-3.14f) avec leur masse associée sont données dans la Table 3.1.

Pour conclure ces discussions sur l'estimation pratique des landmarks, nous résumons les propriétés des deux hyperparamètres qui ont été introduits pour la procédure de bagging et expliquons comment ceux-ci pourraient être choisis à partir des données :

- *Choix de σ_*^2 pour la procédure de bagging* : nous pensons qu'il n'est pas évident de trouver une méthode automatique pour sélectionner le niveau de bruit qui devrait être injecté dans les données. Pour étudier de façon précise les propriétés de la procédure de bagging, il faut tout d'abord définir une mesure de l'erreur de l'estimation des zero-crossings et des maxima d'ondelettes d'un signal. Un exemple de distance entre deux fonctions basée sur la position de leurs zero-crossings peut être trouvée dans Mallat [62]. Celle-ci pourrait être utilisée pour étudier la procédure de bagging d'un point de vue théorique. Notons que cette approche pourrait également servir pour l'étude de l'équivalence entre les modèles de régression nonparamétrique et de bruit blanc (voir section 3.3) pour le problème de l'estimation des landmarks d'un signal.
- *Le seuil de masse minimum (m_0)* : le choix de m_0 est directement relié au problème de l'identification des pics des intensités structurelles qui correspondent à des structures significatives dans un signal. Une procédure pour automatiquement sélectionner m_0 peut certainement être dérivée des approches du type "excess mass" dans le contexte de l'estimation de densité (voir Fisher et Marron [38] pour plus de références sur ce sujet).



(a)

FIG. 3.15: Intensité structurelle $G(x)$ avec de nombreux modes. Les aires grisées montrent la quantité de masse E_j attribuée à chaque mode. Chaque niveau inférieur λ_j est déterminé par les extrema locaux de $G(x)$. Les valeurs de v_1, v_2, \dots dénotent les positions des maxima successifs de $G(x)$ et w_1, w_2, \dots dénotent les positions des minima successifs.

$\text{argmax } \hat{G}_m^{(Bag)}(x), r = 2$	0.0723 0.7266	0.1367 0.8730	0.2363 0.9941	0.3184	0.3770	0.6133
Masse E_j	0.0002 0.2196	0.0348 0.0628	0.0018 0.0006	0.1029	0.0024	0.0567
$\text{argmax } \hat{G}_m^{(Bag)}(x), r = 1$	0.0059 0.4531 0.7520	0.0547 0.4727 0.7910	0.2246 0.5098 0.8359	0.2734 0.6602 0.8770	0.3008 0.6992 0.9453	0.4180 0.7227 0.9922
Masse E_j	0.0001 0.0014 0.0636	0.0570 0.0033 0.0178	0.0250 0.0178 0.0052	0.0002 0.0004 0.0004	0.3405 0.0013 0.0292	0.0005 0.0285 0.0109

TAB. 3.1: Positions des modes de $\hat{G}_m^{(Bag)}(x)$ pour $r = 2$ et $r = 1$ avec leurs masses associées E_j .

3.7.4 Résultats des simulations

Pour illustrer les performances de notre méthode, des données (y_i, x_i) ont été générées à partir du modèle (3.3.27), où x_i sont équidistants sur $[0, 1]$, $x_i = \frac{i}{n}$. Les facteurs sont :

- Les tailles de l'échantillon n .
- Les valeurs de σ^2

Pour chaque combinaisons de ces niveaux de facteurs, la simulation des données est répétée 100 fois en gardant tous les niveaux des facteurs constants, sauf les $\{\epsilon_i\}$ qui sont re-simulés à chaque étape. Dans ce chapitre, nous reportons les résultats pour deux tailles d'échantillon $n = 128$ (une taille modérée) et $n = 512$ (une taille importante) et un RSNR égal à 7 (un niveau de bruit faible), 5 (un niveau de bruit moyen) et 3 (un niveau de bruit important). Pour chaque séquence de données simulées, les lignes de zero-crossings de f_T ont été estimées pour $r = 2$ alors que les zero-crossings de f_W ont été estimés pour $r = 1$. Les lignes de maxima d'ondelettes de f_H ont été estimées pour $r = 1$ et $r = 2$, alors que les lignes de maxima d'ondelettes de f_3 ont été estimées pour $r = 2$. Les vraies lignes de zero-crossings et de maxima d'ondelettes des différentes fonctions de test que nous souhaitons estimer sont représentées dans la Figure 3.9 et la Figure 3.11.

Dans les Figures 3.16-3.17, nous donnons les positions des modes des intensités structurelles des lignes de zero-crossings estimées pour chaque combinaison des niveaux de facteurs et pour chaque séquence de données simulées $Y_j = \{y_{j,1}, \dots, y_{j,n}\}$, $j = 1, \dots, 100$. Dans chacune de ces Figures, les axes horizontaux et verticaux donnent respectivement la position des landmarks sur $[0, 1]$ et j (i.e. l'indice de la simulation), et de plus :

- les 3 Figures (a),(b),(c) de la première ligne donnent les positions des modes des intensités structurelles **sans bagging**, pour $n = 128$ et pour différentes valeurs de σ^2 (à gauche : $RNSR = 7$, au centre : $RSNR = 5$, à droite

$RNSR = 3$).

- les 3 Figures (d),(e),(f) de la deuxième ligne donnent les positions des modes des intensités structurelles **sans bagging**, pour $n = 512$ et pour différentes valeurs de σ^2 (à gauche : $RNSR = 7$, au centre : $RNSR = 5$, à droite $RNSR = 3$).

Les résultats de l'estimation pour f_T sont très satisfaisants (voir Figure 3.16) puisque pour chaque simulation, les modes des intensités structurelles sont localisés au niveau des 5 points d'inflexion de la fonction Time Shifted Sine f_T , sauf pour $n = 128$ et $RNSR = 3$ (seul le point d'inflexion en $x = 0.5$ est correctement détecté à chaque simulation). Rappelons que l'importance d'une structure dans un signal est liée aux longueurs des lignes de zero-crossings de maxima d'ondelettes qui convergent vers celle-ci. Ainsi, on s'attend à ce que notre méthode estime plus facilement les landmarks qui correspondent aux lignes les plus longues dans le plan temps-échelle. Ceci est le cas pour l'estimation des zero-crossings de la fonction Wave f_W pour $n = 128$ (voir Figures 3.17abc), où notre procédure détecte principalement les landmarks aux points $x = 0.25$, $x = 0.5$ et $x = 0.75$ qui correspondent aux limites des trois plus longues lignes du signal f_W (voir Figure 3.17). A partir des Figures 3.17def, on peut constater qu'augmenter la taille de l'échantillon permet l'estimation de structures qui n'étaient pas détectées avec une taille d'échantillon plus modeste (par exemple tous les extrema locaux de f_W sont détectés dans les Figures 3.17de, ce qui n'est pas le cas dans la Figure 3.17ab). D'après ces simulations, il apparaît qu'aucun extremum local ou point d'inflexion erroné n'a été détecté puisqu'à chaque simulation les modes des intensités structurelles sont localisés autour des landmarks de f_H et f_T . Ainsi, la procédure de bagging n'a pas été appliquée à ces exemples simulés. Ceci suggère que l'approche empirique, introduite dans cette section, pour estimer les zero-crossings, est un outil efficace pour supprimer les estimations erronées.

Dans les Figures 3.18-3.23, nous donnons les positions des modes des intensités structurelles pondérées des maxima d'ondelettes estimés de f_H et f_3 pour chaque combinaison des niveaux des facteurs et pour chaque séquence de données simulées $Y_j = \{y_{j,1}, \dots, y_{j,n}\}$, $j = 1, \dots, 100$. Dans chacune de ces Figures, les axes horizontaux et verticaux donnent respectivement la position des landmarks sur $[0, 1]$ et j (i.e. l'indice de la simulation), et de plus :

- les 3 Figures (a),(b),(c) de la première ligne donnent les positions des modes des intensités structurelles **sans bagging**, pour différentes valeurs de σ^2 (à gauche : $RNSR = 7$, au centre : $RNSR = 5$, à droite $RNSR = 3$).
- les 3 Figures (d),(e),(f) de la deuxième ligne donnent les positions des modes des intensités structurelles **sans bagging** dont la masse excède m_0 , pour différentes valeurs de σ^2 (à gauche : $RNSR = 7$, au centre : $RNSR = 5$, à droite $RNSR = 3$).
- les 3 Figures (g),(h),(i) de la troisième ligne donnent les positions des modes des intensités structurelles **après bagging** dont la masse excède m_0 , pour

différentes valeurs de σ^2 (à gauche : $RNSR = 7$, au centre : $RNSR = 5$, à droite $RNSR = 3$).

La fonction f_H a deux discontinuités aux points $x = 0.3$ et $x = 0.72$. Les premières lignes des Figures 3.18-3.21 montrent que de nombreux landmarks erronés sont détectés. Cependant, d'après les deuxièmes lignes des Figures 3.18-3.21, on peut constater que la plupart de ces estimations erronées peuvent être éliminées en rejetant les modes mineurs des intensités structurelles obtenues sans bagging (avec $m_0 = 0.06$). Ainsi, il serait très intéressant de trouver une méthode pour automatiquement sélectionner la masse minimum m_0 afin de décider si un pic d'une intensité structurelle correspond à la présence d'un landmark significatif. L'estimation de la discontinuité au point $x = 0.3$ est très satisfaisante pour chaque combinaison des niveaux de facteurs avec $r = 1$. Le point de discontinuité en $x = 0.72$ est plus difficile à détecter particulièrement avec une taille d'échantillon modeste ($n = 128$) ou avec $RNSR = 3$. De plus, on peut voir que le point de discontinuité en $x = 0.72$ est plus facile à détecter avec $r = 2$. Ceci suggère que la précision de la détection des discontinuités dans un signal est meilleure avec $r = 2$. Cependant, lorsque l'on augmente le nombre de moments nuls, on augmente également le nombre de lignes de maxima d'ondelettes qui convergent vers une singularité. Ainsi, si r est trop grand, des lignes de maxima d'ondelettes peuvent avoir de très faibles amplitudes et seront donc uniquement détectées à des échelles grossières où elles ne sont pas localisées suffisamment près de la singularité vers laquelle elles convergent.

La fonction de test f_3 a 3 singularités au niveau des points $x = 0.2$, $x = 0.4$ et $x = 0.7$ d'ordre $4/5$, $2/5$ et $7/5$. Ainsi, l'estimation des maxima d'ondelettes avec $r = 2$ devrait révéler la présence de ces 3 singularités et la précision de la détection devrait être inversement proportionnelle à l'ordre de ces dernières. Dans la Figure 3.22, nous observons que la procédure détecte principalement les singularités aux points $x = 0.2$ et $x = 0.7$, et les lignes de maxima d'ondelettes qui convergent vers $x = 0.59$ (voir Figure 3.11b). La singularité au point $x = 0.4$ est difficile à détecter avec une taille d'échantillon modeste puisque celle-ci a un “petit support”. Si l'on augmente la taille de l'échantillon, (voir Figure 3.23), la précision de la détection des 3 singularités est satisfaisante avec $RNSR = 7$ et $RNSR = 5$, mais pas avec $RNSR = 3$. Nous pouvons également observer qu'avec $n = 128$, l'élimination des modes mineurs (deuxième ligne de la Figure 3.22) permet l'élimination de la plupart des estimations erronées.

D'après ces simulations, il apparaît également que la procédure de bagging peut réduire la variance des estimateurs de manière significative, et est aussi un moyen de supprimer certains landmarks erronés dus aux fluctuations du bruit. Nous pensons fortement que cette procédure pourrait être améliorée avec des choix automatiques pour σ_*^2 et la masse minimale m_0 . En particulier, la quantité de bruit ajoutée dans les données devrait être inversement proportionnelle au rapport signal-sur-bruit.

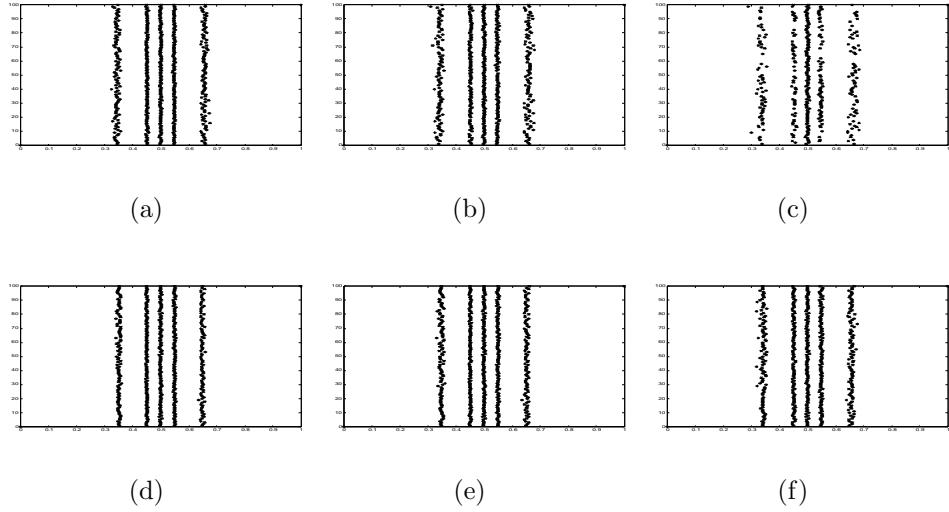


FIG. 3.16: Positions des modes de l'intensité structurelle des zero-crossings estimés de f_T : Première ligne $r = 2$ et $n = 128$: (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne $r = 2$ et $n = 512$: (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$.

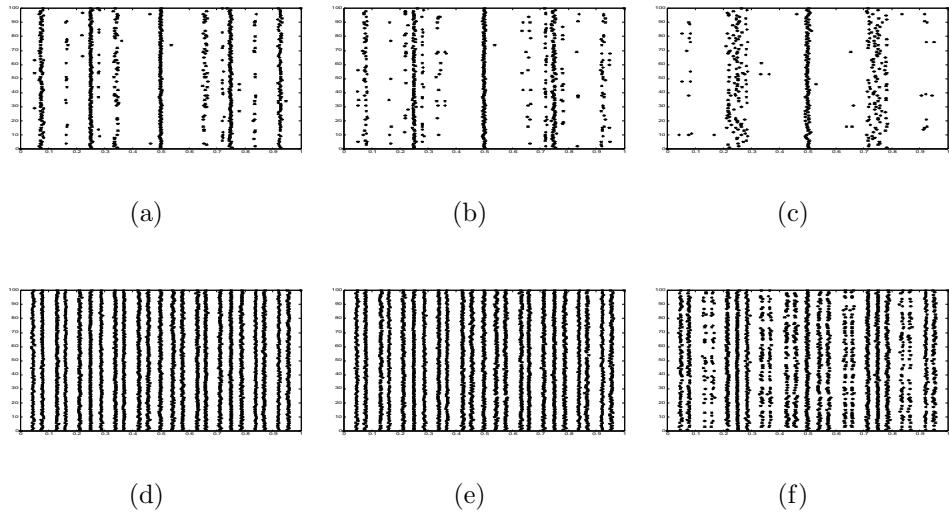


FIG. 3.17: Positions des modes de l'intensité structurelle des zero-crossings estimés de f_W : Première ligne $r = 1$ et $n = 128$: (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne $r = 1$ et $n = 512$: (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$.

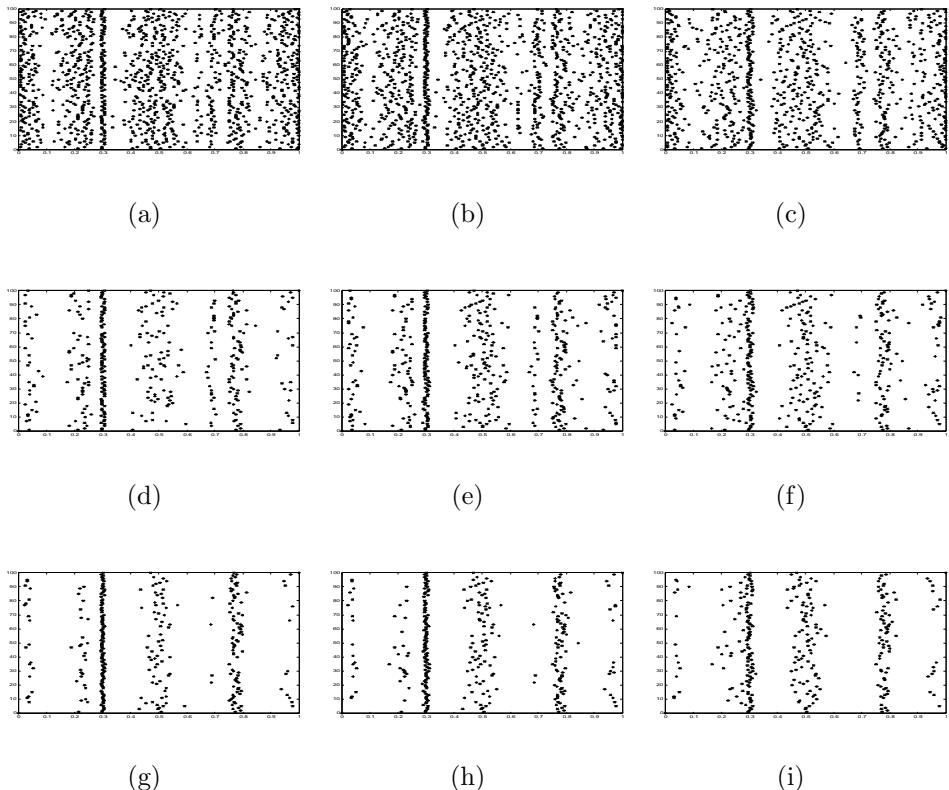


FIG. 3.18: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_H pour $r = 1$ et $n = 128$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.06$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.06$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

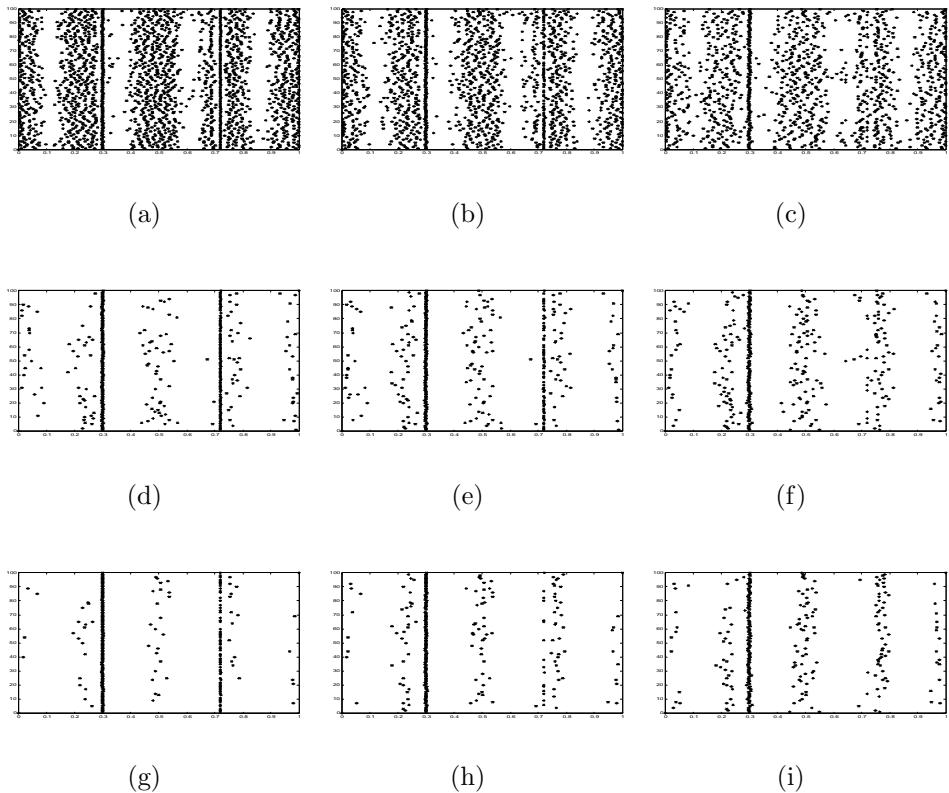


FIG. 3.19: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_H pour $r = 1$ et $n = 512$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.06$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.06$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

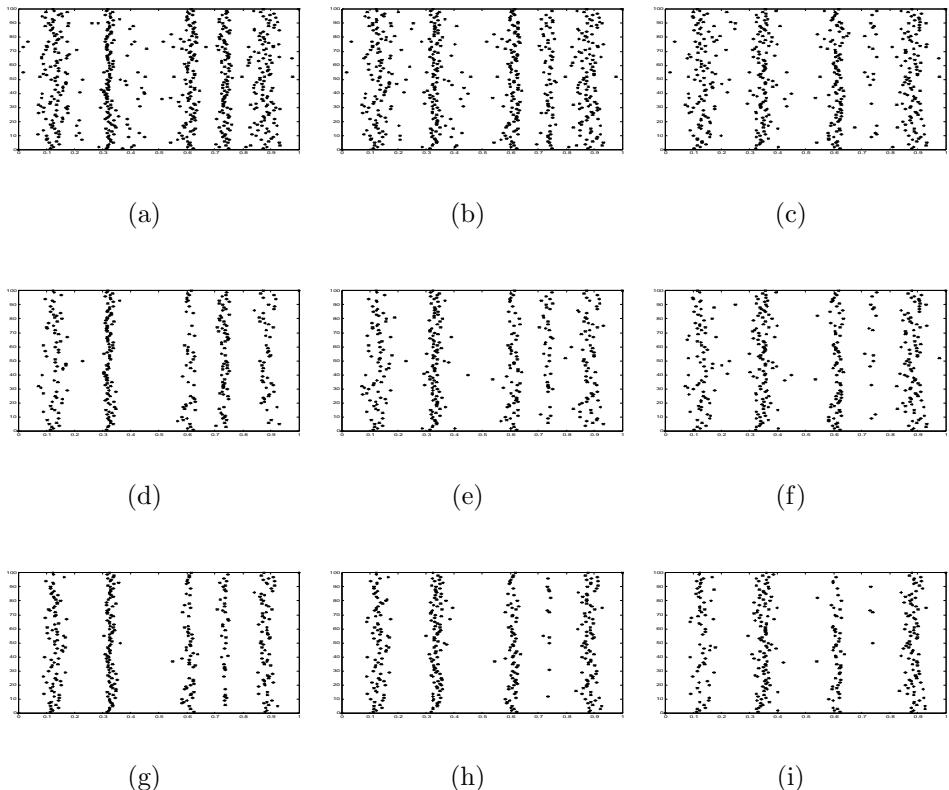


FIG. 3.20: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_H pour $r = 2$ et $n = 128$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.06$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.06$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

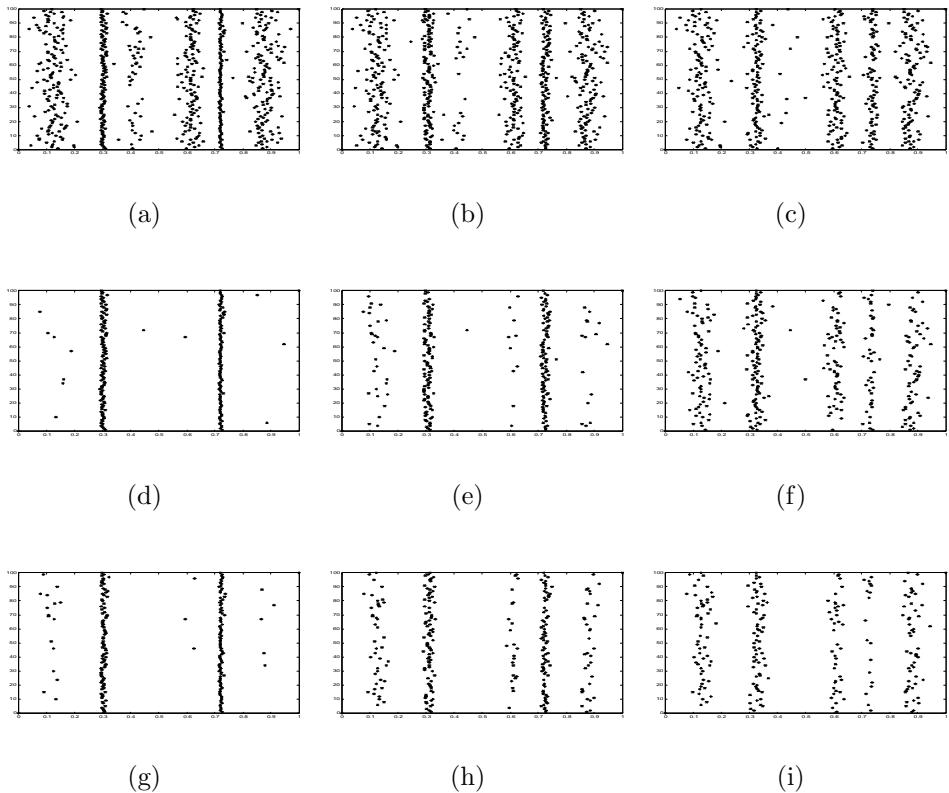


FIG. 3.21: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_H pour $r = 2$ et $n = 512$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.06$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.06$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

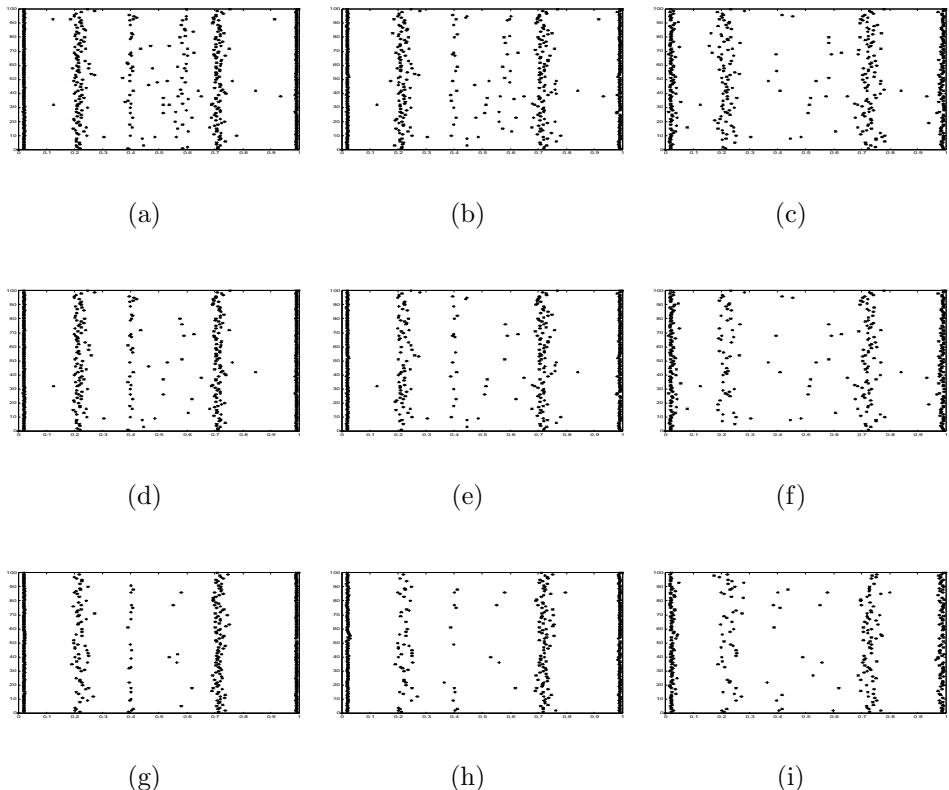


FIG. 3.22: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_3 pour $r = 2$ et $n = 128$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.01$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.01$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

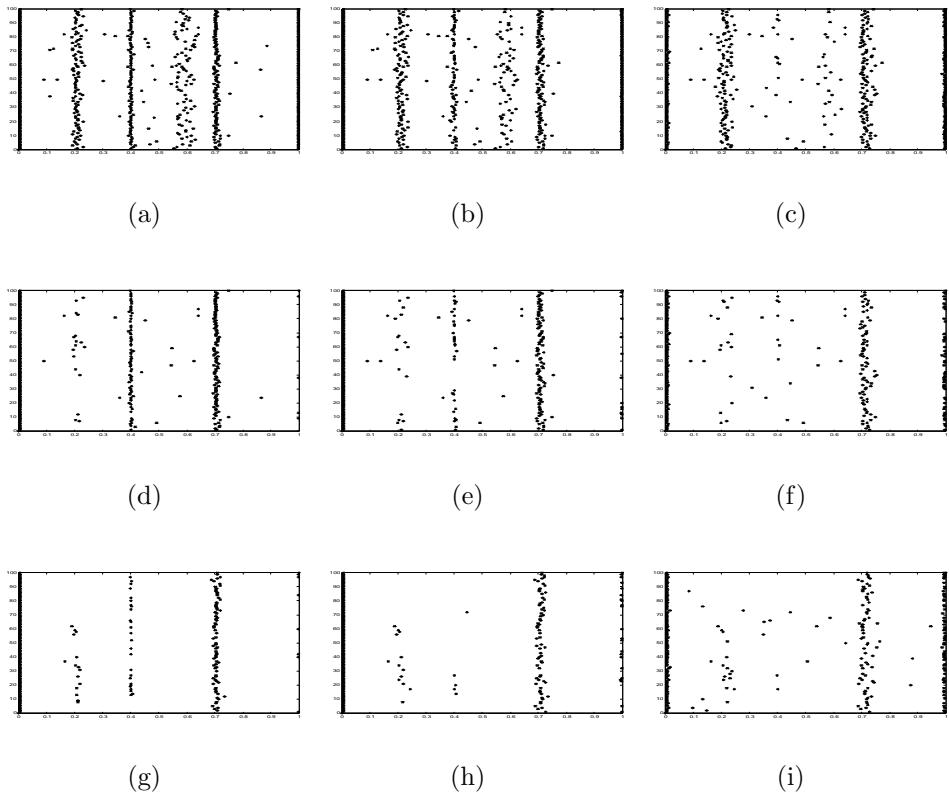


FIG. 3.23: Positions des modes de l'intensité structurelle pondérée des maxima d'ondelettes estimés de f_3 pour $r = 2$ et $n = 512$: Première ligne (sans bagging et sans élimination des modes mineurs) : (a) $RSNR = 7$, (b) $RSNR = 5$, (c) $RSNR = 3$. Deuxième ligne (sans bagging mais avec élimination des modes mineurs, $m_0 = 0.01$) : (d) $RSNR = 7$, (e) $RSNR = 5$, (f) $RSNR = 3$ Troisième ligne (avec bagging et élimination des modes mineurs, $m_0 = 0.01$) : (g) $RSNR = 7$, (h) $RSNR = 5$, (i) $RSNR = 3$

3.7.5 Quelques exemples réels

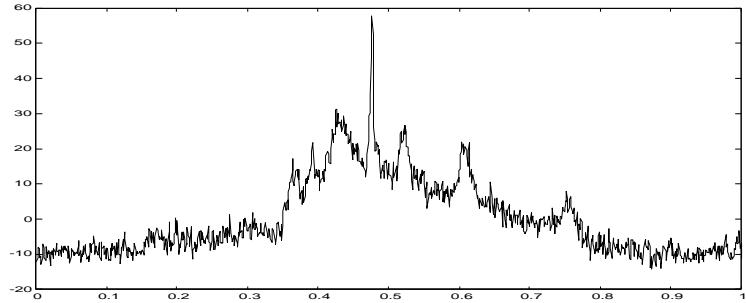
Nous proposons d'illustrer notre méthode sur quelques exemples réels.

3.7.5.1 Détection de transitions brutales

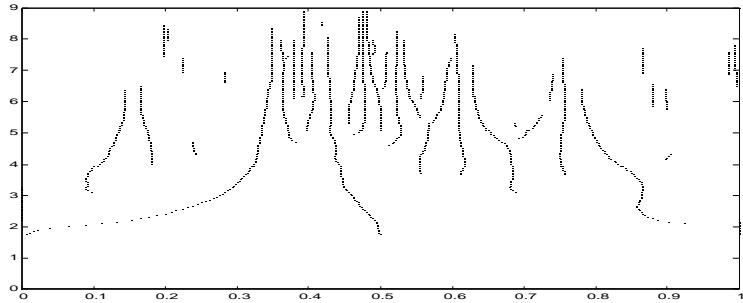
Nous allons montrer que la procédure de détection des maxima d'ondelettes que nous avons décrite précédemment peut être utilisée pour détecter des transitions brutales dans un signal, qui peuvent correspondre à la présence d'un saut ou d'un pic de très forte intensité pendant un intervalle de temps très court. Nous avons appliqué notre méthode à 3 signaux : RaphaelNMR, HochNMR et ESCA qui sont produits par la fonction `ReadSignal` de la boîte à outils *WaveLab* qui permet l'utilisation pratique, sous MATLAB, des méthodes de décomposition à base d'ondelettes de signaux 1D ou 2D (voir Buckheit *et. al.* [12]). *WaveLab* peut être téléchargé à l'URL : <http://www-stat.stanford.edu/wavelab>. Pour chacun de ces signaux, les lignes de maxima d'ondelettes ont été estimées pour une ondelette B-Spline d'ordre $q = 4$ avec $r = 2$ moments nuls, et nous avons calculé les intensités structurelles pondérées associées. Le résultat de l'estimation pour ces différents signaux est donné dans les Figures 3.24, 3.25 et 3.26. Dans ces trois Figures, nous pouvons constater que les modes des intensités structurelles de forte amplitude sont localisés au niveau des transitions brutales des signaux. De plus, bien que de nombreuses lignes de maxima d'ondelettes soient détectées, la pondération des intensités structurelles permet de mettre en évidence les transitions brutales dans un signal.

3.7.5.2 Détection de maxima locaux significatifs

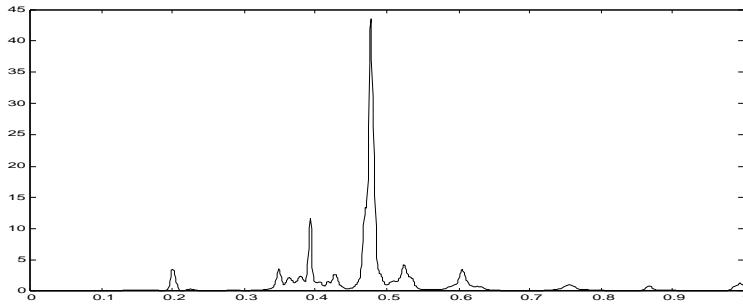
Les deux courbes représentées dans la Figure 4.1a sont des profils de densité de la ganglioside dans la région du cerveau Medulla pour deux vieux rats (le bruit de fond a été supprimé). Ces deux courbes sont extraites de l'ensemble des profils de densité analysés dans le papier de Munoz Maldonado *et. al.* [71], dans lequel les auteurs proposent une méthode pour comparer deux ensembles de courbes, et à titre d'application, comparent la distribution de la ganglioside dans les tissus du cerveau entre des jeunes et des vieux rats. Une description détaillée de ces données peut être trouvée dans le chapitre 4. Notons que les deux courbes g_1 et g_2 données dans la Figure 4.1a ne sont pas observées sur $[0, 1]$ mais sur $[a_1, b_1]$ et $[a_2, b_2]$ respectivement, avec $a_1 = 42.08, b_1 = 85.75, a_2 = 25.81$ et $b_2 = 86.55$. Cependant, en définissant $g_1^* = g_1(a_1 + t(b_1 - a_1))$ et $g_2^* = g_2(a_2 + t(b_2 - a_2))$ pour $t \in [0, 1]$, le problème de l'identification des landmarks significatifs de g_1 et g_2 est équivalent au problème de la détermination des landmarks de g_1^* et g_2^* sur $[0, 1]$. Les nombres de points d'échantillonnage pour les courbes g_1 et g_2 sont respectivement $n_1 = 110$ et $n_2 = 152$. Nous avons décidé d'interpoler ces données avec des Splines cubiques pour former des suites d'observations de $n = 128$ points, et de les recaler ensuite linéairement sur l'intervalle $[0, 1]$ (voir Figure 4.1b). Nous avons pris n égal à une puissance de deux afin de pouvoir estimer la variance avec MAD. Le nombre de points est choisi proche de n_1 et n_2 afin d'éviter de lisser les fonctions, de sorte que



(a)

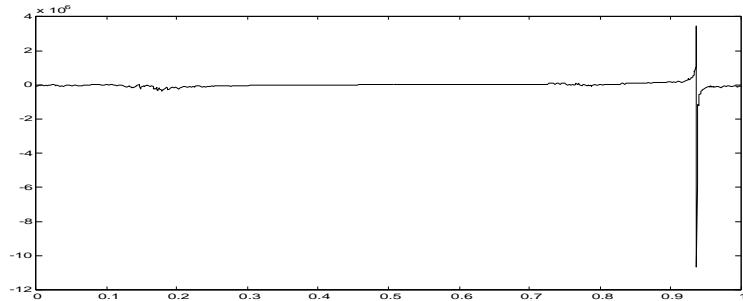


(b)

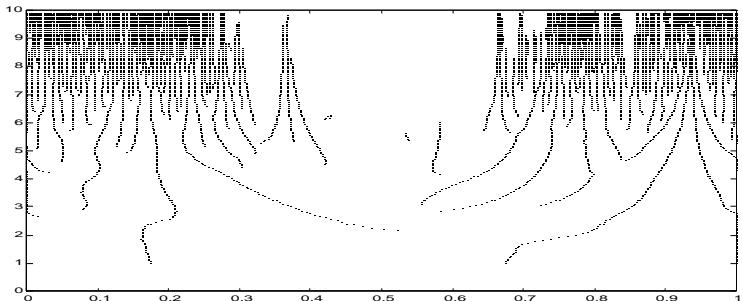


(c)

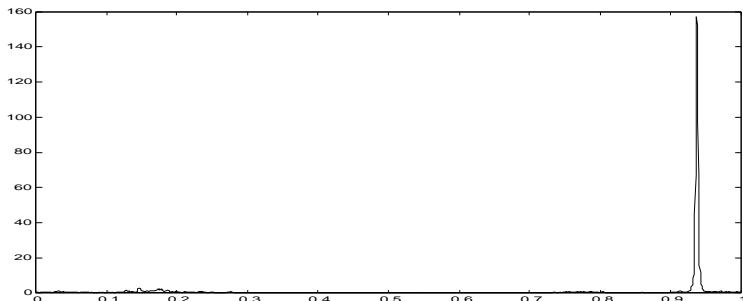
FIG. 3.24: (a) Signal RaphaelNMR (b) Maxima d'ondelettes estimés pour une ondelette B-Spline d'ordre $q = 4$ et $r = 2$. (c) Intensité structurelle pondérée des maxima d'ondelettes estimés



(a)

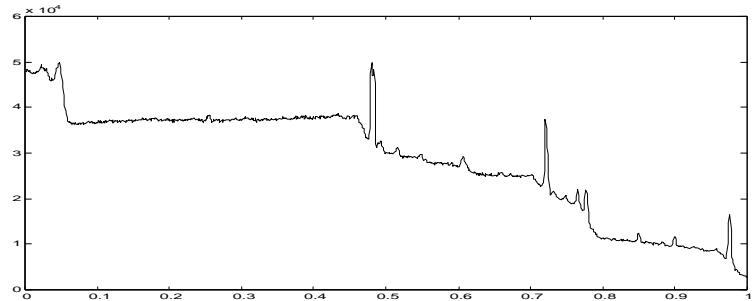


(b)

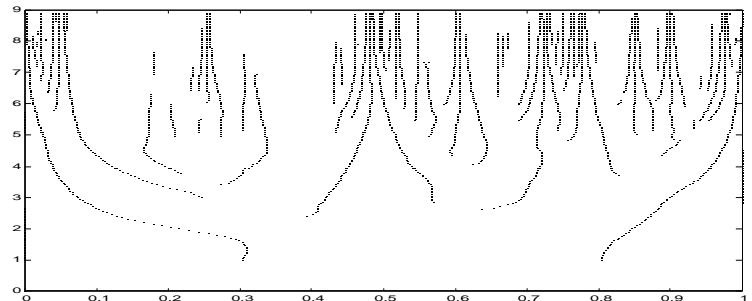


(c)

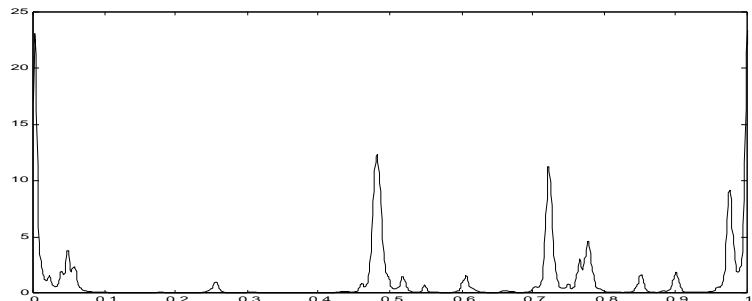
FIG. 3.25: (a) Signal HochNMR (b) Maxima d'ondelettes estimés pour une ondelette B-Spline d'ordre $q = 4$ et $r = 2$. (c) Intensité structurelle pondérée des maxima d'ondelettes estimés



(a)



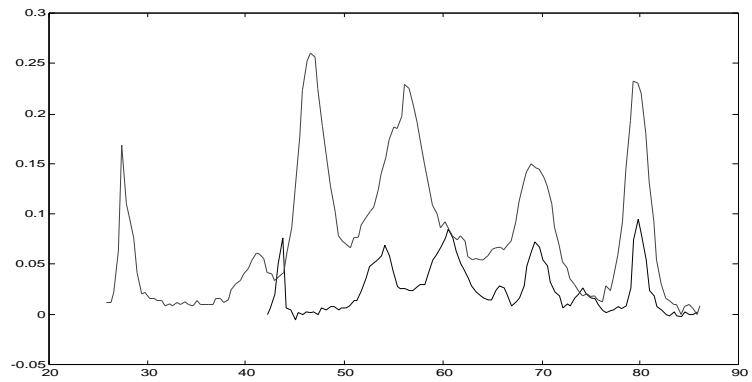
(b)



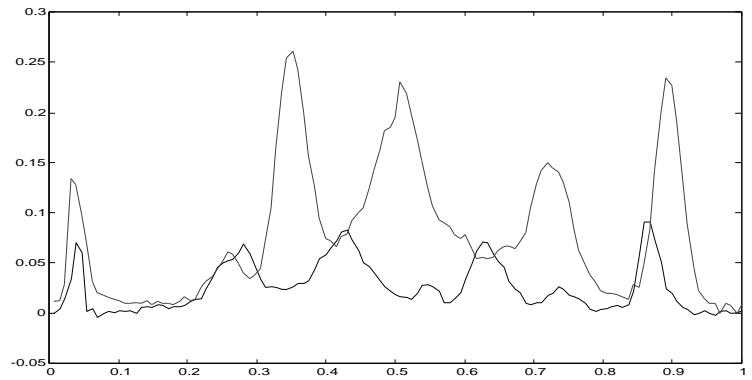
(c)

FIG. 3.26: (a) Signal ESCA (b) Maxima d'ondelettes estimés pour une ondelette B-Spline d'ordre $q = 4$ et $r = 2$. (c) Intensité structurelle pondérée des maxima d'ondelettes estimés

les courbes interpolantes \tilde{g}_1 et \tilde{g}_2 correspondent à la forme des données originales (comparer les Figures 4.1a-4.1b). Dans [71], cinq pics principaux sont identifiés dans chaque profil de densité afin de les aligner. Nous proposons donc d'utiliser notre méthode pour déterminer les maxima significatifs de \tilde{g}_1 et \tilde{g}_2 . Nous avons pris une ondelette B-Spline d'ordre $q = 4$ avec $r = 1$ moment nul. Les lignes de zero-crossings qui correspondent à des maxima locaux sont données dans la Figure 4.2 avec les intensités structurelles correspondantes. Le résultat est très satisfaisant puisque la procédure détecte les cinq pics principaux de \tilde{g}_1 et \tilde{g}_2 respectivement localisés aux points $x = 0.04, 0.28, 0.42, 0.63, 0.86$ et $x = 0.04, 0.35, 0.51, 0.72, 0.89$. De plus, elle détecte également deux pics plus petits de \tilde{g}_1 aux points $x = 0.54, 0.73$ et le petit pic de \tilde{g}_2 au point $x = 0.26$. Les autres maxima locaux dus aux fluctuations du bruit, ne sont pas pris en compte.



(a)



(b)

FIG. 3.27: (a) Profils de densité de deux vieux rats dans la région MD (g_1 : trait plein et g_2 : trait pointillé) (b) Profils de densité de deux vieux rats dans la région MD interpolés avec des Splines cubiques ($n = 128$) et recalés (linéairement) sur l'intervalle $[0, 1]$ (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé)

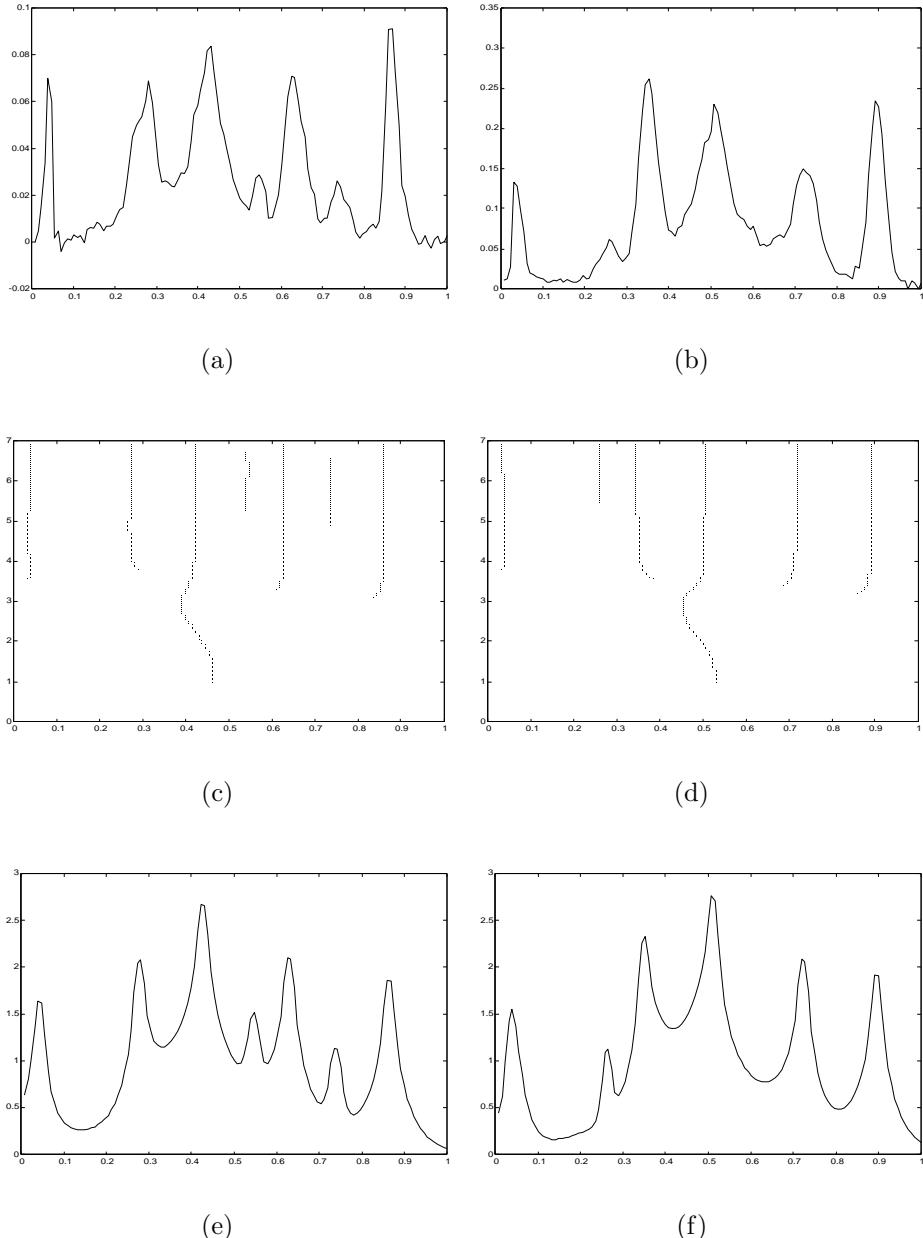


FIG. 3.28: (a) Courbe échantillonnée \tilde{g}_1 . (c) Estimation des zero-crossings de \tilde{g}_1 qui correspondent à des maxima locaux avec l'intensité structurelle associée (e). (b) Courbe échantillonnée \tilde{g}_2 . (d) Estimation des zero-crossings de \tilde{g}_2 qui correspondent à des maxima locaux avec l'intensité structurelle associée (f).

Chapitre 4

Recalage automatique de courbes

Une fois que les landmarks de deux ou plusieurs signaux ont été estimés, les transformations qui alignent ces ensembles de landmarks doivent être calculées. Pour mettre en correspondance deux ensembles de landmarks, il suffit de savoir quelles paires de landmarks doivent être associés. En effet dans le chapitre 2, nous avons vu que dans ce cas le calcul de la déformation qui aligne deux ensembles de landmarks revient à un problème de régression sous contrainte monotone et que de nombreuses méthodes sont disponibles pour le résoudre. Ainsi, tout dépend de la détermination des paires de landmarks qui doivent être associés. En pratique, cette tâche n'est généralement pas évidente : de nombreux landmarks dans un ensemble peuvent ne pas avoir d'homologue dans l'autre ensemble, et la présence de landmarks "erronés" dus aux fluctuations du bruit complique également le problème. Dans ce chapitre, nous proposons une méthode automatique pour aligner deux courbes, éventuellement observées avec du bruit, à partir de leur landmarks. A notre connaissance, il n'existe pas de méthode automatique dans la littérature pour résoudre ce problème. Kneip et Gasser [51] ont proposé un cadre théorique pour l'étude de la consistance des estimateurs des déformations nécessaire au recalage d'un ensemble de courbes. Toutefois, leur approche nécessite en pratique une méthode heuristique qui est difficilement applicable pour de larges ensembles de données, comme nous allons le préciser par la suite.

Les performances de la méthode automatique développée dans ce chapitre sont comparées avec l'algorithme d'alignement global développé par Wang et Gasser ([91], [92]) pour le problème basique du recalage de deux fonctions. La méthodologie que nous proposons peut se généraliser à un ensemble de plusieurs courbes. Deux exemples réels servent d'illustration :

- la comparaison de la distribution de la ganglioside dans les tissus cérébraux de groupes de rats d'âges distincts (exemple 1D donné dans l'introduction, cet ensemble de courbes est analysé dans [71]). Cet ensemble de données sera appelé par la suite : données de la ganglioside.
- un exemple provenant de la physiologie que nous détaillons dans le paragraphe 4.2.2. Ces données ont été enregistrées par les Dr. Amarantini David et Dr.

Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Université de Grenoble, France). Leur étude est motivée par le besoin de mieux comprendre le processus génératrice de mouvement pour différents niveaux d'une force extérieure appliquée sur le genou de plusieurs patients. Cet ensemble de données sera appelé par la suite : données d'orthèse.

L'*analyse de variance* (ANOVA) est certainement l'une des techniques d'analyse de données les plus utilisées en statistique appliquée. Cette approche vise à déterminer comment les variations d'une variable observée peuvent être expliquées par l'observation d'autres quantités. Cette technique, très utile pour l'analyse de données de faibles dimensions, est insuffisante lorsque les variables aléatoires observées sont *fonctionnelles* (par exemple des courbes de croissance chez différents individus). Bien qu'une fonction ne soit jamais observée dans son ensemble, un nombre suffisamment large d'évaluations de celle-ci en différents points est généralement disponible. La taille importante des données observées, due aux appareils modernes d'acquisition, fait que les techniques d'ANOVA multi-dimensionnelle ne sont plus adaptées. Pour la comparaison d'ensemble multiples de signaux, les méthodes d'*analyse de variance fonctionnelles* (FANOVA) sont des alternatives préférables aux techniques d'ANOVA classiques, qui permettent toujours une interprétation simple des résultats. Etant donné le nombre important d'applications nécessitant l'analyse de données fonctionnelles, les modèles de FANOVA ont récemment gagné en popularité et il existe une littérature abondante sur ce sujet. Une revue détaillée peut être trouvée dans le livre de Ramsay et Silverman [77].

Abramovich *et. al.* ont proposé récemment une méthode de FANOVA avec ondelettes qui permet de tester de nombreuses hypothèses pour quantifier les variations au sein d'une population. A titre d'application, Abramovich *et. al.* ont analysé l'ensemble de données d'orthèse. Toutefois, cette analyse a été réalisée sans recalage préalable des données observées. Nous avons donc repris l'étude conduite dans [1] en recalant au préalable les courbes enregistrées, afin d'illustrer l'intérêt de l'alignement de courbes. La méthode de FANOVA avec ondelettes proposé par Abramovich *et. al.* nous a également permis d'analyser les variations au sein des données de la ganglioside. Les résultats obtenus sont comparés avec les conclusions de l'étude conduite dans [71].

4.1 Mise en correspondance automatique de deux ensembles de landmarks

4.1.1 Approches déjà existantes

Généralement, la détermination des landmarks qui doivent se correspondre se fait manuellement, en essayant de repérer le nombre de landmarks qui sont communs à toutes les courbes. Par exemple, pour certains ensembles de données, il paraît clair que les courbes observées ont le même nombre de maxima "principaux"

(voir Munoz Maldonado *et. al.* [71]). Gasser et Kneip [39] ont proposé une méthode pour déterminer les points caractéristiques qui sont communs à un ensemble de m courbes. Supposons que chacune de ces courbes possède exactement un maxima de taille variable dont la position se trouve au voisinage d'un point x_0 . Après avoir obtenu les positions individuelles de tous ces maxima, Gasser et Kneip [39] proposent de calculer la densité de toutes ces positions. La densité obtenue possède alors exactement un mode localisé autour du point x_0 . L'heuristique proposée dans [39] s'appuie sur le fait que les différences *inter-individus* entre les positions des points caractéristiques communs sont relativement petites par rapport aux différences entre les positions des différents landmarks chez un seul individu (ce qui est souvent observé en pratique). Soit f_1, \dots, f_m un ensemble de courbes éventuellement observées avec du bruit, Gasser et Kneip [39] ont alors suggéré la méthodologie suivante pour identifier les landmarks communs de même type (par exemple des maxima) :

- Tout d'abord, les maxima locaux de chacune des courbes $f_i, i = 1, \dots, m$ sont estimés. Dans [39], cette étape est réalisée en calculant les zéros appropriés de la dérivée première des estimateurs à noyau de chacune des courbes.
- Toutes les positions de ces maxima locaux sont triées dans un seul tableau. Une méthode d'estimation de densité par noyau est ensuite appliquée au vecteur obtenu.

Les maxima communs vont alors donner lieu à des pics dans la densité ainsi calculée. Cette méthode permet en particulier de définir de petits intervalles, sur l'axe des abscisses, où les maxima doivent être recherchés dans chacune des courbes. Kneip et Gasser [51] ont proposé un ensemble d'outils statistiques pour étudier l'alignement de courbes à partir de leurs landmarks. En particulier, ils ont introduit la notion de fonctionnelle structurelle, que nous avons utilisée dans le chapitre 3, pour décrire les points structurels d'un signal. Une fonctionnelle structurelle pour décrire les extrema d'une fonction $g : [0, 1] \rightarrow \mathbb{R}$ peut être définie de la manière suivante :

Définition 4.1 Soit α un réel n'appartenant pas à $[0, 1]$, et $[a, b]$ un intervalle inclus dans $]0, 1[$. Une fonctionnelle $T : C^1([0, 1]) \rightarrow]0, 1[\cup \{\alpha\}$ sera appelée une fonctionnelle structurelle du second ordre associée à $[a, b]$, si pour $g \in C^1([0, 1])$:

$$\begin{aligned} T(g) &= \zeta, \text{ si } g \text{ a exactement un extrema dans }]a, b[\text{ au point } \zeta, \\ T(g) &= \alpha, \text{ sinon, i.e. si } g \text{ a soit plusieurs extrema, soit aucun extremum dans }]a, b[. \end{aligned}$$

La donnée de plusieurs fonctionnelles structurelles associées à des intervalles disjoints permet de définir les landmarks d'un ensemble de courbes. $T(g) = \alpha$ signifie que le landmark correspondant à la fonctionnelle structurelle T est manquant, c'est à dire qu'il n'existe pas dans la courbe g . Supposons que l'on dispose d'un ensemble de m courbes f_1, \dots, f_m (sans bruit). A partir de la donnée d'un ensemble de fonctionnelles structurelles T_1, \dots, T_r , il est possible de définir à l'aide d'opérateurs de déformation (voir chapitre 2), des transformations u_1, \dots, u_m qui synchronisent les courbes f_1, \dots, f_m . En supposant que les "vraies" courbes f_1, \dots, f_m n'ont pas

de landmarks manquants pour chaque $T_i, i = 1, \dots, r$ (ce qui est une hypothèse très restrictive), Kneip et Gasser [51] ont étudié les propriétés asymptotiques des estimateurs des fonctions $u_1, \dots, u_m, f_1 \circ u_1, \dots, f_m \circ u_m$ et de la forme moyenne de ces courbes à l'aide d'estimateurs à noyau. En pratique, il est donc nécessaire de déterminer un ensemble de fonctionnelles structurelles, c'est à dire un ensemble de sous-intervalles de $[0, 1]$ dans lesquels rechercher les landmarks. La méthode heuristique décrite dans Gasser et Kneip [39] peut aider le statisticien à déterminer de tels intervalles en fonction de l'intensité et de l'"étalement" des modes de la densité obtenue. Toutefois, comme souligné dans [51], la définition de fonctionnelles structurelles à l'aide de cette heuristique, est délicate en pratique. Il peut y avoir des modes dus au bruit et certains landmarks importants des courbes peuvent ne pas être pris en compte. Enfin, cette approche est applicable lorsqu'un nombre suffisant de courbes est observé, et ne peut donc pas être utilisée pour le recalage de seulement deux signaux.

4.1.2 Une méthode automatique

Nous allons maintenant expliquer comment les intensités structurelles des lignes de zero-crossings de deux courbes f_1 et f_2 peuvent être utilisées pour construire une méthode pour décider automatiquement si un landmark de f_1 (resp. f_2) correspond à un landmark de f_2 (resp. f_1). Afin de présenter cette méthode, nous proposons d'illustrer sur un exemple réel, le problème de l'identification des paires de landmarks qui doivent se correspondre. Cet exemple a déjà servi d'illustration des méthodes de détection de landmarks présentées dans le chapitre 3. Les deux courbes tracées dans la Figure 4.1a sont des profils de densité dans la région Medulla pour deux vieux rats (le bruit de fond a été supprimé comme suggéré dans Munoz Maldonado *et. al.* [71]). Ces deux échantillons font partie des données de la ganglioside. Remarquons tout d'abord, que les courbes g_1 et g_2 de la Figure 4.1a ne sont pas observées sur $[0, 1]$ mais sur $[a_1, b_1]$ et $[a_2, b_2]$ respectivement, avec $a_1 = 42.08, b_1 = 85.75, a_2 = 25.81$ et $b_2 = 86.55$. Cependant, en définissant $\tilde{g}_1 = g_1(a_1 + t(b_1 - a_1))$ et $\tilde{g}_2 = g_2(a_2 + t(b_2 - a_2))$ pour $t \in [0, 1]$, le problème de l'alignement des courbes g_1 et g_2 est équivalent à celui du recalage des fonctions \tilde{g}_1 et \tilde{g}_2 qui sont définies sur $[0, 1]$. Le nombre de points pour les courbes g_1 et g_2 est respectivement $n_1 = 110$ et $n_2 = 152$. Nous avons choisi d'interpoler ces données à l'aide de Spline cubique pour former deux suites d'observations de $n = 128$ points, et ensuite de les aligner sur l'intervalle $[0, 1]$ à l'aide du recalage linéaire expliqué précédemment (voir Figure 4.1b). Nous avons pris n égal à une puissance de deux, afin de pouvoir utiliser facilement l'une des nombreuses méthodes de débruitage à base d'ondelettes disponible dans la littérature pour estimer les courbes \tilde{g}_1 et \tilde{g}_2 (voir e.g. Antoniadis *et. al.* [3] pour une revue détaillée de ces méthodes dont une implémentation sous Matlab peut être obtenue à l'URL : <http://www-lmc.imag.fr/SMS/software/GaussianWaveDen>). Le nombre de points n est choisi proche de n_1 et n_2 afin d'éviter de lisser les fonctions g_1 et g_2 . Ainsi, la forme des courbes échantillonées \tilde{g}_1 et \tilde{g}_2 correspond à celle des données originales (voir Figure 4.1b).

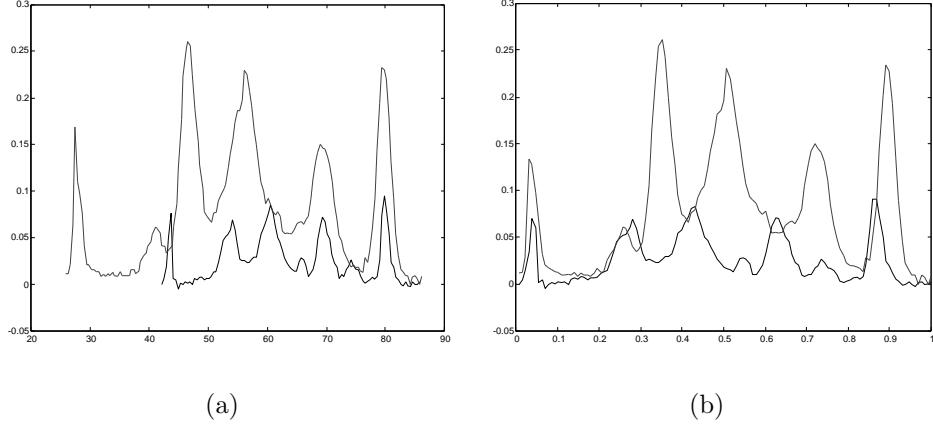


FIG. 4.1: (a) Profils de densité de deux vieux rats dans la région MD (g_1 : trait plein et g_2 : trait pointillé) (b) Profils de densité de deux vieux rats dans la région MD interpolés avec des Splines cubiques ($n = 128$) et recalés (linéairement) sur l'intervalle $[0, 1]$ (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé)

Supposons que pour aligner les deux fonctions \tilde{g}_1 et \tilde{g}_2 , on souhaite mettre en correspondance les extrema locaux significatifs de chacune des courbes. Pour déterminer la position de tels landmarks, nous avons utilisé la procédure décrite dans le chapitre 3 pour estimer les lignes de zero-crossings qui correspondent aux maxima locaux de \tilde{g}_1 et \tilde{g}_2 (calculées pour une ondelette B-Spline d'ordre $q = 4$ et $r = 1$ moment nul). Les lignes de zero-crossings estimées sont données dans la Figure 4.2 avec les intensités structurelles associées. Dans cet exemple, l'intensité structurelle \hat{G}_{z1}^+ dans la Figure 4.2e a sept modes, alors que l'intensité structurelle \hat{G}_{z2}^+ de la Figure 4.2f a six modes. A première vue, il n'est pas facile de décider si le i ème mode de \hat{G}_{z1}^+ doit correspondre au j ème mode de \hat{G}_{z2}^+ pour $i = 1, \dots, 7$ et $j = 1, \dots, 6$. Dans Munoz Maldonado *et. al.* [71], cinq pics principaux sont identifiés dans chaque profil de densité pour recaler les courbes. On peut effectivement clairement affirmer que les courbes \tilde{g}_1 et \tilde{g}_2 ont cinq maxima principaux qui se trouvent aux points $x = 0.04, 0.28, 0.42, 0.63, 0.86$ et $x = 0.04, 0.35, 0.51, 0.72, 0.89$, et que les formes de ces pics suggèrent que ces cinq landmarks doivent être alignés. Lorsque l'on regarde les lignes de zero-crossings qui convergent vers ces cinq pics, on peut constater que les formes des lignes qui correspondent au i ème maxima principal de \tilde{g}_1 et \tilde{g}_2 sont similaires pour $i = 1, \dots, 5$. Dans un certain sens, pour identifier les paires de landmarks qui doivent se correspondre, il serait naturel d'essayer d'associer les lignes de zero-crossings dont la forme et la longueur dans le plan temps-échelle sont similaires. Puisque la largeur et la hauteur des modes des intensités structurelles sont reliées à la forme des lignes de zero-crossings, nous proposons d'aligner \hat{G}_{z1}^+ et \hat{G}_{z2}^+ par *programmation dynamique temporelle* (DTW) pour mesurer la similarité entre ces lignes de manière automatique. Etant donné que le DTW est une technique qui

\hat{G}_{z1}^+ et \hat{G}_{z2}^+ soient alignés après transformation, ce qui indiquerait donc les paires de landmarks qui doivent se correspondre.

4.1.2.1 Alignement des intensités structurelles par DTW

La programmation dynamique temporelle (DTW), qui a été mise au point pour l’alignement d’une courbe par rapport à l’autre, est largement documentée dans la littérature d’ingénierie. L’article de Sakoe et Chiba [82] constitue une bonne référence pour quelques idées basiques sur ce sujet. En reconnaissance vocale, un mot ou une phrase peut être exprimé comme une suite de points caractéristiques obtenue à partir d’un enregistrement et d’une méthode appropriés. Ce vecteur de caractéristiques, plutôt que le signal audio original, est ensuite analysé. Le processus de reconnaissance consiste à comparer le mot enregistré avec d’autres mots dans un ensemble de référence. Si les caractéristiques du mot enregistré correspondent suffisamment à celles d’un des mots dans l’ensemble de référence, alors le mot (ou la phrase) est reconnu. Pour cette comparaison, les fluctuations de l’axe temporel entre le mot enregistré et un mot de référence doivent être éliminées. Une version discrète du problème de recalage de deux vecteurs de caractéristiques correspondant à des enregistrements audio, est donnée dans Sakoe et Chiba [82], et nous décrivons leur algorithme dans la section 4.1.3.3.

L’alignement de courbes bruitées par DTW, a été étudié par Wang et Gasser ([91],[92]) en utilisant des estimateurs à noyau. D’un point de vue continu, l’alignement de deux fonctions θ_1 et θ_2 par DTW peut être formulé comme un problème d’alignement global. Pour la fonction de coût quadratique classique, une formulation possible de ce problème est la suivante : soit $\theta_1, \theta_2 : [0, 1] \rightarrow \mathbb{R}$, aligner les deux courbes θ_1 et θ_2 , consiste à chercher le minimum, parmi tous les difféomorphismes croissants h de $[0, 1] \rightarrow [0, 1]$, de la fonctionnelle :

$$L_{\theta_1, \theta_2}(h) = \int_0^1 (\theta_1(x) - \theta_2 \circ h(x))^2 dx. \quad (4.4.1)$$

L’une des difficultés principales de cette approche est de garantir que la solution du problème variationnel (4.4.1) est une fonction croissante, et dérivable afin d’avoir une déformation suffisamment régulière. Piccioni *et. al.* [72] ont étudié en détail le problème variationnel associé à ce type de fonctionnelle. Pour étudier l’existence du minimum, les auteurs dans [72] suggèrent de reformuler ce problème variationnel sur l’ensemble de toutes les mesures de probabilités sur $[0, 1]$. Nous proposons de suivre leur méthodologie pour étudier le problème de l’alignement de deux fonctions par DTW où le but est d’aligner les modes communs de deux intensités structurelles. Pour ce faire, notons par M_1 l’ensemble de toutes les mesures de probabilités sur l’ensemble des Boréliens de $[0, 1]$. C’est un espace compact lorsqu’il est muni de la topologie de convergence faible. Pour un élément $\nu \in M_1$, nous noterons sa fonction

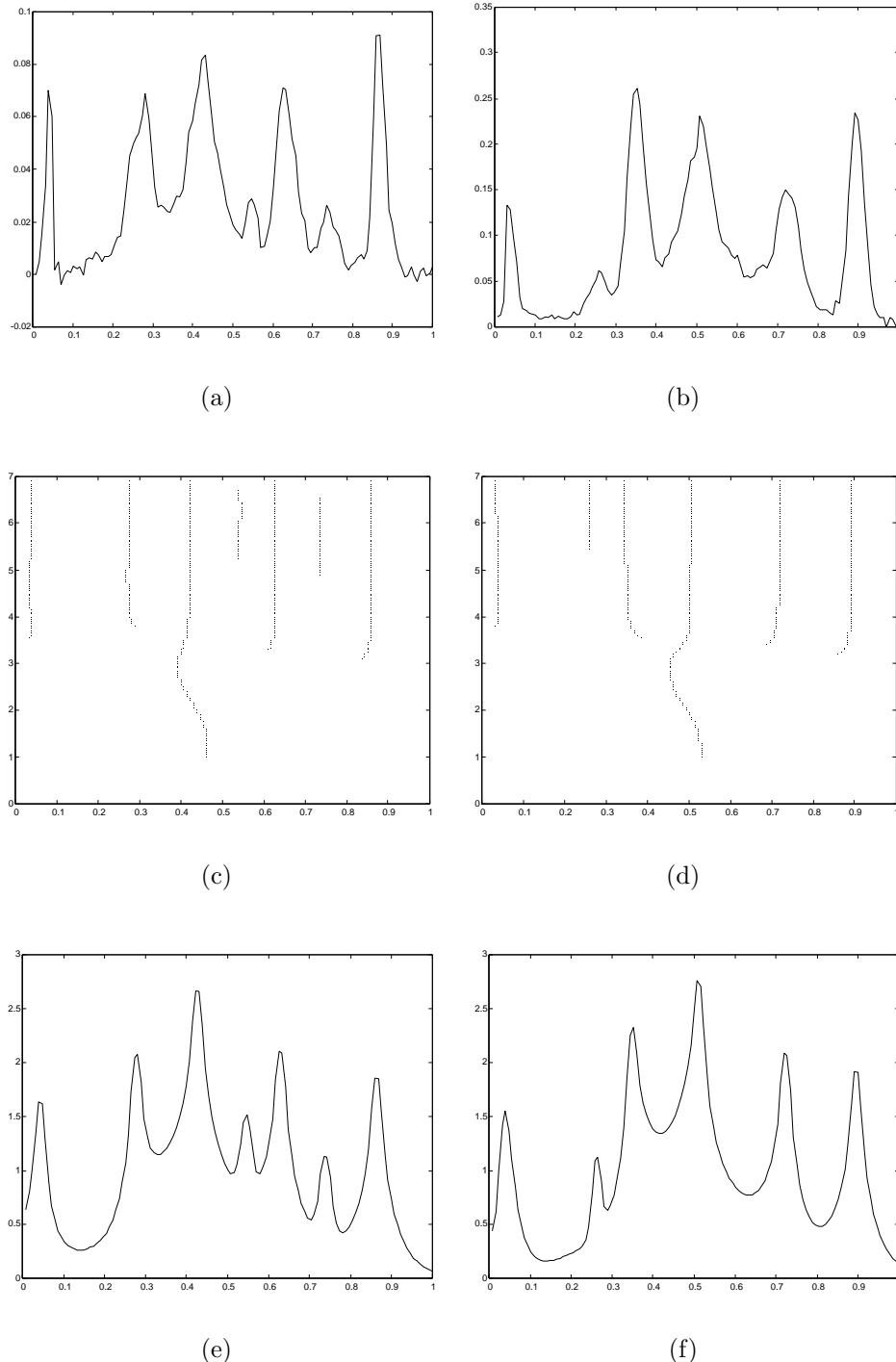


FIG. 4.2: (a) Courbe échantillonnée \tilde{g}_1 . (c) Estimation des zero-crossings de \tilde{g}_1 qui correspondent à des maxima locaux avec l'intensité structurelle associée (e). (b) Courbe échantillonnée \tilde{g}_2 . (d) Estimation des zero-crossings de \tilde{g}_2 qui correspondent à des maxima locaux avec l'intensité structurelle associée (f).

de répartition par :

$$F(\nu)(x) = \int_0^x d\nu, \quad x \in [0, 1].$$

Soit θ_1 et θ_2 deux fonctions bornées qui sont dans $L^2([0, 1], \mathbb{R}, dx)$, où dx représente la mesure de Lebesgue sur $[0, 1]$. Alors, pour $\nu \in M_1$ nous posons :

$$L_{\theta_1, \theta_2}(F(\nu)) = \int_0^1 \left(\frac{\theta_1(x)}{\|\theta_1\|_\infty} - \frac{\theta_2 \circ F(\nu)(x)}{\|\theta_2\|_\infty} \right)^2 dx. \quad (4.4.2)$$

Si θ_2 est continue sur $[0, 1]$, il peut être montré que $L_{\theta_1, \theta_2}(\cdot)$ est continue et a donc un minimum puisque M_1 est un compact (voir la preuve de la Proposition 4.2 dans [72]). La normalisation avec la norme-sup $\|\cdot\|_\infty$ dans (4.4.2) vise à réduire les différences en amplitude entre les courbes lors du calcul de la fonction qui minimise $L_{\theta_1, \theta_2}(\cdot)$. Puisque que la transformation de coût minimal tend à aligner les extrema locaux de θ_1 et θ_2 (ceci est expliqué dans Wang et Gasser [91],[92]), nous proposons d'aligner les intensités structurelles des lignes de zero-crossings de deux fonctions f_1 et f_2 possédant différents landmarks, pour décider quelles paires de points caractéristiques doivent être associés. Cette idée est formulée dans la proposition suivante, où une version légèrement modifiée de l'intensité structurelle est proposée afin d'obtenir des fonctions continues et bornées sur $[0, 1]$ (les raisons de cette modification sont données dans la remarque 4.1 qui suit cette proposition) :

Proposition 4.1 *Soit $f_1, f_2 \in L^2(\mathbb{R})$ et $\psi = (-1)^r \theta^{(r)}$, $r \geq 1$ une ondelette causale. Supposons qu'il existe p_1 lignes de zero-crossings $z_{1,i}(s)$, $i = 1, \dots, p_1$ et p_2 lignes de zero-crossings $z_{2,i}(s)$, $i = 1, \dots, p_2$ associées à f_1 et f_2 respectivement. Pour $x \in \mathbb{R}$, définissons l'intensité structurelle des zero-crossings $G_{z_1}(x)$ et $G_{z_2}(x)$ comme :*

$$\begin{aligned} G_{z_1}(x) &= \sum_{i=1}^{p_1} \int_0^{s_{z_{1,i}}} |\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right) ds \\ G_{z_2}(x) &= \sum_{i=1}^{p_2} \int_0^{s_{z_{2,i}}} |\log(s)|^3 \theta\left(\frac{x - z_{2,i}(s)}{s}\right) ds, \end{aligned}$$

où $[0, s_{z_{j,i}}]$ est le support de la ligne de zero-crossings $z_{j,i}(\cdot)$. Alors, $G_{z_1}(\cdot)$ et $G_{z_2}(\cdot)$ sont des fonctions continues sur $[0, 1]$.

Soient $\tau_1 = (\tau_{1,1}, \dots, \tau_{1,p_1^*})$ l'ensemble des maxima locaux de G_{z_1} et $\tau_2 = (\tau_{2,1}, \dots, \tau_{2,p_2^*})$ l'ensemble des maxima locaux de G_{z_2} . Posons $\mu = \operatorname{argmin}_{\nu \in M_1} L_{G_{z_1}, G_{z_2}}(F(\nu))$ et soit $\epsilon > 0$. Alors, une ϵ -correspondance entre les deux ensembles de landmarks τ_1 et τ_2 est définie par le procédé suivant :

- pour $i = 1, \dots, p_1^*$, soit $\tau_{2,i}^* = (\tau_{2,j_1}, \dots, \tau_{2,j_i})$ l'ensemble de landmarks qui sont tels que $F(\mu)(\tau_{2,k}) \in [\tau_{1,i} - \epsilon, \tau_{1,i} + \epsilon]$, $\tau_{2,k} \in \tau_{2,i}^*$ et qui n'ont pas encore été associés à un autre landmark $\tau_{1,k}$, $k = 1, \dots, i-1$.
- si $\tau_{2,i}^*$ n'est pas vide alors le landmark τ_{2,j_1} est associé au landmark $\tau_{1,i}$.

Preuve : Nous avons seulement besoin de prouver que $G_{z_1}(\cdot)$ et $G_{z_2}(\cdot)$ sont des fonctions continues sur $[0, 1]$ ce qui implique que $L_{G_{z_1}, G_{z_2}}(\cdot)$ a un minimum. Soit :

$$g_{z_1,i}(x) = \int_0^{s_{z_1,i}} |\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right) ds, \quad x \in [0, 1].$$

Pour presque tout $s \in [0, s_{z_1,i}]$, $x \mapsto |\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right)$ est continue sur $[0, 1]$ et pour tout $x \in [0, 1]$, $|\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right) \leq \|\theta\|_\infty |\log(s)|^3$. Puisque $s \mapsto |\log(s)|^3$ est intégrable sur $[0, s_{z_1,i}]$, le théorème de continuité sous le signe intégral implique que $g_{z_1,i}$ est continue sur $[0, 1]$. Puisque $G_{z_1}(x) = \sum_{i=1}^{p_1} g_{z_1,i}(x)$, cette fonction est donc continue sur $[0, 1]$. La preuve pour montrer que $G_{z_2}(\cdot)$ est continue sur $[0, 1]$ est identique. \square

Remarque 4.1 Dans la Proposition 4.1, la définition des intensités structurelles est légèrement différente de celle donnée dans la chapitre 3. Rappelons que notre première définition était : $G_{z_1}(x) = \sum_{i=1}^{p_1} \int_0^{s_{z_1,i}} \frac{1}{s} \theta\left(\frac{x - z_{1,i}(s)}{s}\right) ds = \int_{-\log(s_{z_1,i})}^{+\infty} \theta\left(\frac{x - z_{1,i}(e^{-u})}{e^{-u}}\right) du$. Ainsi, lorsque l'on considère le logarithme du paramètre d'échelle ordinaire s , tous les noyaux $\theta\left(\frac{x - z_{1,i}(e^{-u})}{e^{-u}}\right)$, $u \geq -\log(s_{z_1,i})$ ont la même amplitude. Cette définition est motivée par le fait que les chercheurs en vision par ordinateur considèrent qu'un choix naturel de paramètre d'échelle, pour un signal continu, est le logarithme du paramètre d'échelle ordinaire (voir e.g. Lindeberg [58]). En particulier, un argument d'échelle qui requiert que les structures à des échelles différentes soient traitées de manière similaire, implique directement qu'une manière naturelle de définir la durée de vie des structures dans le plan temps-échelle est de prendre comme niveau de résolution le logarithme du paramètre d'échelle ordinaire. Ainsi, avec ce choix, la forme des modes de l'intensité structurelle doit normalement refléter la forme et la longueur des lignes de zero-crossings dans le plan temps-échelle. Cependant, $s \mapsto \frac{1}{s}$ n'est pas intégrable sur $[0, 1]$ et G_{z_1} n'est donc pas bornée. Afin de rendre $G_{z_1}(\cdot)$ continue sur $[0, 1]$, nous avons remplacé la fonction de poids $s \mapsto \frac{1}{s}$ par $s \mapsto |\log(s)|^3$ dans la Proposition 4.1. Etant donné que les courbes de ces fonctions de poids sont proches l'une de l'autre sur $[0, 1]$, la forme des intensités structurelles introduites dans la Proposition 4.1 est très proche de celle définie précédemment. Nous avons introduit cette nouvelle définition de l'intensité structurelle afin d'étudier de manière théorique les propriétés de l' ϵ -correspondance. Toutefois, en pratique, les intensités structurelles sont calculées pour des échelles strictement positives et quelque soit la définition ceci conduit donc toujours à des fonctions continues et bornées. Ainsi, dans les figures suivantes, les différentes intensités structurelles sont calculées avec la fonction de poids "naturelle" $\frac{1}{s}$ (nous avons constaté dans les différentes simulations effectuées, que remplacer $\frac{1}{s}$ par $|\log(s)|^3$ ne change pas la forme des intensités structurelles et ne modifie donc pas le résultat de l' ϵ -correspondance).

Avant d'illustrer cette technique de mise en correspondance automatique sur quelques exemples réels et par quelques simulations, nous devons essayer de savoir si l' ϵ -correspondance est une méthode consistante quand les courbes f_1 et f_2 sont

inconnues et observées avec du bruit. Dans le chapitre 3, il a été montré que si l'on définit les landmarks d'un signal comme les modes de l'intensité structurelle des lignes de zero-crossings, alors les maxima locaux de l'intensité structurelle des zero-crossings estimés converge en probabilité vers les landmarks du signal. Si \hat{G}_{z_1} et \hat{G}_{z_2} sont les intensités structurelles des lignes de zero-crossings de deux signaux f_1 et f_2 contaminés par du bruit, une idée naturelle pour estimer la fonction de déformation qui aligne les landmarks de f_1 et f_2 est d'utiliser l' ϵ -correspondance entre les maxima locaux de \hat{G}_{z_1} et \hat{G}_{z_2} afin de déterminer les paires de landmarks à aligner. Afin de savoir si cette approche peut identifier la vraie fonction de déformation, nous devons tout d'abord étudier la convergence de $\text{argmin}_{\nu \in M_1} L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(F(\nu))$.

Proposition 4.2 Assumons les hypothèses de la Proposition 4.1. Supposons que pour $s_{\min} \geq s \geq s_*$, p_1 lignes de zero-crossings $\hat{z}_{1,i}(s)$ ont été estimées et convergent en probabilité vers $z_{1,i}(s)$, $i = 1, \dots, p_1$, et p_2 lignes de zero-crossings $\hat{z}_{2,i}(s)$ ont été estimées et convergent en probabilité vers $z_{2,i}(s)$, $i = 1, \dots, p_2$ quand $n \rightarrow \infty$, où s_* est définie dans la Proposition 3.13. Pour $x \in [0, 1]$ définissons :

$$\begin{aligned} G_{z_1}(x) &= \sum_{i=1}^{p_1} \int_0^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right) ds. \\ G_{z_2}(x) &= \sum_{i=1}^{p_2} \int_0^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x - z_{2,i}(s)}{s}\right) ds. \\ \hat{G}_{z_1}(x) &= \sum_{i=1}^{p_1} \int_{s_*}^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x - \hat{z}_{1,i}(s)}{s}\right) ds. \\ \hat{G}_{z_2}(x) &= \sum_{i=1}^{p_2} \int_{s_*}^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x - \hat{z}_{2,i}(s)}{s}\right) ds. \end{aligned}$$

Soit $\hat{\mu}_n = \text{argmin}_{\nu \in M_1} L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(F(\nu))$. Alors :

- pour toute suite $x_n \rightarrow x \in [0, 1]$, $\hat{G}_{z_1}(x_n) \rightarrow_p G_{z_1}(x)$ et $\hat{G}_{z_2}(x_n) \rightarrow_p G_{z_2}(x)$ quand $n \rightarrow \infty$.
- il existe une sous-suite $\{\hat{\mu}_{n_k}\}_{k \in \mathbb{N}}$ de $\{\hat{\mu}_n\}_{n \in \mathbb{N}}$ qui converge vers $\hat{\mu} \in M_1$ tel que $P(\hat{\mu} \in \text{argmin}_{\nu \in M_1} L_{G_{z_1}, G_{z_2}}(\nu)) \rightarrow 1$ quand $n \rightarrow \infty$.

Preuve : Pour $i = 1, \dots, p_1$ et $x \in [0, 1]$, définissons $g_{z_{1,i}}(x) = \int_0^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x - z_{1,i}(s)}{s}\right) ds$. Soit $z_{1,i,n}(s)$ une suite de fonctions telles que pour tout $s \in [0, s_{\min}]$, $|z_{1,i,n}(s) - z_{1,i}(s)| \rightarrow 0$ quand $n \rightarrow \infty$. Soit s_n une suite d'échelles qui converge vers zéro, et x_n une suite qui converge vers $x \in [0, 1]$. Définissons $g_{z_{1,i,n}}(x_n) = \int_{s_n}^{s_{\min}} |\log(s)|^3 \theta\left(\frac{x_n - z_{1,i,n}(s)}{s}\right) ds$. Etant donné que pour tout $n \in \mathbb{N}$ et presque tout $s \in [0, s_{\min}]$, $|\log(s)|^3 \theta\left(\frac{x_n - z_{1,i,n}(s)}{s}\right) \leq \|\theta\|_\infty |\log(s)|^3$, le théorème de convergence dominé et la continuité de θ impliquent que $g_{z_{1,i,n}}(x_n) \rightarrow g_{z_{1,i}}(x)$ quand $n \rightarrow \infty$ (puisque $s \mapsto |\log(s)|^3$ est intégrable sur $[0, s_{\min}]$). D'après nos hypothèses, nous avons que pour $s_{\min} \geq s > s_*$, $|\hat{z}_{1,i}(s) - z_{1,i}(s)| \rightarrow_p 0$ et que $s_* \rightarrow 0$ quand $n \rightarrow \infty$. Donc, par le théorème de l'application continue (continuous mapping theorem voir Billingsley [7], Corollary 2 p.31), nous

obtenons que $g_{z_{1,i}}(x_n) = \int_{s_*}^{s_{min}} |\log(s)|^3 \theta(\frac{x_n - z_{1,i}(s)}{s}) ds \rightarrow_p g_{z_{1,i}}(x)$, ce qui implique finalement que $x_n \rightarrow x \in [0, 1]$, $\hat{G}_{z_1}(x_n) \rightarrow_p G_{z_1}(x)$ et $\hat{G}_{z_2}(x_n) \rightarrow_p G_{z_2}(x)$ quand $n \rightarrow \infty$.

Prouvons maintenant la deuxième partie du théorème. Soient θ_1, θ_2 des fonctions continues sur $[0, 1]$, et soient $\theta_{1,n}, \theta_{2,n}$ deux suites de fonctions continues sur $[0, 1]$ qui sont telles que pour tout $x \in [0, 1]$, $|\theta_{j,n}(x) - \theta_j(x)| \rightarrow 0$, $j = 1, 2$ quand $n \rightarrow \infty$. Supposons que pour toute suite $x_n \rightarrow x \in [0, 1]$, $|\theta_{2,n}(x_n) - \theta_2(x)| \rightarrow 0$. De plus, supposons qu'il existe une constante $M > 0$ telle que pour tout $n \in \mathbb{N}$ et tout $x \in [0, 1]$, $|\theta_{j,n}(x)| \leq M$, $j = 1, 2$. Soit $\nu_n \rightarrow \nu \in M_1$. Alors, pour presque tout $x \in [0, 1]$, $F(\nu_n)(x) \rightarrow F(\nu)(x)$. Ainsi, d'après nos hypothèses sur $\theta_{1,n}$ et $\theta_{2,n}$, nous obtenons que pour presque tout $x \in [0, 1]$, $(\theta_{1,n}(x) - \theta_{2,n}(F(\nu_n)(x)))^2 \rightarrow (\theta_1(x) - \theta_2(F(\nu)(x)))^2$. Puisque pour tout $n \in \mathbb{N}$ et tout $x \in [0, 1]$, $(\theta_{1,n}(x) - \theta_{2,n}(F(\nu_n)(x)))^2 \leq 4M^2$, le théorème de convergence dominé implique que $L_{\theta_{1,n}, \theta_{2,n}}(F(\nu_n)) \rightarrow L_{\theta_1, \theta_2}(F(\nu))$. Comme dans la preuve de la Proposition 4.1, nous pouvons montrer, en utilisant le théorème de continuité sous le signe intégral, que \hat{G}_{z_1} et \hat{G}_{z_2} sont des fonctions continues sur $[0, 1]$. Puisque pour tout $x \in [0, 1]$, $j = 1, 2$, $\hat{G}_{z_j}(x) \leq \|\theta\|_\infty \int_{s_*}^{s_{min}} |\log(s)|^3 ds \leq \|\theta\|_\infty \int_0^{s_{min}} |\log(s)|^3 ds$, et étant donné que pour tout suite $x_n \rightarrow x \in [0, 1]$, $\hat{G}_{z_2}(x_n) \rightarrow_p G_{z_2}(x)$, le théorème de l'application continue implique que pour toute suite $\nu_n \rightarrow \nu \in M_1$, $L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(\nu_n) \rightarrow_p L_{G_{z_1}, G_{z_2}}(\nu)$ quand $n \rightarrow \infty$. Maintenant, posons $\hat{\mu}_n = \operatorname{argmin}_{\nu \in M_1} L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(\nu)$. Puisque M_1 est compact, il existe une sous-suite $\{\hat{\mu}_{n_k}\}_{k \in \mathbb{N}}$ de $\{\hat{\mu}_n\}_{n \in \mathbb{N}}$ qui converge vers $\hat{\mu} \in M_1$. Par définition de $\hat{\mu}_{n_k}$, nous avons que pour tout $\nu \in M_1$, $L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(F(\hat{\mu}_{n_k})) \leq L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(F(\nu))$. Donc, avec probabilité tendant vers un quand $n \rightarrow \infty$, nous obtenons que pour tout $\nu \in M_1$, $L_{G_{z_1}, G_{z_2}}(F(\hat{\mu})) \leq L_{G_{z_1}, G_{z_2}}(F(\nu))$ ce qui complète la preuve. \square .

Prouver des théorèmes limites pour des estimateurs de type “argmin” est une tâche difficile, bien que l'approche soit généralement simple conceptuellement. Dans la preuve de la Proposition 4.2, nous avons utilisé la méthodologie de l'epi-convergence pour étudier quelques propriétés des estimateurs $\hat{\mu}_n = \operatorname{argmin}_{\nu \in M_1} L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(\nu)$. Nous présentons brièvement la notion d'epi-convergence pour des fonctions semi-continues inférieurement (l-sc) et expliquons pourquoi il s'agit d'un outil très utile pour étudier les estimateurs de type “argmin”. Rappelons qu'une fonction f est l-sc si :

$$f(x) \leq \liminf_{n \rightarrow \infty} f(x_n),$$

pour toute séquence $x_n \rightarrow x$. En particulier, la continuité semi-inférieure implique que pour tout ensemble compact K , $\inf_{x \in K} f(x) = f(x_0)$ pour un $x_0 \in K$. Soit f_n une suite de fonctions l-sc. Une définition possible de l'epi-convergence de f_n vers une fonction l-sc f est la suivante :

Définition 4.2 Soit f_n, f des fonctions l-sc. f_n epi-converge vers f si pour tout $x \in \mathbb{R}$ nous avons :

$$\begin{aligned} \liminf_{n \rightarrow \infty} f_n(x_n) &\geq f(x) \text{ pour toute suite } x_n \rightarrow x, \text{ et} \\ \limsup_{n \rightarrow \infty} f_n(x_n) &\leq f(x) \text{ pour une suite } x_n \rightarrow x. \end{aligned}$$

L'epi-convergence pour l'analyse variationnelle et l'optimisation est étudiée en détail dans le livre de Rockafellar et Wets [80], et d'autres références sur l'epi-convergence en loi et la convergence d'estimateurs de type "argmin" peuvent être trouvées dans Knight [52]. Dans Rockafellar et Wets [80], il est montré que l'epi-convergence est un outil bien adapté pour étudier la convergence de $\inf f_n$ ou de l'ensemble des $\text{argmin } f_n$. Cependant, il est également expliqué qu'à moins que $\text{argmin } f$ soit réduit à un point, il serait irréaliste d'essayer de trouver des conditions qui garantissent que $\lim_{n \rightarrow \infty} \text{argmin } f_n = \text{argmin } f$ quand f_n epi-converge vers f . Ainsi, l'on doit plutôt se contenter de l'inclusion $\limsup_{n \rightarrow \infty} \text{argmin } f_n \subset \text{argmin } f$, qui signifie que pour toute suite $x_n \in \text{argmin } f_n$, il existe une sous-suite x_{n_k} qui converge vers x avec $x \in \text{argmin } f$. Donc, puisqu'il n'est pas garanti que le problème variationnel (4.4.2) a une solution unique, il semble très difficile de trouver des conditions qui garantissent que $\text{argmin}_{\nu \in M_1} L_{\hat{G}_{z_1}, \hat{G}_{z_2}}(\nu) \rightarrow_p \text{argmin}_{\nu \in M_1} L_{G_{z_1}, G_{z_2}}(\nu)$. Ainsi, l'étude de la consistance de l' ϵ -correspondance semble être une tâche loin d'être évidente. Toutefois, plutôt qu'étudier ce problème d'un point de vue théorique, nous proposons d'appliquer notre méthode d'alignement automatique à partir de landmarks à quelques exemples réels, et d'effectuer quelques simulations pour illustrer les très bonnes propriétés de cette approche en pratique.

4.1.3 Implémentation pratique et simulations

4.1.3.1 Discrétisation du problème variationnel

En pratique, le problème variationnel (4.4.2) peut être résolu efficacement à l'aide de la programmation dynamique. Supposons que deux vecteurs de caractéristiques de longueur n : $\Theta_1 = (\Theta_{1,1}, \dots, \Theta_{1,n})$ et $\Theta_2 = (\Theta_{2,1}, \dots, \Theta_{2,n})$ doivent être alignés. Une approximation discrète du problème variationnel (4.4.2) est alors la suivante : les vecteurs de déformation \tilde{w} de la forme

$$\tilde{w} = ((1, j(1)), \dots, (n, j(n)))$$

qui alignent au mieux les vecteurs Θ_1 et Θ_2 sont ceux tels que la dissimilarité (ou coût)

$$L(\Theta_1, \Theta_2, \tilde{w}) = \sum_{k=1}^n \left(\frac{\Theta_1(k)}{\max(\Theta_1)} - \frac{\Theta_2(j(k))}{\max(\Theta_2)} \right)^2$$

est suffisamment petite. Les vecteurs de déformation \tilde{w} doivent satisfaire plusieurs contraintes :

1. Monotonie : $j(k) \leq j(k+1)$.
2. Bords : $j(1) = 1$ et $j(n) = n$.
3. Pente : $0 \leq |j(k+1) - j(k)| \leq M$,

où $M > 0$ est une constante. La troisième condition est une contrainte de pente qui vise à éliminer les déformations trop brusques car irréalistes. Le recherche d'une

déformation de coût minimal peut se faire efficacement à partir d’algorithmes qui trouvent le plus court chemin dans un graphe. Pour l’alignement des intensités structurelles, nous avons pris $M = +\infty$ ce qui autorise n’importe quelle pente positive.

4.1.3.2 Un exemple réel

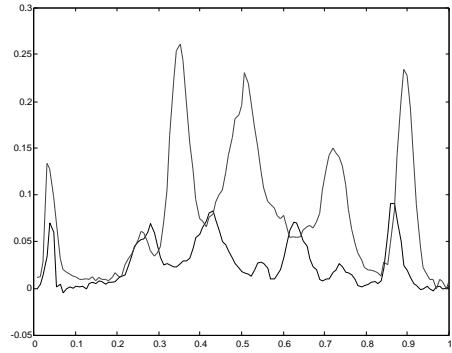
Revenons maintenant au problème de l’alignement des profils de densité pour deux vieux rats dans la région Medulla (voir Figure 4.1a). L’alignement par DTW des intensités structurelles \hat{G}_{z1}^+ et \hat{G}_{z2}^+ des zero-crossings estimés qui correspondent aux maxima locaux de \tilde{g}_1 et \tilde{g}_2 est donné dans la Figure 4.3. Dans Munoz Maldonado *et. al.* [71], les auteurs expliquent que, pour aligner les profils de densité, les cinq maxima principaux de \tilde{g}_1 aux points $x = 0.04, 0.28, 0.42, 0.63, 0.86$ ont été mis en correspondance avec les cinq maxima principaux de \tilde{g}_2 aux points $x = 0.04, 0.35, 0.51, 0.72, 0.89$. Dans la Figure 4.3c, on peut constater que s’il l’on applique l’ ϵ -correspondance avec $\epsilon = 0.01$ entre les modes de \hat{G}_{z1}^+ et \hat{G}_{z2}^+ , alors les vecteurs de landmarks qui doivent être associés sont $\tau_1^+ = (0.04, 0.28, 0.42, 0.63, 0.86)$ et $\tau_2^+ = (0.04, 0.35, 0.51, 0.72, 0.89)$. Ainsi, dans cet exemple, l’ ϵ -correspondance aligne correctement les cinq pics principaux de \hat{G}_{z1}^+ et \hat{G}_{z2}^+ et correspond de plus avec l’alignement que l’on aurait naturellement fait “avec les yeux”.

Dans un problème d’alignement de courbes, il peut être souvent souhaitable d’aligner à la fois les maxima et les minima communs. Dans la Figure 4.4, nous avons représenté les intensité structurelles \hat{G}_{z1}^- et \hat{G}_{z2}^- des zero-crossings estimés qui correspondent aux minima significatifs de \tilde{g}_1 et \tilde{g}_2 ainsi que leur alignement par DTW. Avec $\epsilon = 0.01$, les vecteurs des minima locaux qui doivent être alignés sont $\tau_1^- = (0.12, 0.34, 0.52, 0.79)$ et $\tau_2^- = (0.13, 0.41, 0.63, 0.83)$. Afin d’aligner les extrema de \tilde{g}_1 et \tilde{g}_2 , il est nécessaire de construire les vecteurs de landmarks τ_1 et τ_2 qui contiennent à la fois des minima et des maxima pour calculer la transformation $u = G_{(\tau_1, \tau_2)}$ qui aligne les courbes \tilde{g}_1 et \tilde{g}_2 , où G est un opérateur de déformation approprié. Etant donné que l’on peut espérer qu’un maxima significatif soit suivi d’un minima significatif, il peut être tentant de définir $\tau_1 = \tau_1^+ \cup \tau_1^-$ et $\tau_2 = \tau_2^+ \cup \tau_2^-$ de telle sorte que les éléments de τ_1 et τ_2 soient triés en ordre croissant, i.e :

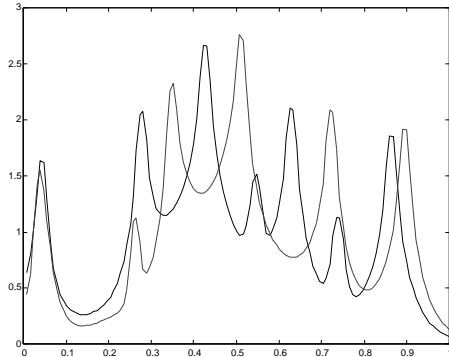
$$\begin{aligned}\tau_1 &= (0.04, 0.12, 0.28, 0.34, 0.42, 0.52, 0.63, 0.79, 0.86) \\ \tau_2 &= (0.04, 0.13, 0.35, 0.41, 0.51, 0.63, 0.72, 0.83, 0.90)\end{aligned}.$$

Dans notre exemple, cette méthode fonctionne puisqu’elle préserve la correspondance entre les maxima et minima locaux. Toutefois, si pour $i > j$ donnés, nous avons $\tau_1^+(i) < \tau_1^-(j)$ et $\tau_2^+(i) > \tau_2^-(j)$ cette procédure donne alors deux vecteurs de landmarks qui ne respectent pas la contrainte de monotonicité pour la transformation u . Ainsi, pour pouvoir aligner à la fois les maxima et les minima locaux de deux courbes, l’approche suivante a été utilisée :

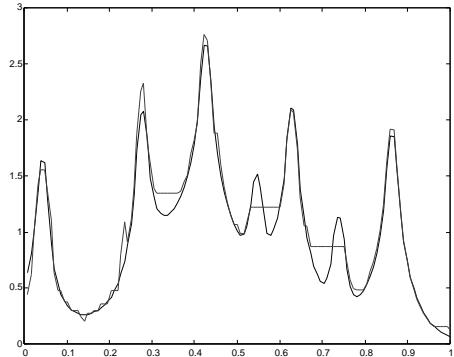
- soit $\hat{G}_{z1}^* = \hat{G}_{z1}^+ - \hat{G}_{z1}^-$ et $\hat{G}_{z2}^* = \hat{G}_{z2}^+ - \hat{G}_{z2}^-$.
- définissons $\mu^* = \operatorname{argmin}_{\nu \in M_1} L_{\hat{G}_{z1}^*, \hat{G}_{z2}^*}(F(\nu))$ et $w^*(x) = F(\mu^*)(x)$.



(a)



(b)



(c)

FIG. 4.3: (a) Profil de densité pour deux vieux dans la région Medulla, recalés sur l'intervalle $[0, 1]$ et échantillonnés à $n = 128$ points (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé) (b) Intensités structurelles non-alignées \hat{G}_{z1}^+ (trait plein) et \hat{G}_{z2}^+ (trait pointillé) (c) Alignement des intensités structurelles \hat{G}_{z1}^+ (trait plein) et \hat{G}_{z2}^+ (trait pointillé) par DTW

- soit $\tau_1^{*,+}$ et $\tau_2^{*,+}$ les vecteurs obtenus par ϵ -correspondance entre les *maxima locaux* de \hat{G}_{z1}^* et \hat{G}_{z2}^* , et $\tau_1^{*,-}$ et $\tau_2^{*,-}$ les vecteurs obtenus par ϵ -correspondance entre *minima locaux* de \hat{G}_{z1}^* et \hat{G}_{z2}^* , avec le chemin de coût minimal w^* .
- définissons $\tau_1^* = \tau_1^{*,+} \cup \tau_1^{*,-}$ et $\tau_2^* = \tau_2^{*,+} \cup \tau_2^{*,-}$ de telle sorte que les éléments de τ_1^* et τ_2^* soient triés par ordre croissant. Prendre alors τ_1^* et τ_2^* comme les vecteurs de landmarks qui doivent se correspondre.

La Figure 4.5 donne une illustration de cette procédure pour les intensité structurelles de \tilde{g}_1 et \tilde{g}_2 . Avec $\epsilon = 0.01$, la correspondance entre les extrema locaux de \hat{G}_{z1}^* et \hat{G}_{z2}^* donne :

$$\begin{aligned}\tau_1^* &= (0.04, 0.10, 0.27, 0.34, 0.42, 0.52, 0.63, 0.79, 0.87) \\ \tau_2^* &= (0.04, 0.12, 0.35, 0.41, 0.51, 0.63, 0.72, 0.83, 0.90)\end{aligned}.$$

Cette méthode, que nous noterons ϵ -correspondance *MinMax*, préserve donc bien la correspondance entre les maxima et les minima locaux, et l'on peut constater que les vecteurs τ_j^* et τ_j ($j = 1, 2$) sont très proches l'un de l'autre.

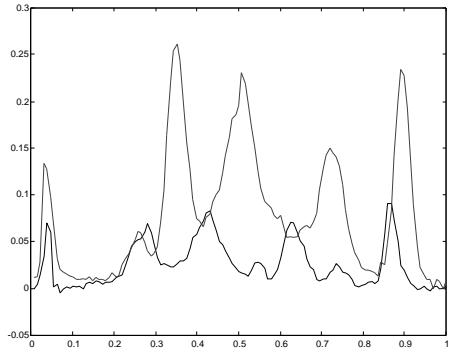
Maintenant que nous avons déterminé les paires de landmarks qui doivent se correspondre, nous pouvons construire une fonction de déformation pour aligner les deux profils de densité. Pour calculer cette déformation, nous avons utilisé la méthode de Ramsay [75] avec $\lambda = 0.001$ (voir chapitre 2). L'estimateur par ondelette invariant par translation avec seuillage hard (TI-H) de Coifman et Donoho [26] a été utilisé pour débruiter \tilde{g}_1 et \tilde{g}_2 avec l'ondelette Symmlet 8. Nous avons choisi la procédure TI-H car les simulations effectuées dans Antoniadis *et. al.* [3] pour comparer différents estimateurs par ondelettes, montrent que TI-H est significativement meilleure que d'autres méthodes pour estimer des fonctions régulières. Le résultat, donné dans la Figure 4.6, montre que les extrema significatifs de \tilde{g}_1 et \tilde{g}_2 sont correctement alignés après déformation de \tilde{g}_2 .

4.1.3.3 Simulations

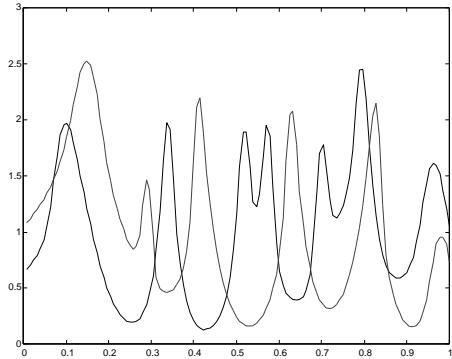
Nous proposons d'effectuer à présent quelques simulations numériques pour évaluer les performances de notre méthode automatique d'alignement à partir de landmarks, et les comparer avec celles d'un algorithme d'alignement global par DTW proposé par Wang et Gasser ([91],[92]). Avant de décrire le modèle utilisé pour les simulations, nous présentons brièvement l'approche suggérée dans [91] et [92].

Soient f_1 et f_2 deux fonctions observées à partir de données bruitées $\{y_{j,1}, \dots, y_{j,n}\}$ aux mêmes instants $t_i, i = 1, \dots, n$ (i.e $y_{j,i} = f_j(t_i) + \epsilon_{j,i}, i = 1, \dots, n$ pour $j = 1, 2$). La méthode du noyau de Gasser et Müller [41] est utilisée pour estimer f_1 et f_2 et leur dérivé première :

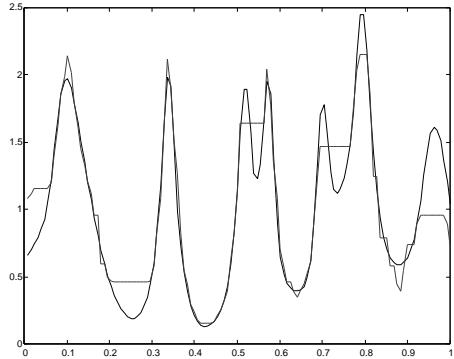
$$\hat{f}_j^{(k)}(t) = \frac{1}{b_k^{k+1}} \sum_{i=1}^n \int_{s_{i-1}}^{s_i} y_{j,i} K_k\left(\frac{t-u}{b}\right) du,$$



(a)

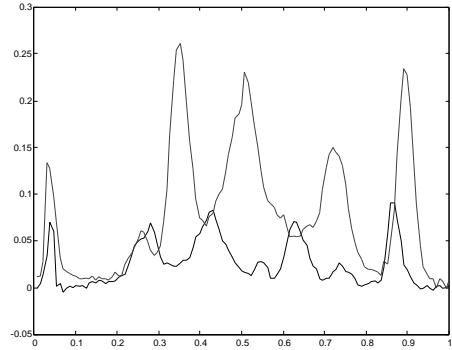


(b)

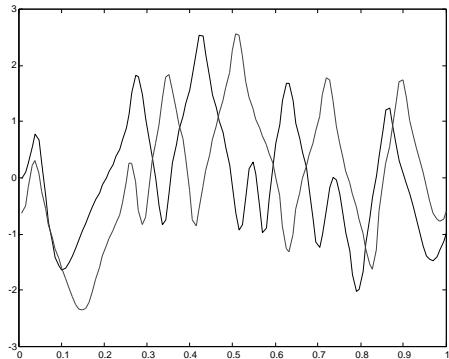


(c)

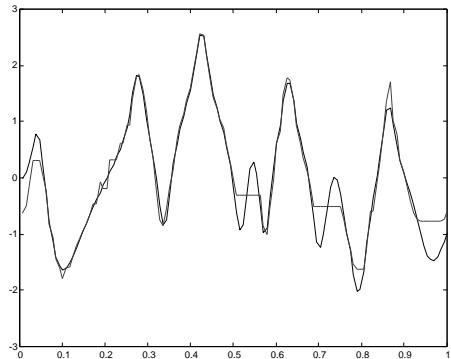
FIG. 4.4: (a) Profil de densité pour deux vieux dans la région Medulla, recalés sur l'intervalle $[0, 1]$ et échantillonnés à $n = 128$ points (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé) (b) Intensités structurelles non-alignées \hat{G}_{z1}^- (trait plein) et \hat{G}_{z2}^- (trait pointillé) (c) Alignement des intensités structurelles \hat{G}_{z1}^- (trait pointillé) et \hat{G}_{z2}^- (trait pointillé) par DTW.



(a)

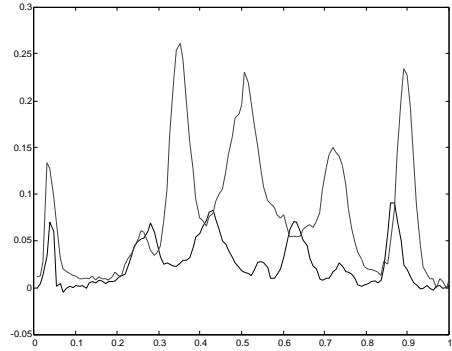


(b)

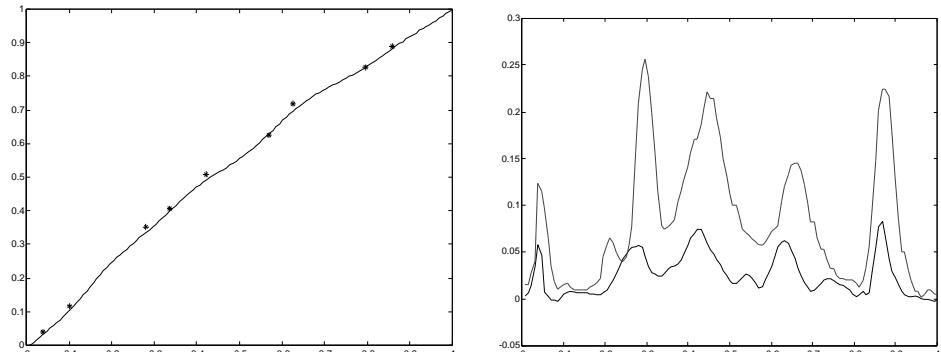


(c)

FIG. 4.5: (a) Profil de densité pour deux vieux dans la région Medulla, recalés sur l'intervalle $[0, 1]$ et échantillonnés à $n = 128$ points (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé) (b) Intensités structurelles non-alignées \hat{G}_{z1}^* (trait plein) et \hat{G}_{z2}^* (trait pointillé) (c) Alignement des intensités structurelles \hat{G}_{z1}^* (trait plein) et \hat{G}_{z2}^* (trait pointillé) par DTW.



(a)



(b)

(c)

FIG. 4.6: (a) Profil de densité pour deux vieux dans la région Medulla, recalés sur l'intervalle $[0, 1]$ et échantillonnés à $n = 128$ points (\tilde{g}_1 : trait plein et \tilde{g}_2 : trait pointillé) (b) Fonction de déformation u avec les paires de landmarks à aligner (*) (c) Débruitage et alignement de \tilde{g}_1 (trait plein) et \tilde{g}_1 (trait pointillé).

où k est l'ordre des dérivées, K_k est un noyau d'ordre $k + 2$ et b_k est une taille de fenêtre globale qui est indépendante de t mais qui dépend de k . La suite s est définie par $s_i = (t_i + t_{i+1})/2, i = 1, \dots, n - 1, s_0 = 0, s_n = 1$. La taille de fenêtre optimale b_k est choisi par la méthode de plug-in : estimer la fenêtre optimale qui minimise l'estimation de l'erreur quadratique (ou $L2$) asymptotique (voir Gasser, Kneip et Kohler [40]). Dans nos simulations, nous avons utilisé la routine `g1kern.f`, qui peut être téléchargée à l'URL : www.unizh.ch/biostat/, pour calculer les estimateurs à noyau. L'estimation d'une fonction de déformation \hat{u} est alors définie comme le minimum, parmi tous les difféomorphismes croissants $u : [0, 1] \rightarrow [0, 1]$, de la fonctionnelle :

$$L_{\hat{f}_1, \hat{f}_2, \alpha}(u) = \int_0^1 \alpha^2 \left(\frac{\hat{f}_1(x)}{\|\hat{f}_1\|_\infty} - \frac{\hat{f}_2(u(x))}{\|\hat{f}_2\|_\infty} \right)^2 + (1 - \alpha)^2 \left(\frac{\hat{f}'_1(x)}{\|\hat{f}'_1\|_\infty} - \frac{\hat{f}'_2(u(x))}{\|\hat{f}'_2\|_\infty} \right)^2 + 2\phi(u(x))dx,$$

où $\alpha \in [0, 1]$ et $\phi(u(x)) = 0.001 \frac{(u(x)-x)^2}{1+(u(x)-x)^2}$ pénalise les fonctions de déformation trop irrégulières. En suivant les recommandations pratiques discutées dans [92], nous résolvons le problème variationnel ci-dessus par DTW pour trois valeurs différentes de $\alpha = 0.3, 0.5, 0.7$ et choisissons ensuite comme meilleur alignement celui qui donne le plus petit coût $L_{\hat{f}_1, \hat{f}_2, \alpha}(u)$. Nous imposons également la contrainte de pente suivante : $|u(x) - x| \leq 1/2$ et utilisons la taille de fenêtre optimale b_2 (pour l'estimation de f''_j) afin d'estimer f'_j de manière à ce que les dérivées des courbes soient estimées avec un “sur-lissage”.

L'évaluation des deux méthodes d'alignement est effectuée pour le problème basique du recalage d'une fonction sur une autre. La fonction test $s(t)$ tracée dans la Figure 4.7 est celle utilisée par Wang et Gasser [91] dans leurs simulations, définie par :

$$\begin{aligned} s(t) &= s_1(t) + 4t \sin(31.4(0.45 - t)), t \leq 0.45, \\ s(t) &= s_1(t) + 10(s_1(t) - s_2(t))(0.45 - t), 0.45 \leq t \leq 0.55, \\ s(t) &= s_2(t) + 4(1 - t) \sin(31.4(t - 0.55)), t \geq 0.55, \end{aligned}$$

avec

$$\begin{aligned} s_1(t) &= 0.25(t - 5)^2 - 8(0.45 - t), \\ s_2(t) &= -0.25(t + 4)^2 + 8(t - 0.55). \end{aligned}$$

Nous considérons le modèle :

$$\begin{aligned} f_1(t) &= s(t), \\ f_2(t) &= a * s(h(t)) + d, \end{aligned}$$

où h est une fonction strictement croissante de $[0, 1]$ dans $[0, 1]$, et a et d sont des constantes. Les fonctions f_1 et f_2 ont 10 extrema que nous noterons $\tau_1 = (\tau_{1,j}, j = 1, \dots, 10)$ et $\tau_2 = (\tau_{2,j} = h(\tau_{1,j}), j = 1, \dots, 10)$ respectivement.

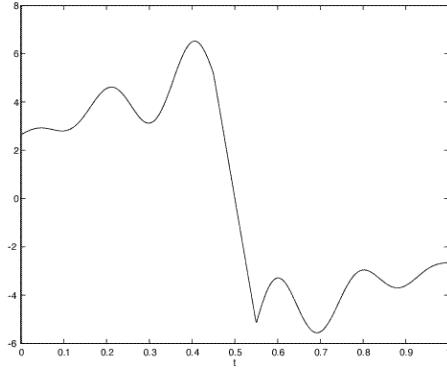


FIG. 4.7: Fonction test

Dans nos simulations, nous avons considéré des fonctions de déformation non-linéaires :

$$h_i(t) = t + \frac{\alpha_i}{2\pi} \sin(2\pi(\beta_i t + \gamma_i)),$$

où les paramètres $\alpha_i, \beta_i, \gamma_i$ sont générés aléatoirement :

$$\begin{aligned}\alpha_i &= \max\{-0.7, \min\{0.35N(0, 1), 0.7\}\}, \\ \beta_i &= \min\{1 + 0.2N(0, 1), 1.4\}, \\ \gamma_i &= 0.5(U(0, 1) - 0.5),\end{aligned}$$

où $N(0, 1)$ désigne une variable aléatoire de loi normale standard et $U(0, 1)$ une variable aléatoire de loi uniforme sur $(0, 1)$. Comme expliqué dans Wang et Gasser [91], typiquement $|\beta_i| \leq 1.4$ et ainsi

$$h'_i(t) = 1 + \alpha_i \beta_i \cos(2\pi(\beta_i t + \gamma_i)) \geq 1 - |\alpha_i \beta_i| > 0.$$

Nous avons donc des fonctions strictement croissantes et nous pouvons éventuellement ré-échelonner h_i afin d'avoir $h_i(0) = 0$ et $h_i(1) = 1$. Ensuite, notre modèle est à nouveau :

$$\begin{aligned}f_{1,i}(t) &= s(t) \quad ; \quad f_{2,i}(t) = a_i s(h_i(t)) + d_i, \\ \text{avec } a_i &= \max(1 + N(0, 1), 0.5) \quad ; \quad d_i = 4N(0, 1).\end{aligned}$$

Pour $i = 1, \dots, 100$, deux suites de données bruitées de longueur n sont alors générées :

$$y_{j,i,k} = f_{j,i}(t_k) + \sigma_{j,i} \epsilon_{j,i,k}, \quad k = 1, \dots, n,$$

où $t_k = \frac{k}{n}$, $\epsilon_{j,i,k}$ sont des variables i.i.d. de loi normale, de moyenne nulle et de variance 1, et $\sigma_{j,i}$ sont des niveaux de bruits inconnus. Les facteurs dans nos simulations

sont la taille d'échantillon n et les valeurs de $\sigma_{j,i}$. Nous donnons les résultats pour deux tailles d'échantillon, $n = 128$ (une taille modeste) et $n = 512$ (une taille plus importante) et un rapport signal sur bruit RSNR égal à 7 (un faible niveau de bruit), 5 (un niveau de bruit moyen) et 3 (un fort niveau de bruit), où :

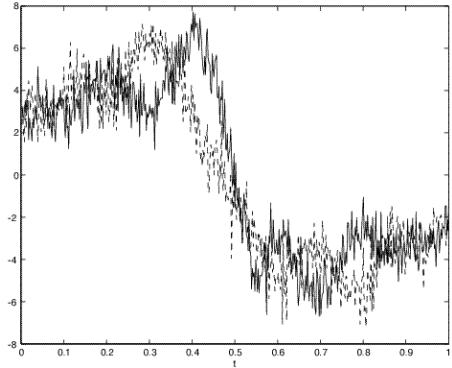
$$RSNR(f_{j,i}, \sigma_{j,i}) = \frac{\sqrt{\int_0^1 (f_{j,i}(x) - \bar{f}_{j,i})^2 dx}}{\sigma_{j,i}}, \text{ avec } \bar{f}_{j,i} = \int_0^1 f_{j,i}(x) dx.$$

Pour $i = 1, \dots, 100$, soit \hat{h}_i la fonction de déformation obtenue par la méthode de Wang et Gasser [91]. Pour la i -ème simulation et $j = 1, 2$, définissons $G_{z_{j,i}}^+$ et $\hat{G}_{z_{j,i}}^+$ respectivement comme les intensités structurelles des vrais zero-crossings et des zero-crossings estimés qui correspondent aux maxima locaux, et définissons $G_{z_{j,i}}^-$, $\hat{G}_{z_{j,i}}^-$ respectivement comme les intensités structurelles des vrais zero-crossings et des zero-crossings estimés qui correspondent aux minima locaux. Ensuite, posons $G_{z_{j,i}}^* = G_{z_{j,i}}^+ - G_{z_{j,i}}^-$ et définissons $\tau_{1,i}^*$ et $\tau_{2,i}^*$ comme les vecteurs des vrais landmarks obtenus par ϵ -correspondance MinMax (avec $\epsilon = 4/n$) entre les extrema locaux de $G_{z_{1,i}}^*$ et $G_{z_{2,i}}^*$. De façon similaire, soit $\hat{G}_{z_{j,i}}^* = \hat{G}_{z_{j,i}}^+ - \hat{G}_{z_{j,i}}^-$ et définissons $\hat{\tau}_{1,i}^*$ et $\hat{\tau}_{2,i}^*$ comme les vecteurs des landmarks estimés obtenus par ϵ -correspondance MinMax (avec $\epsilon = 4/n$) entre les extrema locaux de $\hat{G}_{z_{1,i}}^*$ et $\hat{G}_{z_{2,i}}^*$. Ensuite, nous posons $u_i(t) = G_{(\tau_{1,i}^*, \tau_{2,i}^*)}(t)$ et $\hat{u}_i(t) = G_{(\hat{\tau}_{1,i}^*, \hat{\tau}_{2,i}^*)}(t)$, $t \in [0, 1]$, où G est un opérateur de déformation (nous avons pris la méthode de Ramsay [75] avec $\lambda = 0.001$ pour calculer les fonctions de déformation). Dans le calcul de u_i et \hat{u}_i , nous avons également imposé $u_i(0) = 0$, $u_i(1) = 1$, $\hat{u}_i(0) = 0$, $\hat{u}_i(1) = 1$. Afin de comparer les performances de différentes ondelettes, nous reportons les résultats pour une ondelette B-Spline d'ordre $q = 4$ et pour une ondelette Gaussienne avec $\beta = 0.5$ (voir chapitre 3). Un exemple simulé, pour lequel la qualité de l'estimation de h_i et u_i est très bonne, est donné dans la Figure 4.8 (avec $n = 512$, $RSNR = 5$ et une ondelette Gaussienne pour estimer les landmarks).

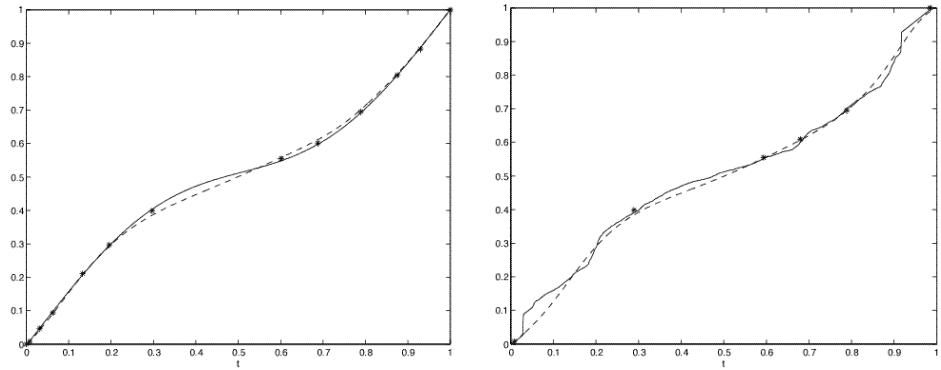
Pour comparer notre méthode avec celle de Wang et Gasser, nous avons utilisé deux critères d'erreur pour la i -ème simulation ($i = 1, \dots, 100$) :

$$\begin{aligned} WMSE_i &= \int_0^1 \left(f_{1,i}(t) - \frac{\hat{f}_{2,i}(\hat{w}_i(t)) - d_i}{a_i} \right)^2 dt, \\ \rho_i &= \frac{1}{10} \sum_{j=1}^{10} (\hat{w}_i(\tau_{1,j}) - h_i(\tau_{1,j}))^2, \end{aligned}$$

où \hat{w}_i correspond soit à \hat{h}_i soit à \hat{u}_i , et $\hat{f}_{2,i}$ est une estimation de $f_{2,i}$ obtenue par un méthode du noyau ou bien par la procédure TI-H. $WMSE_i$ est une mesure globale de la qualité de l'alignement, alors que ρ_i une mesure de l'erreur de l'alignement des extrema. Dans la Figure 4.9, nous avons représenté sous forme d'histogramme la moyenne et l'écart-type empirique de $WMSE_i$, $i = 1, \dots, 100$ pour différentes valeurs de n et $RSNR$, et les différentes méthodes d'estimation :



(a)



(b)

(c)

FIG. 4.8: (a) Fonctions $f_{1,i}$ (trait plein) et $f_{2,i}$ (trait pointillé) observées avec du bruit ($n = 512$, $RSNR = 5$) (b) Vraie fonction de déformation h_i (trait plein) et fonctions de déformation u_i (trait pointillé) obtenue à partir des vrais landmarks (*) associés par ϵ -correspondance MinMax (c) Fonction de déformation estimée \hat{h}_i (trait plein) par DTW et fonction de déformation \hat{u}_i (trait pointillé) obtenue à partir de landmarks estimés (*) associés par ϵ -correspondance MinMax.

- “WG” : méthode de Wang et Gasser avec $\hat{f}_{2,i}$ obtenue par un estimateur à noyau.
- “Auto TI-H B” : notre méthode avec une ondelette B-Spline et $\hat{f}_{2,i}$ obtenue par la procédure TI-H.
- “Auto Kernel B” : notre méthode avec une ondelette B-Spline et $\hat{f}_{2,i}$ obtenue par un estimateur à noyau.
- “Auto TI-H G” : notre méthode avec une ondelette Gaussienne et $\hat{f}_{2,i}$ obtenue par la procédure TI-H.
- “Auto Kernel G” : notre méthode avec une ondelette Gaussienne et $\hat{f}_{2,i}$ obtenue par un estimateur à noyau.

Dans la Figure 4.10, nous avons représenté sous forme d’histogramme la moyenne et l’écart-type empirique de $\rho_i, i = 1, \dots, 100$ pour différentes valeurs de n et $RSNR$, et les différentes méthodes d’estimation :

- “WG” : méthode de Wang et Gasser.
- “B-Spline” : notre méthode avec une ondelette B-Spline.
- “Gaussienne” : notre méthode avec une ondelette Gaussienne.

De manière évidente, les performances de notre méthode dépendent du nombre de landmarks estimés qui sont mis en correspondance. Nous avons donc également indiqué dans les Figures 4.11 et 4.12, les positions (dans $[0, 1]$) des landmarks estimés de f_1 qui sont mis en correspondance avec des landmarks estimés de f_2 à chaque simulation et pour différentes combinaisons des valeurs des facteurs.

Nous pouvons constater que les résultats obtenus avec une ondelette Gaussienne et une ondelette B-Spline sont sensiblement les mêmes. La moyenne empirique de $WMSE_i$ obtenue avec “WG” est toujours plus petite que celle obtenue avec “Auto TI-H B” ou “Auto TI-H G”. Pour $n = 128, RSNR = 7,5$ la moyenne et l’écart-type de $WMSE_i$ pour “Auto Kernel G” et pour “WG” ont la même amplitude, alors que “Auto Kernel B” donne le meilleur résultat. Pour $n = 512, RSNR = 7$, “Auto Kernel B” et “Auto Kernel G” font mieux que “WG” pour le critère $WMSE$. Ce cas correspond à la meilleure approximation des landmarks de f_1 mis en correspondance avec des landmarks estimés de $f_{2,i}, i = 1, \dots, 100$ (voir Figures 4.11b et 4.12b : à chaque simulation, six landmarks (dans $]0, 1[$) sont utilisés pour calculer la déformation). Pour $RSNR = 3$, la moyenne de $WMSE_i$ pour “WG” est beaucoup plus petite que celle obtenue avec notre méthode, ce qui peut être expliqué par les Figures 4.11ef et 4.12ef (à chaque simulation, seulement deux landmarks dans $]0, 1[$ (autour de 0.4 et 0.7) sont utilisés pour calculer \hat{u}_i).

L’erreur de l’alignement des extrema ($\rho_i, i = 1, 100$, voir Figure 4.10) est plus petite pour notre méthode avec $n = 128, RSNR = 7,5, 3$ et $n = 512, RSNR = 7$. La moyenne et l’écart-type de ρ_i obtenus avec les deux méthodes sont du même ordre pour $n = 512, RSNR = 5$, alors que la méthode de Wang et Gasser est meilleure pour $n = 512, RSNR = 3$. Ainsi, les graphes de la Figure 4.10 montrent que la

fonction de déformation obtenue avec notre méthode donne un meilleur alignement des landmarks de deux fonctions bruitées. Lorsqu'un nombre suffisant de landmarks peuvent être estimés et associés, les deux méthodes donnent des résultats similaires pour une mesure d'alignement global, quand f_1 et f_2 sont estimées par une méthode du noyau. L'un des désavantages de notre méthode est le choix d'une procédure de débruitage qui donnera généralement des fonctions estimées dont les landmarks ne sont pas positionnés exactement au niveau des modes des intensités structurelles des zero-crossings estimés. Ceci peut expliquer la raison pour laquelle "Auto Kernel" est meilleure que "Auto TI-H". L'estimation des landmarks est peut-être plus précise avec des estimateurs à noyau. Remarquons que généralement, les estimateurs par ondelettes donnent de mauvais résultats aux bornes de l'intervalle $[0, 1]$, ce qui peut aussi donner des différences en terme d'erreur quadratique entre "Auto TI-H" et "Auto Kernel". Puisque notre méthode semble être un outil efficace pour automatiquement associer les landmarks de deux fonctions, il serait intéressant de définir une fonctionnelle $L_{f_1, f_2}(u)$ qui combinerait une mesure de l'erreur d'alignement global entre f_1 et $f_2 \circ u$, avec une mesure de l'erreur d'alignement des landmarks.

Pour finir cette discussion, nous proposons d'évaluer les performances de l' ϵ -correspondance. Etant donné que f_2 est obtenue par une déformation de f_1 , ces deux courbes ont le même nombre d'extrema. Ainsi, une méthode d'alignement à partir de landmarks devrait mettre en correspondance les positions des extrema successifs de f_i , $i = 1, 2$. Pour $i = 1, \dots, 100$, définissons $\tau_{2,i} = (h_i(\tau_{1,1}), \dots, h_i(\tau_{1,10}))$ et soit $u_i^*(t) = G_{(\tau_1, \tau_2, i)}(t)$, $t \in [0, 1]$, où G est un opérateur de déformation. Alors, u_i^* est la fonction de déformation obtenue par alignement à partir de landmarks, si l'on suppose que les positions des extrema de f_1 et f_2 qui doivent être associés sont connues. Nous avons tout d'abord évalué les performances de l' ϵ -correspondance lorsque f_1 et f_2 sont observées sans bruit, afin de savoir si notre méthode automatique peut identifier les paires de landmarks qui doivent être mis en correspondance. La déviation absolue maximale $MXDV_i^* = \max_{t \in [0, 1]} |u_i^*(t) - u_i(t)|$ a été calculée pour $i = 1, \dots, 100$, et la moyenne et l'écart-type empiriques de $MXDV_i^*$ sont donnés dans le Tableau 4.1. Dans la Figure 4.13, nous avons indiqué les positions (dans $[0, 1]$) des extrema locaux de $G_{z_{1,i}}^*$ (i.e. landmarks de f_1) qui sont mis en correspondance avec des extrema de $G_{z_{2,i}}^*$ (i.e. landmarks de $f_{2,i}$) à chaque simulation ($i = 1, \dots, 100$) pour $n = 128$ et $n = 512$ (n étant le nombre de points d'échantillonnage de f_1 et $f_{2,i}$). Il peut être constaté dans la Figure 4.13 que pour chaque déformation h_i , notre méthode associe les 10 extrema de f_1 (dans $]0, 1[$) avec les 10 extrema de $f_{2,i}$. De plus, la moyenne et l'écart-type de $MXDV_i^*$ dans le Tableau 4.1 sont inférieurs à 0.01 pour $n = 128, 512$ ce qui est très satisfaisant. Ceci confirme que l' ϵ -correspondance semble être un outil efficace pour automatiquement associer les paires de landmarks qui auraient été marqués manuellement. Pour évaluer la consistance de l' ϵ -correspondance, nous avons également calculé l'erreur $MXDV_i = \max_{t \in [0, 1]} |\hat{u}_i(t) - u_i(t)|$ pour $i = 1, \dots, 100$, et donnons sa moyenne et son écart-type empirique dans le Tableau 4.1. Pour $n = 512$ et $RSNR = 7, 5$ les résultats sont satisfaisants puisque la moyenne de $MXDV$ est proche de 0.2 et l'écart-type

est entre 0.1 et 0.15. Ces deux cas correspondent aux meilleures approximations des landmarks de f_1 associés avec des landmarks estimés de f_2 (voir Figures 4.11bd et 4.12bd). Dans les autres cas ($n = 128, RSNR = 7, 5, 3$ et $n = 512, RSNR = 3$), notre méthode utilise principalement deux landmarks (autour de 0.4 et 0.7) pour effectuer le recalage et les résultats sont moins satisfaisants (la moyenne de $MXDV$ est plus grande que 0.3).

TAB. 4.1: Moyenne (M) et Ecart-type (STD) empiriques de $MXDV_i^*$ et $MXDV_i, i = 1, \dots, 100$ pour différentes valeurs de n et $RSNR$

		$MXDV^*$	$MXDV$	$MXDV$	$MXDV$
		$RSNR = 7$	$RSNR = 5$	$RSNR = 3$	
$n = 128$	M	$9.52e^{-3}$	$3.05e^{-2}$	$3.52e^{-2}$	$5.6e^{-2}$
	STD	$3.63e^{-3}$	$2.27e^{-2}$	$2.45e^{-2}$	$3.3e^{-2}$
$n = 512$	M	$8.76e^{-3}$	$1.82e^{-2}$	$2.37e^{-2}$	$2.96e^{-2}$
	STD	$5.89e^{-3}$	$1.01e^{-2}$	$1.42e^{-2}$	$2.29e^{-2}$

4.2 Recalage d'un ensemble de courbes

Supposons que l'on dispose d'un ensemble de m courbes f_1, \dots, f_m éventuellement observées avec du bruit, qui doivent être alignées. Afin de simplifier les notations, nous supposerons que ces courbes ont le même support e.g $[0, 1]$ (on peut toujours se ramener à cette situation par un recalage linéaire des données originales). Il s'agit alors de déterminer un ensemble de fonctions de déformation $u_i : [0, 1] \rightarrow [0, 1], i = 1, \dots, m$ telles que les courbes $f_1 \circ u_1, \dots, f_m \circ u_m$ soient représentées dans un référentiel commun. Nous avons jusqu'à présent proposé une méthode automatique d'alignement de deux fonctions à partir de landmarks. Afin de pouvoir synchroniser un ensemble de fonctions, il est donc nécessaire d'avoir une courbe f_e ou un ensemble de landmarks de référence, sur lesquels seront recalées toutes les fonctions. La courbe de référence doit suffisamment bien refléter la forme typique d'une fonction de l'échantillon, et doit avoir plus ou moins les mêmes caractéristiques que la plupart des courbes. Comme expliqué dans Wang et Gasser [91], le choix d'une courbe de référence est alors un compromis entre précision et temps de calcul. Afin d'expliquer les différentes méthodes que nous avons utilisées, nous supposerons que les landmarks des fonctions f_1, \dots, f_m ont été estimés, et qu'une méthode automatique a été définie afin de mettre en correspondance les landmarks de deux courbes f_i et f_j . Plusieurs possibilités peuvent alors être envisagées :

1. Si l'on suppose que l'ensemble des courbes ont une forme commune et constituent un groupe homogène, on pourrait en principe choisir une courbe au hasard comme fonction de référence. Cette méthode est particulièrement attractive lorsque m est grand car peu coûteuse en temps de calcul. Toutefois, si une courbe atypique est sélectionnée, la qualité de l'analyse statistique peut être affectée et l'estimation des fonctions de déformation sera plus difficile.

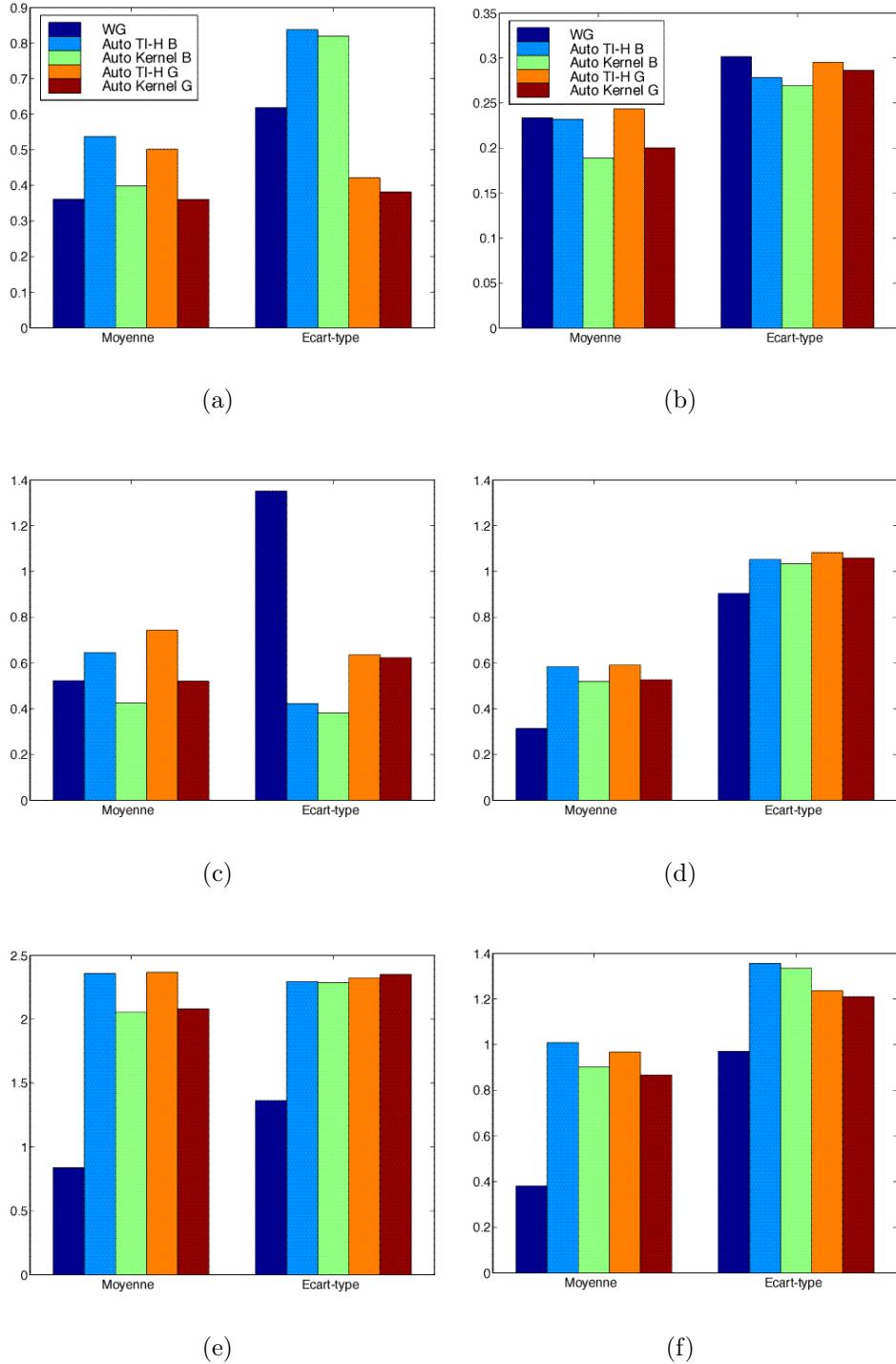


FIG. 4.9: Moyenne et écart-type empiriques de $WMSE_i, i = 1, \dots, 100$ pour différentes valeurs de n et $RSNR$: (a) $n = 128, RSNR = 7$, (b) $n = 512, RSNR = 7$, (c) $n = 128, RSNR = 5$, (d) $n = 512, RSNR = 5$, (e) $n = 128, RSNR = 3$, (f) $n = 512, RSNR = 3$.

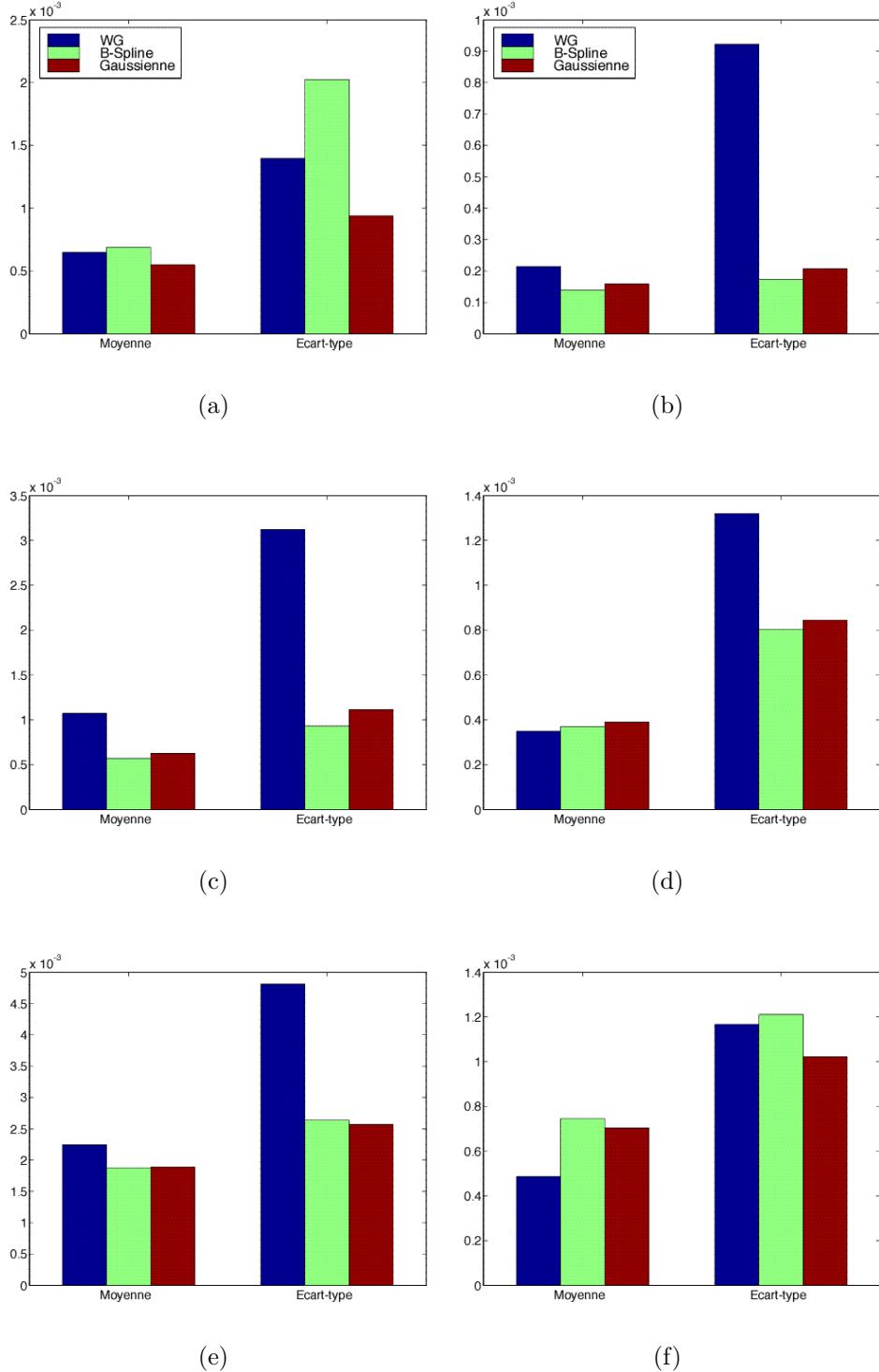


FIG. 4.10: Moyenne et écart-type empiriques de $\rho_i, i = 1, \dots, 100$ pour différentes valeurs de n et $RSNR$: (a) $n = 128, RSNR = 7$, (b) $n = 512, RSNR = 7$, (c) $n = 128, RSNR = 5$, (d) $n = 512, RSNR = 5$, (e) $n = 128, RSNR = 3$, (f) $n = 512, RSNR = 3$.

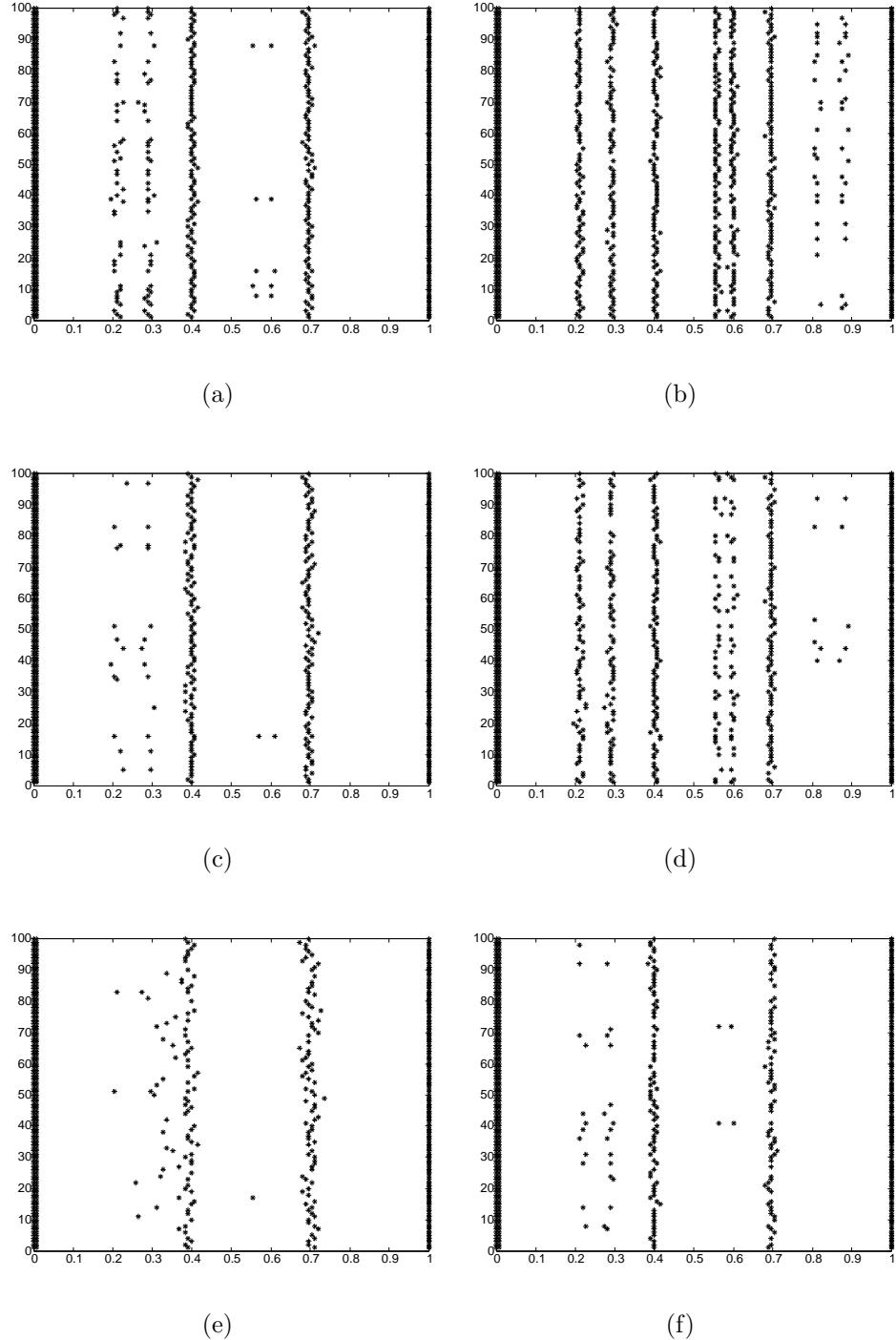


FIG. 4.11: Ondelette B-Spline : positions des landmarks estimés (dans $[0, 1]$) de f_1 qui sont mis en correspondance avec des landmarks estimés de $f_{2,i}$ ($i = 1, \dots, 100$). Dans chaque figure, l'axe vertical donne l'indice de la i -ème simulation, et l'axe horizontal est l'intervalle $[0, 1]$: (a) $n = 128$, $RSNR = 7$, (b) $n = 512$, $RSNR = 7$, (c) $n = 128$, $RSNR = 5$, (d) $n = 512$, $RSNR = 5$, (e) $n = 128$, $RSNR = 3$, (f) $n = 512$, $RSNR = 3$.

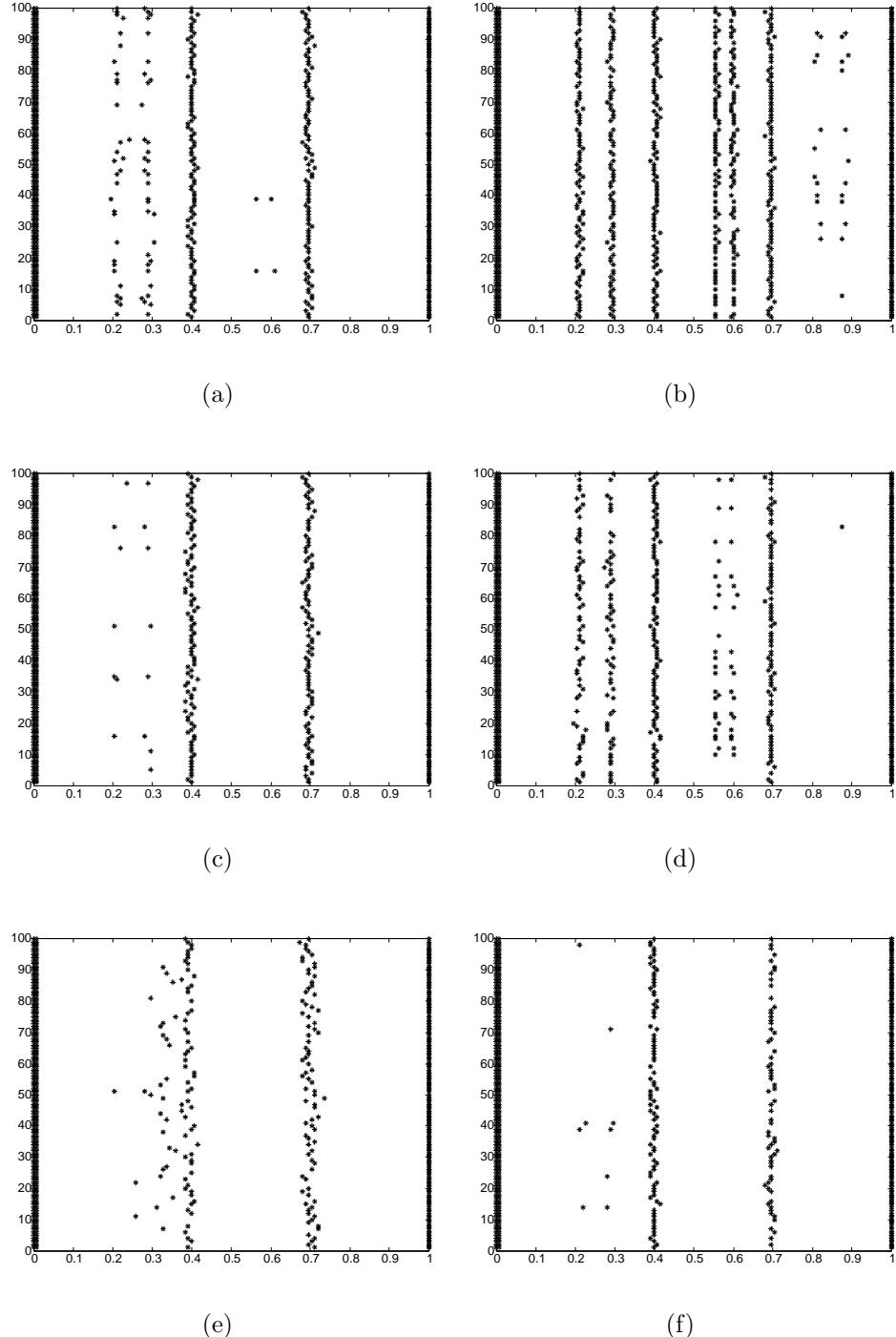


FIG. 4.12: Ondelette Gaussienne : positions des landmarks estimés (dans $[0, 1]$) de f_1 qui sont mis en correspondance avec des landmarks estimés de $f_{2,i}$ ($i = 1, \dots, 100$). Dans chaque figure, l'axe vertical donne l'indice de la i -ème simulation, et l'axe horizontal est l'intervalle $[0, 1]$: (a) $n = 128, RSNR = 7$, (b) $n = 512, RSNR = 7$, (c) $n = 128, RSNR = 5$, (d) $n = 512, RSNR = 5$, (e) $n = 128, RSNR = 3$, (f) $n = 512, RSNR = 3$.

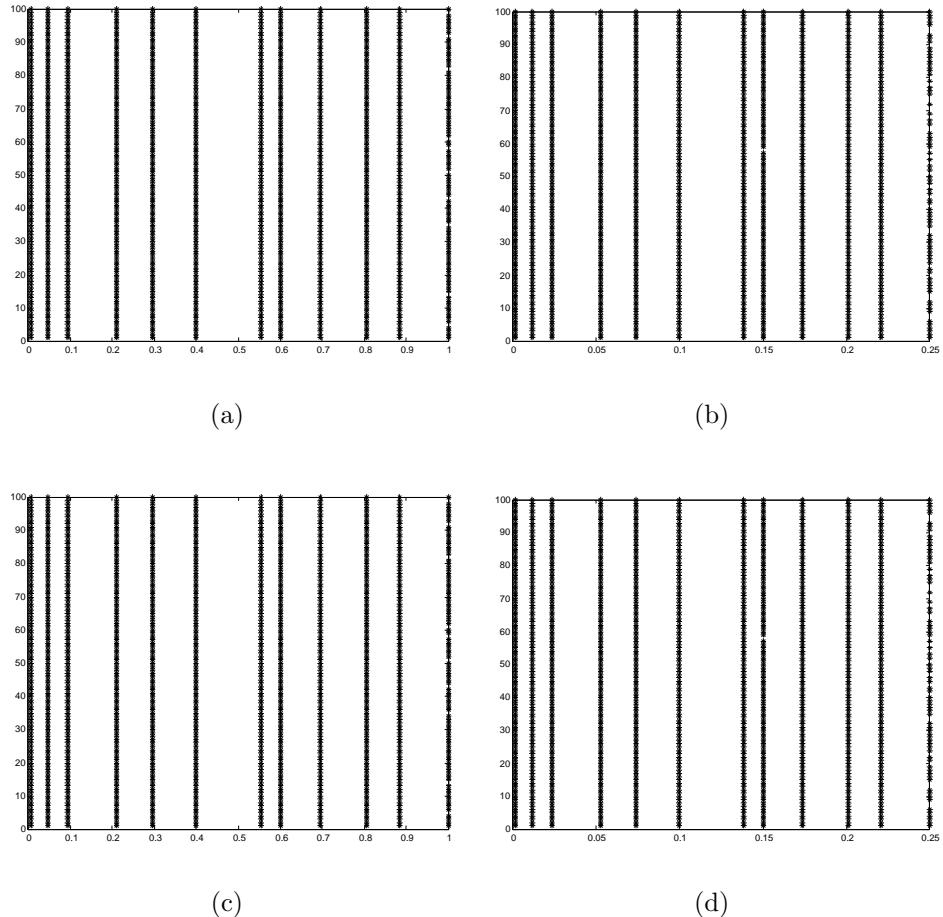


FIG. 4.13: Positions des landmarks (dans $[0, 1]$) de f_1 qui sont mis en correspondance avec des landmarks de $f_{2,i}$ ($i = 1, \dots, 100$). Dans chaque figure, l'axe vertical donne l'indice de la i -ème simulation, et l'axe horizontal est l'intervalle $[0, 1]$. Ondelette B-Spline : (a) $n = 128$, (b) $n = 512$. Ondelette Gaussienne : (a) $n = 128$, (b) $n = 512$

2. Prendre chaque courbe $f_i, i = 1, \dots, m$ comme courbe de référence. Pour chaque chaque courbe $f_j, j = 1, \dots, m, j \neq i$, calculer le vecteur $\tau_{i,j}$ qui contient les landmarks de f_i qui doivent être associés avec des landmarks de f_j . Calculer $\tau_i = \cap_{j=1, j \neq i}^m \tau_{i,j}$ qui est le vecteur contenant les landmarks de f_i qui sont “communs” à toutes les courbes, et notons alors par $\tau_{j,i}^*$ le vecteur de landmarks de f_j qui sont associés aux points du vecteur τ_i . Définissons alors f_{i_0} comme la courbe qui maximise la longueur du vecteur $\tau_i, i = 1, \dots, m$. Le vecteur τ_{i_0} contient donc le nombre maximal de landmarks communs à toutes les courbes. Définissons alors le vecteur τ^* comme la moyenne par composante, des vecteurs $\tau_{j,i_0}^*, j = 1, \dots, m$ (où $\tau_{i_0,i_0}^* = \tau_{i_0}$). Dans le cas où plusieurs vecteurs τ_{i_0} de longueur maximale M_0 peuvent être choisis, nous prenons celui qui maximise le coût : $\sum_{j=1}^m \sum_{k=1}^{M_0} (\tau_{j,i_0}^*[k] - \tau^*[k])^2$ où $\tau[k]$ représente la k -ème composante du vecteur τ .
- La synchronisation des fonctions f_1, \dots, f_m peut alors s’effectuer comme suit : soit G un opérateur de déformation. Les fonctions de déformation u_1, \dots, u_m pour recalier les courbes f_1, \dots, f_m sont alors définies par $u_i(t) = G_{(\tau^*, \tau_{j,i_0}^*)}(t), t \in [0, 1]$. Le principal désavantage de cette méthode est son coût algorithmique lorsque m est grand (l’ ϵ -correspondance nécessite que le problème variationnel (4.4.2) soit résolu $m(m - 1)$ fois).
3. Lorsque m est grand, on peut sélectionner au hasard un sous-ensemble de courbes de taille 2^j . Partitionner ensuite ce sous-ensemble en 2^{j-1} paires. Pour chaque paire de fonctions f_1, f_2 , déterminer les ensembles de landmarks τ_1 et τ_2 qui doivent être mis en correspondance et calculer la moyenne par composante τ^* de ces deux vecteurs. Pour un opérateur de déformation donné, calculer alors les fonctions de déformation $u_i(t) = G_{(\tau^*, \tau_i^*)}(t), t \in [0, 1], i = 1, 2$, puis déterminer la fonction moyenne $(f_1 \circ u_1 + f_2 \circ u_2)/2$. Nous obtenons donc ainsi 2^{j-1} fonctions moyennes. Partitionner ensuite ce groupe en 2^{j-2} paires et calculer à nouveau une fonction moyenne pour chaque paire. Répéter cette procédure jusqu’à obtenir une seule fonction moyenne qui sera prise comme courbe de référence f_e . Calculer ensuite les landmarks de f_e , puis déterminer pour $i = 1, \dots, m$ le vecteur $\tau_{i,e}$ de landmarks de f_i qui doivent être associés avec des landmarks de f_e . Poser ensuite $u_i(t) = G_{(\tau_{e,i}, \tau_{i,e})}(t)$ où $\tau_{e,i}$ désigne les landmarks de f_e associés à des landmarks de f_i .

Notons que si les courbes f_1, \dots, f_m sont inconnues et observées avec du bruit elles doivent être estimées avant d’être recalées. Dans la suite, nous avons choisi de débruiter les courbes à l’aide de la procédure TI-H (translation invariant and hard thresholding) de Coifman et Donoho [26] avec l’ondelette Symmlet 8. Comme expliqué précédemment, Nous avons choisi TI-H car cette procédure est significativement meilleure que d’autres méthodes à base d’ondelettes pour estimer des fonctions régulières (voir Antoniadis *et. al.* [3]).

Maintenant se pose la question de savoir quelle procédure parmi les possibilités (1)-(3) doit être choisie en pratique. Il n’est pas judicieux de choisir la première

possibilité car le résultat de l'analyse peut être complètement différent selon le choix d'une courbe atypique ou au contraire représentative de l'allure des courbes. Si m est petit, (2) est la procédure qu'il faut choisir. Si m est grand et que l'on ne dispose pas d'information a priori sur la qualité de l'échantillon de courbes, alors la méthode (3) est recommandée. Comme suggéré dans Wang et Gasser [91], il est aussi possible de combiner (2) et (3) : choisir un sous-ensemble de courbes de taille k et appliquer (2) à cet échantillon pour calculer une courbe de référence. Dans le reste de ce chapitre, nous allons nous intéresser à deux exemples réels d'analyse de données fonctionnelles qui nécessitent une étape de recalage qui permettra d'illustrer les procédures (2) et (3). L'alignement des données de la ganglioside et des données d'orthèse est décrit dans les sections 4.2.1 et 4.2.2. Nous indiquons ensuite, dans la section 4.3, quelles informations peuvent être extraites de ces données à l'aide de l'analyse de variance fonctionnelle.

4.2.1 Alignement des données de la ganglioside

Le but de l'étude conduite dans Munoz Maldonado *et. al.* [71] est la compréhension du processus physique du vieillissement dans le cerveau qui pourrait finalement expliquer les changements de performance. Les biologistes pensent que la distribution la ganglioside, une molécule présente dans les tissus du cerveau, est susceptible de changer avec l'âge dans certaines parties de ces tissus. La distribution de la ganglioside, représentée par un profil de densité, a donc été relevée par chromatographie en couche mince (TLC) dans trois zones du cerveau : Medulla (MD), Locus Coeruleus (LC) et Hippocampus (HY), chez cinq vieux rats (âgés de 12 mois) et cinq jeunes rats (âgés de 2 mois). Pour chaque rats, un échantillon de solvants organiques et de ganglioside est déposé au bas d'une plaque solide sur laquelle est répartie une fine couche de gel de silice. Le mélange remonte alors verticalement le long du gel par capillarité, dans la direction du solvant. L'action de la capillarité déplace les composants de la ganglioside le long de la plaque de gel de silice à différentes vitesses, selon la taille et la charge de leurs molécules. L'intensité des spots le long d'une colonne verticale dans la plaque de gel de silice est alors quantifiée par densitométrie et enregistrée à l'aide d'un scanner. Une description plus précise sur la méthode d'acquisition de ces données peut être trouvée dans [71]. Le but de l'analyse statistique proposée dans [71] est alors de savoir si pour une région du cerveau, le profil de densité de la ganglioside varie selon l'âge. Les données originales (sans bruit de fond) sont représentées dans la Figure 4.14.

Après avoir supprimé le bruit de fond dans les données, la modélisation suivante est proposée dans [71] pour les courbes d'une zone donnée du cerveau : les deux populations, vieux et jeunes rats, sont distribuées de manière indépendantes. La réponse de chaque expérience est une courbe $Y_{ij}(t)$, $t \in [0, T_{ij}]$ qui correspond à un profil de densité. L'indice $i = 1, 2$ indique de quel groupe il s'agit (vieux ou jeunes rats), et $j = 1, \dots, n_i$ est l'indice de la j -ème expérience dans le groupe i . Le modèle

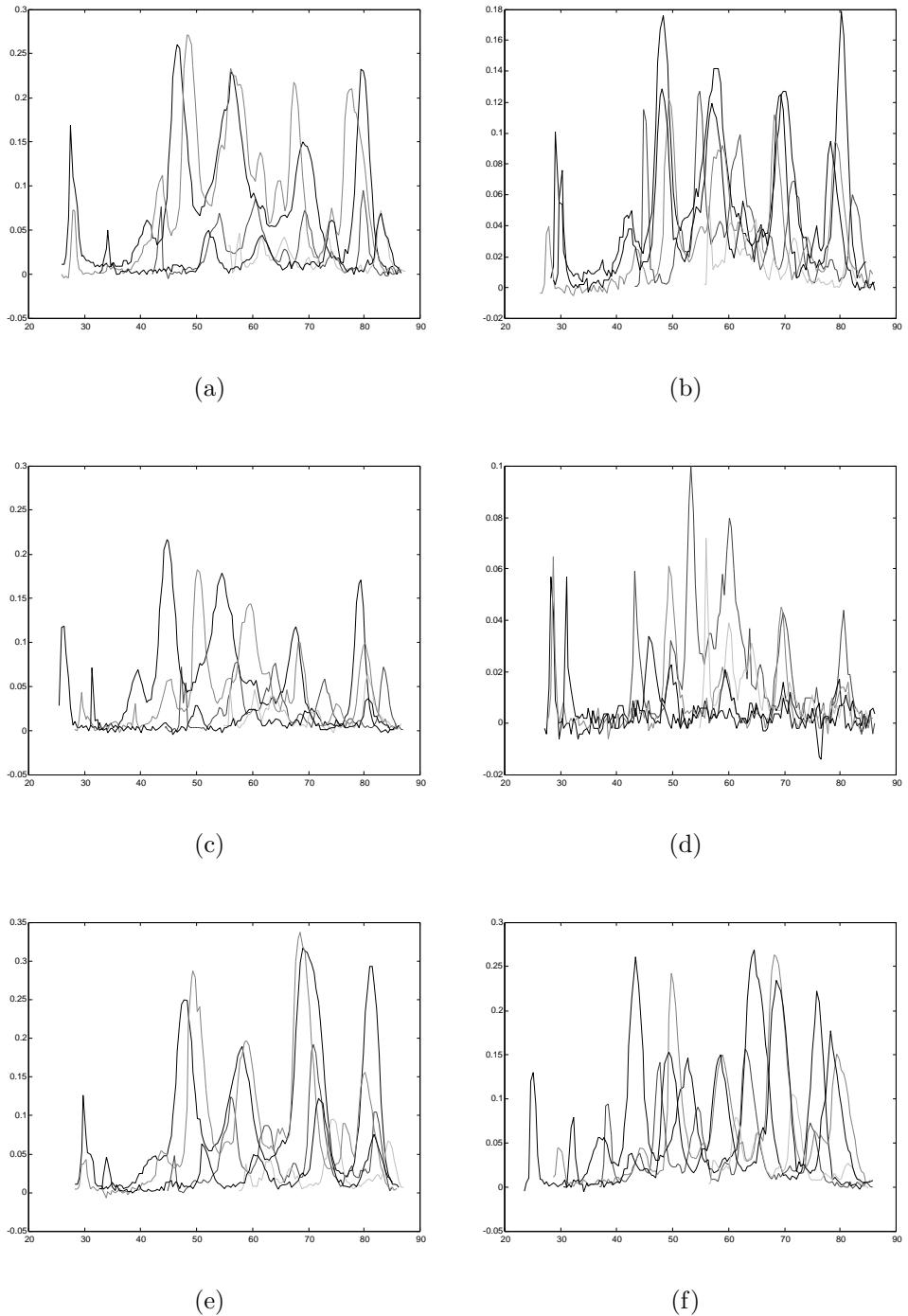


FIG. 4.14: Profil de densité de la ganglioside dans la région : MD (a) 5 Vieux rats (b) 5 Jeunes rats, LC (b) 5 Vieux rats (c) 5 Jeunes rats, HY (d) 5 Vieux rats (e) 5 Jeunes rats

suivant est alors suggéré :

$$Y_{ij}(t) = a_{ij} \frac{d}{dt} F_{ij} \circ h_{ij}(t) + \sigma \epsilon_{ij}(t), \quad (4.4.3)$$

où

- a_{ij} sont des réalisations i.i.d. d'une variable aléatoire utilisée pour modéliser la variation dans la quantité de tissu utilisé pour le TLC.
- $\epsilon_{ij}(t)$ est un bruit blanc et σ un paramètre de bruit. Notons que pour le recalage et le débruitage de ces données, nous avons pris un modèle légèrement différent en considérant que le paramètre de bruit varie selon les expériences et les groupes (i.e. $\sigma = \sigma_{ij}$ dans 4.4.3).
- F_{i1}, \dots, F_{in_i} sont des variables aléatoires i.i.d. qui sont des fonctions de répartitions continûment dérивables, de moyenne F_i et de support commun $[0, T]$ pour $i = 1, 2$. Il n'y a pas de description plus précise de la loi de ces variables aléatoires dans [71].
- Les h_{ij} ($i = 1, 2; j = 1, \dots, n_i$) sont des fonctions aléatoires i.i.d. strictement croissantes et continûment dérивables. Soit $h_{ij} = w_{ij}^{-1}$; w_{ij} sont alors appelées les fonctions de déformation $w_{ij} : [0, T] \rightarrow [0, T_{ij}]$. Pour l'identifiabilité du modèle, il est supposé que $E\{w_{ij}(t)\} = w_0(t)$, où $w_0(t)$ est une fonction connue.

La variabilité de l'action de la capillarité le long de la plaque de gel de silice est modélisée à travers les fonctions de déformation aléatoires. Cette variabilité expérimentale est principalement due à des variations dans les matériaux qui composent la plaque de gel de silice et à des variations dans la composition chimique des solvants utilisés pour faire le mélange de ganglioside. Ainsi, les fonctions de déformations contiennent peu d'informations sur les différences entre les populations, et il est donc essentiel de recaler les profils de densité pour pouvoir les comparer. Les différentes fonctions F_{ij} modélisent les variations dans la composition du cerveau parmi les individus d'une même population. Comme expliqué dans [71], les densités $f_1(t) = \frac{dF_1(t)}{dt}$ et $f_2(t) = \frac{dF_2(t)}{dt}$ sont les moyennes par population des profils de densités dans une situation idéale caractérisée par l'absence de variation dans la quantité de tissu utilisée pour chaque expérience ($a_{ij} = 1$), ou dans l'action de capillarité le long de la plaque de gel de silice ($h_{ij}(t) = w_{ij}(t) = t$). Le modèle 4.4.3 est motivé par le fait que l'*aire* sous un pic des profils de densité, et non la *hauteur* de ce pic, reflète la quantité d'un composant de la ganglioside (voir [71]). Ainsi, l'aire d'un profil densité entre deux points quelconques doit être préservée par le processus de recalage qui synchronise les courbes. Ceci implique que les profils de densités synchronisés sont donnés par $w'_{ij}(t) \hat{Y}_{ij} \circ w_{ij}(t)$ et non simplement par $\hat{Y}_{ij} \circ w_{ij}(t)$, où \hat{Y}_{ij} est une estimation des données bruitées Y_{ij} . Le but de l'étude conduite dans [71] est alors de quantifier les différences entre $f_1(t)$ et $f_2(t)$ après avoir recalé les courbes de la Figure 4.14.

4.2.1.1 Pré-traitement des données

Chaque profil de densité de la Figure 4.14 est une suite d’observations, dont le nombre varie entre 74 et 155 points selon les courbes, collectées sur toute la longueur de la plaque de gel de silice. Comme dans la section 4.1.2, nous avons choisi d’interpoler ces données à l’aide de Splines cubiques pour former des suites d’observations de $n = 128$ points, que nous avons ensuite alignées sur l’intervalle $[0, 1]$ à l’aide d’un recalage linéaire (voir Figure 4.15). Le nombre de points n est proche du nombre de données originales pour chaque profil de densité et la forme des courbes échantillonnées sur $[0, 1]$ correspond à celle des données originales (comparer les Figures 4.14 et 4.15). Le recalage linéaire sur $[0, 1]$ revient à mettre en correspondance les positions de la première et de la dernière observations de chaque profil de densité avec 0 et 1 respectivement. Le support commun des courbes après recalage est donc l’intervalle $[0, 1]$. Afin de supprimer les différences d’amplitude entre les courbes, les profils de densité sont normalisés après recalage de sorte que l’aire sous les courbes soit égale à 1.

4.2.1.2 Recalage à partir des maxima

Les caractéristiques essentielles des profils de densités sont les positions des pics qui correspondent aux différents composantes de la ganglioside. Nous avons donc choisi d’aligner les maxima principaux de ces différentes courbes. Pour chaque groupe, le nombre de fonctions est peu important (seulement 5) et la procédure de recalage (2) peut donc être utilisée (voir section 4.2). Les maxima significatifs de chaque courbe sont estimés à partir des zero-crossings d’une ondelette B-Spline d’ordre $q = 4$ et avec $r = 1$ moment nul. La mise en correspondance des maxima est ensuite effectuée par ϵ -correspondance avec $\epsilon = 4/128$. Les fonctions de déformations à partir de landmarks ont été calculées à l’aide de la méthode de Ramsay [75] avec $\lambda = 10^{-5}$. Pour chaque fonction de déformation u , nous avons également imposé $u(0) = 0$ et $u(1) = 1$. Le résultat de l’alignement des données de la ganglioside est donné dans la Figure 4.16. Avant alignement, les courbes ont été débruitées à l’aide de la procédure TI-H. Toutefois, comme le bruit est très faible, le débruitage change très peu la forme des courbes originales. On peut constater que les maxima qui sont communs à toutes les courbes sont correctement alignés après recalage. La qualité visuelle de l’alignement dépend des caractéristiques communes des courbes. Ainsi l’alignement des profils de densités de la région HY est très satisfaisant car toutes les courbes ont clairement 5 pics principaux dont les formes sont similaires. L’alignement pour les courbes de la zone LC n’est pas autant satisfaisant. Ceci est du au fait que les profils de densités LC n’ont pas tous le même nombre de maxima significatifs. Afin d’améliorer la qualité de l’alignement, certains points d’inflexions ont été utilisés comme landmarks dans [71] lorsque certaines caractéristiques des courbes n’étaient pas préservées lors l’alignement à partir des maxima. Toutefois, les points d’infexion qui doivent être éventuellement utilisés sont déterminés manuellement dans [71], et nous n’avons donc pas considéré cette possibilité.

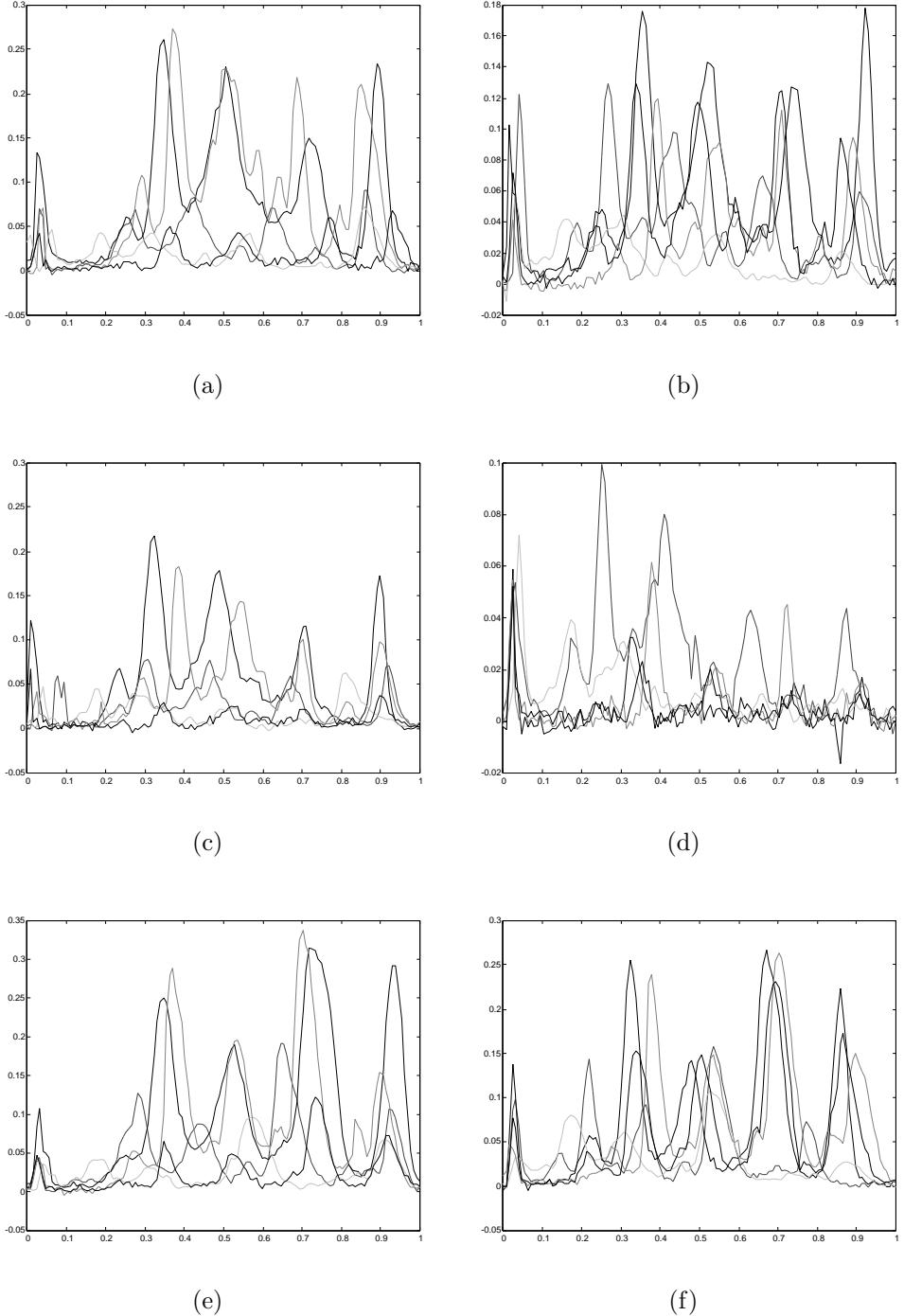


FIG. 4.15: Profil de densité de la ganglioside interpolés avec des Splines cubiques et alignés sur l'intervalle $[0, 1]$ à l'aide d'un recalage linéaire : MD (a) 5 Vieux rats (b) 5 Jeunes rats, LC (b) 5 Vieux rats (c) 5 Jeunes rats, HY (d) 5 Vieux rats (e) 5 Jeunes rats

Remarque 4.2 Pour l’alignement des intensités structurelles, nous avons constaté que la minimisation du problème variationnel (4.4.2) pouvait conduire à des transformations qui ne sont pas strictement croissantes et qui possèdent des “plateaux” ce qui peut complètement “fausser” la mise en correspondance des landmarks. En effet, si G_{z_1} et G_{z_2} sont deux intensités structurelles à aligner et si la déformation $F(\mu)$ (où $\mu = \operatorname{argmin}_{\nu \in M_1} L_{G_{z_1}, G_{z_2}}(F(\nu))$) possède des plateaux, le nombre de modes de $G_{z_2} \circ F(\nu)$ peut être inférieur au nombre de modes de G_{z_2} . Afin d’obtenir des transformations monotones et régulières, nous avons alors choisi de calculer la transformation u qui aligne deux densités structurelles à l’aide de la méthode de Wang et Gasser, dont la description est donnée dans la section 4.1.3.3. L’ ϵ -correspondance s’effectue alors en remplaçant $F(\mu)$ par u dans la Proposition 4.1.

4.2.2 Alignement des données d’orthèse

Les données d’orthèse ont été enregistrées par les Dr. Amarantini David et Dr. Martin Luc (Laboratoire Sport et Performance Motrice, EA 597, UFRAPS, Université de Grenoble, France). Leur étude est motivée par le besoin de mieux comprendre le processus génératrice de mouvement pour différents niveaux d’une force extérieure appliquée sur le genou de plusieurs patients. Dans cette expérience, le moment résultant (qui est un effort exprimé en Newton.mètre) est calculé à l’articulation du genou droit lors d’un mouvement de “piétinement” qui comprend deux phases caractéristiques : une phase d’appui du membre inférieur droit (env. 65 %) et une phase de vol au cours de laquelle le poids du corps est supporté par le membre inférieur gauche. Le moment résultant correspond à la somme algébrique des moments des groupes musculaires agonistes et antagonistes de l’articulation du genou. Un mouvement de “piétinement” est une tâche bien adaptée pour étudier comment les groupes musculaires du genou réagissent à une perturbation extérieure qui peut compliquer le mouvement. Quatre conditions expérimentales ont été testées chez sept jeunes volontaires de sexe masculin : une condition “Contrôle” où les sujets effectuaient la tâche sans contrainte, une condition “Orthèse” où chaque sujet portait une orthèse faite de telle sorte qu’elle n’affecte pas le mouvement à priori et deux conditions pour lesquelles des ressorts de différentes longueurs et raideurs étaient ajoutés (nous noterons ces conditions “Ressort 1” et “Ressort 2”). L’ensemble de l’expérience inclut 10 essais de 20 secondes pour chaque condition expérimentale et pour chaque sujet. L’échantillonnage des données est commencé 5 secondes après le début du “piétinement”, et dure 10 secondes pour chaque essai. Ainsi, les mouvements d’anticipation et de jointure induits par le début du mouvement ne sont pas enregistrés. Pour chacun des sept sujets, dix cycles d’un mouvement de “piétinement” ont été échantillonnés de cette façon pour chaque condition expérimentale. Le moment résultant à l’articulation du genou est enregistré avec une fréquence d’échantillonnage de 200 Hz. De plus amples détails sur la méthode d’acquisition de ces données et sur le calcul du moment résultant peuvent être trouvés dans Cahourt *et. al.* [13].

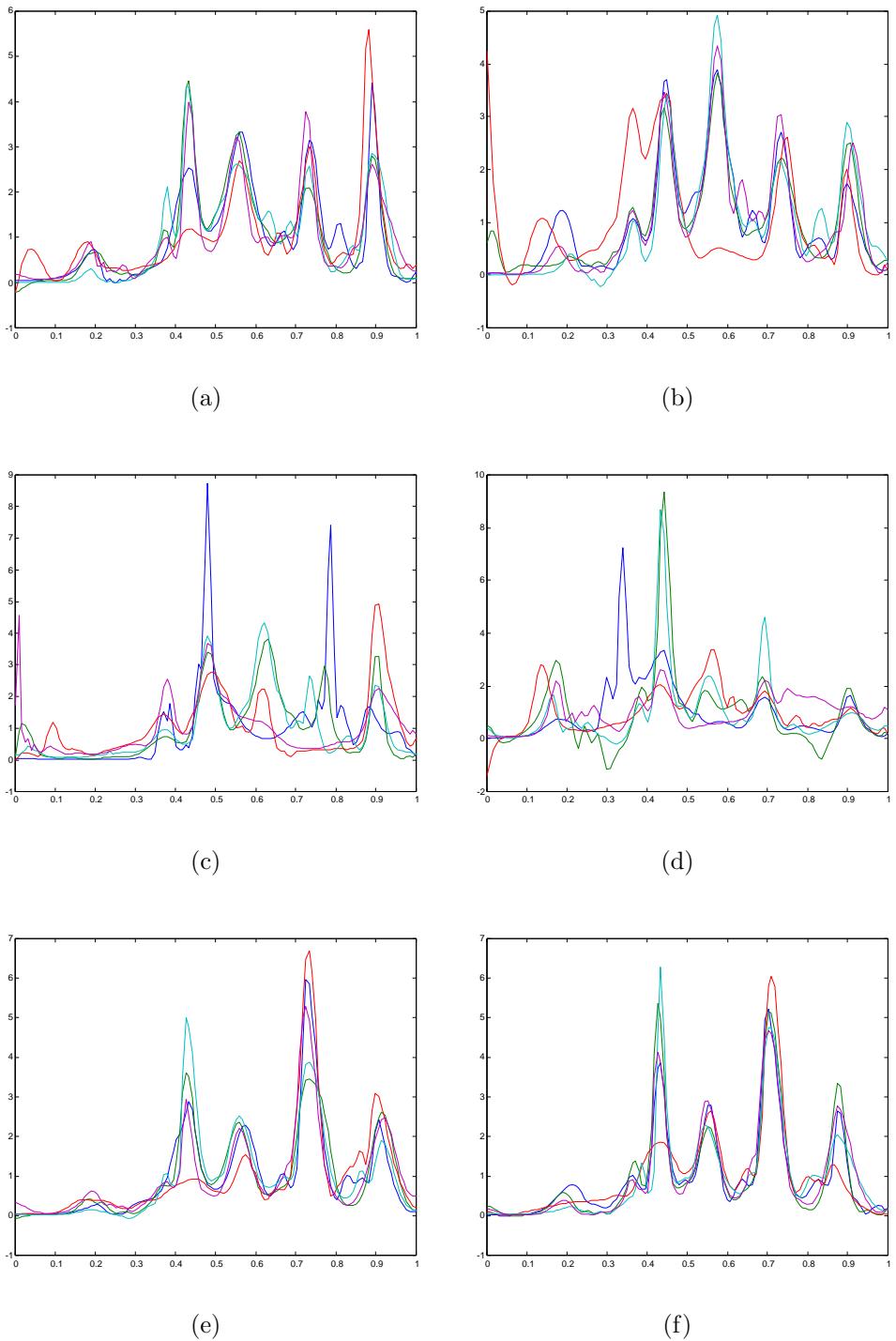


FIG. 4.16: Alignement des données de la ganglioside à partir de leurs maxima : MD
(a) 5 Vieux rats (b) 5 Jeunes rats, LC (b) 5 Vieux rats (c) 5 Jeunes rats, HY (d) 5
Vieux rats (e) 5 Jeunes rats

Pour chaque essai, le moment résultant est calculé pour 256 points régulièrement espacés et dilaté de sorte que l'intervalle de temps corresponde à l'intervalle unité $[0, 1]$. Une observation typique du moment résultant est donc une fonction unidimensionnelle du temps t , où $t \in [0, 1]$. L'ensemble des données d'orthèse se compose donc de 280 courbes qui sont les enregistrements pour les sept sujets sous les quatre conditions expérimentales décrites précédemment, répétés dix fois pour chaque sujet. La Figure 4.17 montre l'ensemble des données obtenues : moment résultant en fonction du cycle d'enregistrement. Pour chaque sujet et chaque condition, on observe que les courbes qui correspondent aux 10 essais répétés présentent des caractéristiques similaires. Ceci est plus visible sur la Figure 4.18 où sont représentées les estimations (par la procédure TI-H) des moments résultants des essais répétés pour trois sujets sous la condition “Contrôle” (première ligne de la Figure 4.17). Pour chaque sujet et chaque condition, les courbes ont la même forme et en général le même nombre d'extrema significatifs. Toutefois, il existe des variations temporelles entre ces courbes qui s'expliquent par le fait que les sujets n'effectuent pas exactement le même mouvement à la même vitesse lors de la répétition des essais pour une condition expérimentale. Afin de pourvoir comparer ces courbes, il est donc nécessaire de corriger ces variations temporelles. Nous avons donc choisi de recalier les courbes pour chaque sujet et chaque condition expérimentale à partir de l'alignement de leur extrema. Les maxima et les minima significatifs des courbes sont estimés à l'aide des zero-crossings d'une ondelette B-Spline d'ordre $q = 4$ et $r = 1$ moment nul. Etant donnée que 28 ensembles de 10 courbes doivent être synchronisées, la procédure de recalage (3) a été utilisée (voir section 4.2) afin de réduire le temps de calcul. La mise en correspondance des extrema est effectuée par ϵ -correspondance MinMax avec $\epsilon = 4/256$ (voir section 4.1.3.2). Les fonctions de déformations à partir de landmarks ont été calculées à l'aide de la méthode de Ramsay [75] avec $\lambda = 10^{-5}$. Pour chaque fonction de déformation u , nous avons également imposé $u(0) = 0$ et $u(1) = 1$. Le résultat de l'alignement des données d'orthèse est donné dans la Figure 4.19. Avant de recalier les courbes, la procédure TI-H a été utilisée pour les débruiter. Afin de juger de la qualité de l'alignement, nous avons également représenté le recalage des courbes débruitées de la Figure 4.18. Nous pouvons constater dans la Figure 4.18, que les extrema des courbes débruitées sont correctement alignés après recalage, ce qui n'est pas le cas pour les courbes débruitées sans recalage.

4.3 Analyse de variance fonctionnelle

Lorsque l'on dispose d'un ensemble données fonctionnelles, il est souvent intéressant de pouvoir expliquer les variations au sein d'une population ou bien de quantifier les différences entre deux sous-ensembles de sujets. Dans le cas des données de la ganglioside, il s'agit de comparer les densités de profils moyens f_1 et f_2 selon l'âge dans chacune des zones du cerveau. Une estimation de f_1 et f_2 peut être obtenue en effectuant la moyenne des données de la ganglioside après recalage. Les profils de densités moyens selon l'âge et la zone du cerveau sont représentés dans la Figure 4.20.

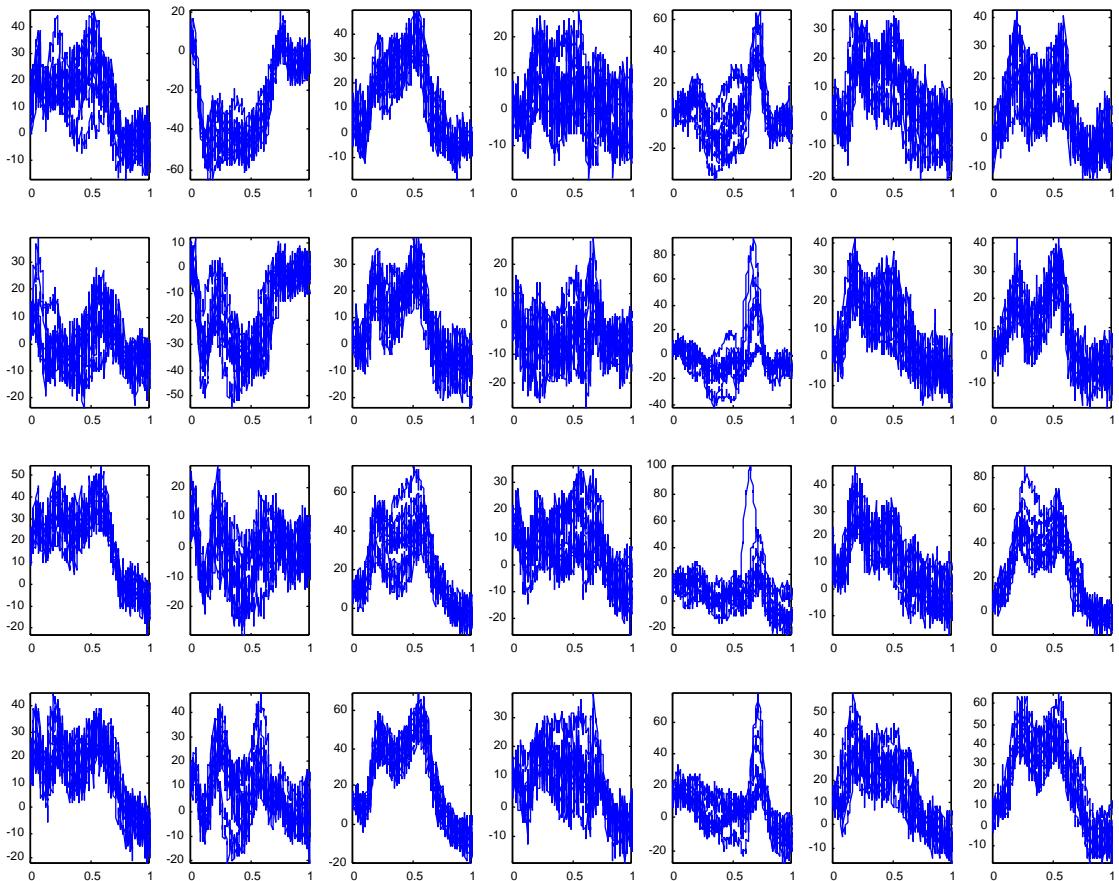


FIG. 4.17: Ensemble des données d'orthèse : les lignes correspondent aux quatre conditions expérimentales et les colonnes aux sept sujets

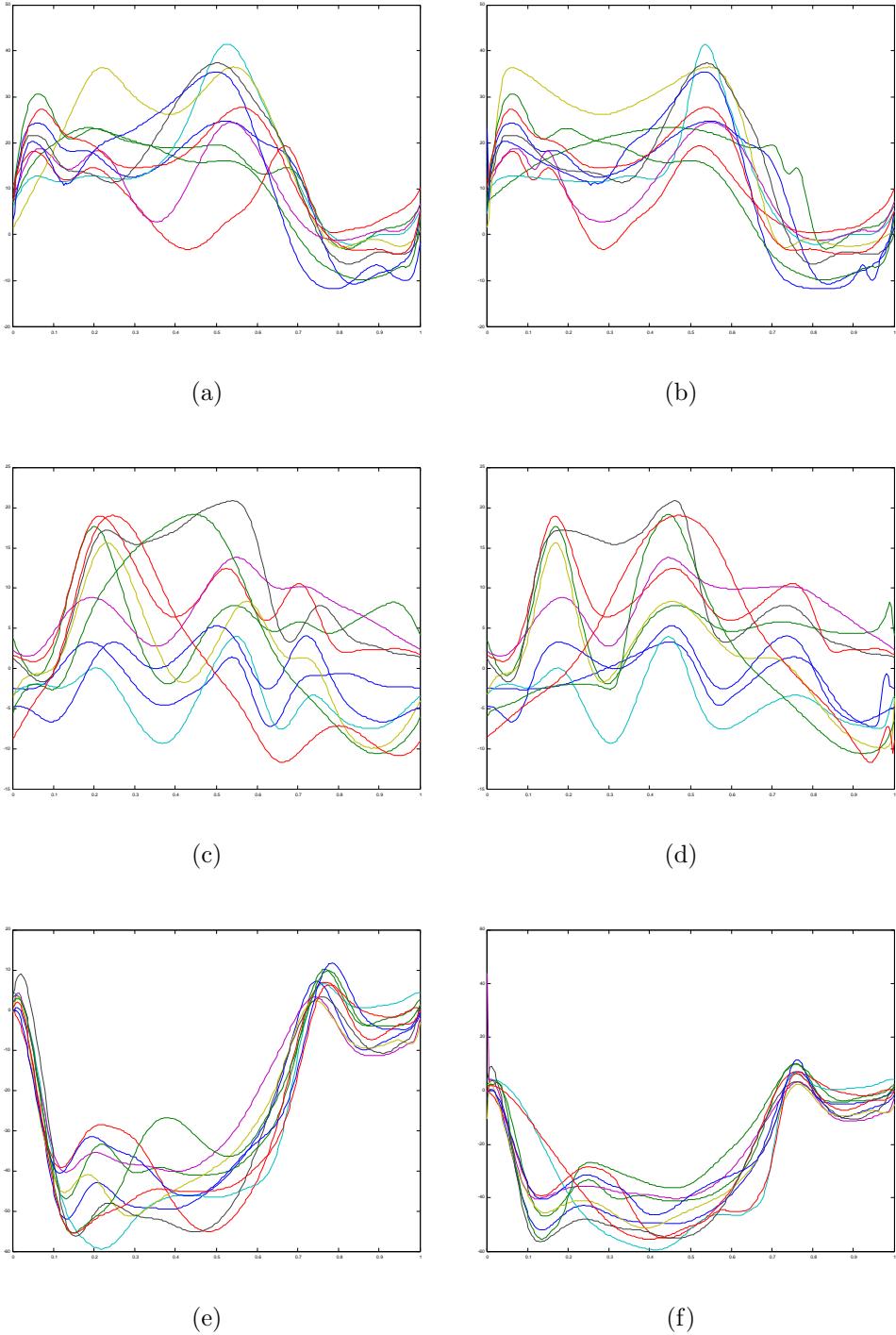


FIG. 4.18: Exemples d’alignement de quelques courbes des données d’orthèse à partir de leurs extréma. (a), (c), (e) : débruitage des moments résultants des essais répétés pour trois sujets sous la condition “Contrôle” sans recalage. (b), (d), (f) : débruitage puis recalage des moments résultants des essais répétés pour trois sujets sous la condition “Contrôle”.

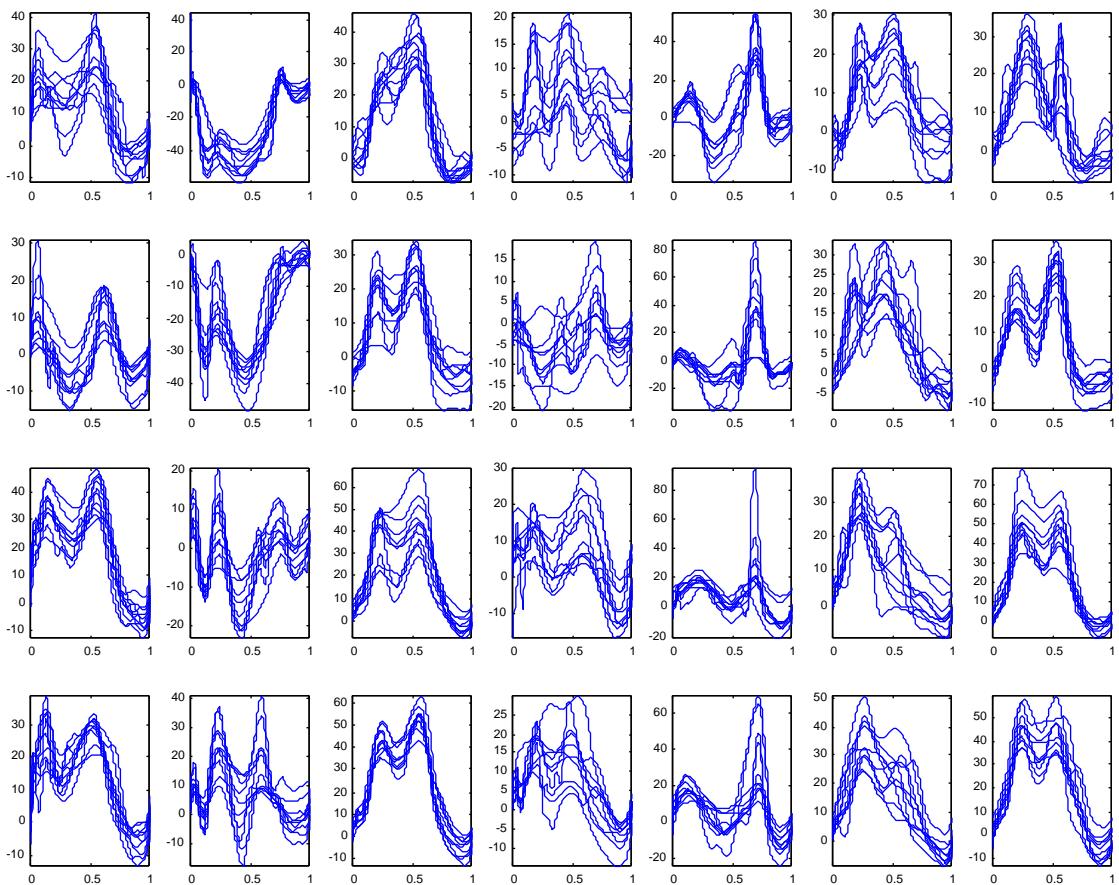


FIG. 4.19: Ensemble des données d'orthèse débruitées et recalées à partir de leurs extréma : les lignes correspondent aux quatre conditions expérimentales et les colonnes aux sept sujets

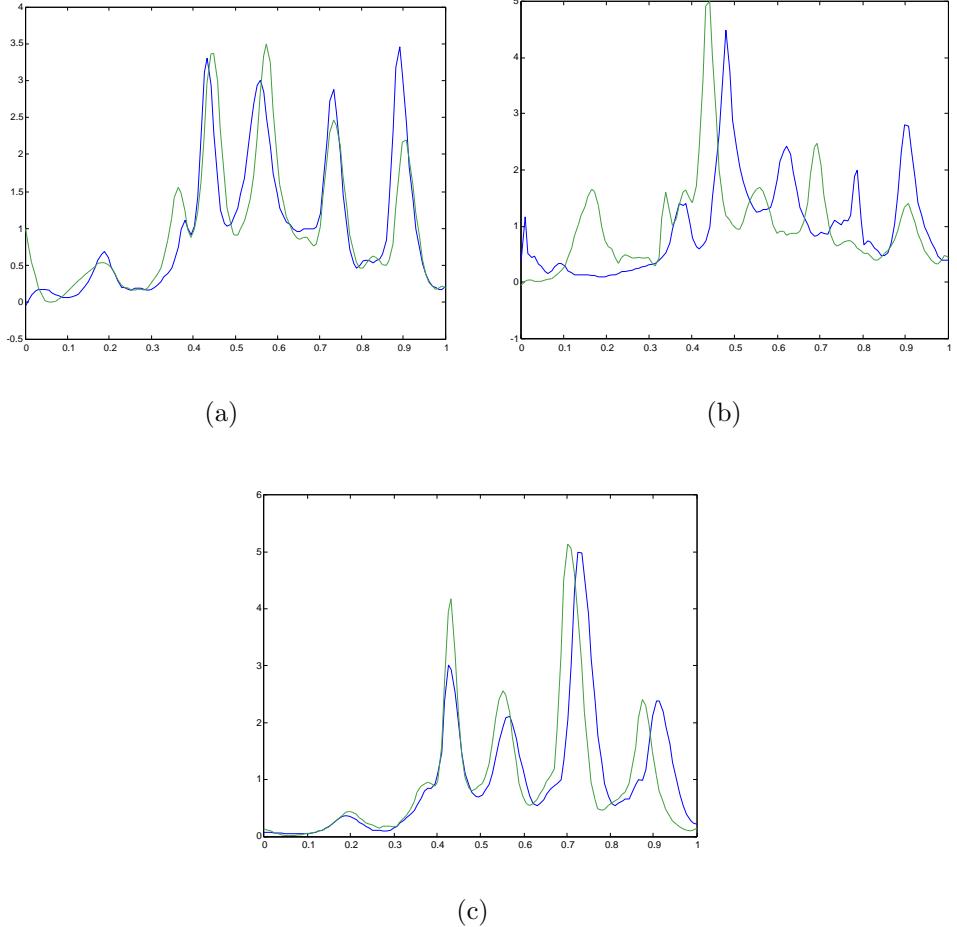


FIG. 4.20: Profil de densité moyen chez les jeunes rats (traits pointillés) et les vieux rats (traits pleins) dans différentes régions du cerveau. (a) MD (b) LC (c) HY.

Afin de tester l'absence de différences dans la composition de ganglioside avec l'âge, les auteurs dans [71] ont proposé de considérer les hypothèses suivantes :

$$\begin{aligned} H_0 &: f_1(t) = f_2(t), \quad t \in \Theta, \\ H_1 &: f_1(t) \neq f_2(t), \quad \text{pour } t_0 \in \Theta, \end{aligned}$$

où Θ est grille de points qui est une discréétisation de l'intervalle $[0, 1]$. Une mesure de similarité entre deux ensembles de courbes dont le support commun est $[0, 1]$ est utilisé pour suggérer trois statistiques (basées sur la moyenne et la variance des corrélations obtenues pour chaque groupe) permettant de tester l'hypothèse nulle d'absence de différence entre les deux groupes. L'indice de similarité choisi est le coefficient de corrélation empirique de Pearson appliqué aux deux vecteurs obtenus en évaluant les profils de densité synchronisés sur la grille commune Θ . En permutant aléatoirement les courbes entre les deux groupes, les p-values de chacune des

trois statistiques sous l'hypothèse nulle sont calculées. Les trois statistiques utilisées rejettent l'hypothèse nulle d'absence de différence entre les vieux et les jeunes rats pour les régions MD et LC au seuil 5%, et ne rejettent pas l'hypothèse nulle pour la région HY au seuil 5%. De plus amples détails sur le calcul de ces statistiques et les résultats obtenus sont disponibles dans [71].

L'approche utilisé dans [71], qui consiste à considérer chaque courbe comme un long vecteur multivarié, présente deux défauts importants : la continuité des données à des points voisins est complètement ignorée et la dimension des vecteurs (ici autour de 100) est typiquement beaucoup plus grande que la taille de l'échantillon (ici 10). Une autre approche naïve consiste à considérer un problème d'analyse de données fonctionnelles comme un problème d'ANOVA univarié pour chaque point $t \in [0, 1]$ et d'effectuer une série de tests en chacun de ces points. L'un des désavantages cruciaux de cette approche est qu'un nombre énorme d'hypothèses (le nombre de points pour chaque courbe peut être d'un millier) doivent être testées simultanément ce qui conduit à un sérieux problème de multiplicité. Ignorer la multiplicité conduit à des erreurs de type I incontrôlées (voir Fan et Li [36] et Abramovich *et. al.* pour une discussion plus détaillée sur les désavantages de ce type d'approches). Une discussion sur les difficultés de la généralisation des idées des procédures de tests multivariés dans le contexte de l'analyse de données fonctionnelles peut également être trouvée dans Faraway [37].

Lorsque l'on doit analyser un ensemble de données fonctionnelles, les méthodes d'analyse de variance fonctionnelle (FANOVA) (voir Ramsay et Silverman [77], et Stone *et. al.* [83] pour une revue de ces techniques) sont mieux adaptées, car elles traitent les observations comme des fonctions et non pas de simples vecteurs multivariés. Bien que de nombreuses méthodes pour construire des modèles de FANOVA et estimer leurs composants aient été proposées (voir e.g. Wahba *et. al.* [88], Stone *et. al.* [83], Huang [47], Lin [57], Gu [44]), il existe peu de travail sur le développement de techniques de tests d'hypothèses dans des modèles de FANOVA. Récemment, Abramovich *et. al.* [1] ont considéré le problème du test d'hypothèses dans un modèle de FANOVA à effets fixés, et obtenu des tests adaptatifs et non-adaptatifs, asymptotiquement optimaux (au sens minimax) pour les effets principaux et les interactions. Les tests développés dans [1] sont basés sur des décompositions en ondelettes des observations qui conduisent à des représentations creuses d'un signal, et permettent donc une diminution significative de la dimension des données originales. Les décompositions en ondelettes permettent également de caractériser différentes conditions de régularité d'une fonction dans de nombreux espaces fonctionnels (voir e.g. Meyer [66]).

4.3.1 Tests d'hypothèses dans un modèle de FANOVA

Nous proposons de décrire brièvement le modèle utilisé dans [1] et d'expliquer les tests proposés, puis d'expliquer comment ces techniques peuvent s'appliquer aux données de la ganglioside et aux données d'orthèse. Supposons que l'on observe une série de réalisations d'un processus stochastique qui peut s'écrire :

$$dY_i(t) = m_i(t) dt + \epsilon dW_i(t), \quad i = 1, \dots, r; \quad t \in [0, 1], \quad (4.4.4)$$

où ϵ est le coefficient de diffusion, r est un entier, m_i sont des fonctions inconnues et W_i sont des processus de Wiener standard indépendants. En pratique, on observe toujours des échantillons discrets de taille n avec un bruit de variance σ^2 , mais comme mentionné dans le chapitre 3, Brown et Low [10] ont prouvé que, sous certaines conditions générales, le modèle discret est asymptotiquement équivalent au modèle de bruit blanc (4.4.4) quand $\epsilon = \sigma/\sqrt{n}$.

Chacune des r fonctions dans le modèle (4.4.4) admet la décomposition unique suivante :

$$m_i(t) = m_0 + \mu(t) + a_i + \gamma_i(t), \quad i = 1, \dots, r; \quad t \in [0, 1], \quad (4.4.5)$$

où m_0 est une constante (*la moyenne totale*), $\mu(t)$ est soit égale à zéro soit une fonction non-constante du temps t (*le principal effet de t*), a_i est soit égal à zéro soit une fonction non-constante de i (*l'effet principal de i*) et $\gamma_i(t)$ est soit égale zéro soit une fonction non-nulle qui ne peut pas être décomposée comme la somme d'une fonction de i et d'une fonction de t (*la composante d'interaction*). Les composants de la décomposition (4.4.5) vérifient également les conditions d'identifiabilité suivantes :

$$\begin{aligned} \int_{[0,1]} \mu(t) dt &= 0, & \sum_{i=1}^r a_i &= 0, \\ \sum_{i=1}^r \gamma_i(t) &= 0, & \int_{[0,1]} \gamma_i(t) dt &= 0, & \forall i &= 1, \dots, r; \quad t \in [0, 1]. \end{aligned}$$

Dans ce modèle de FANOVA à effets fixés, il est intéressant de pouvoir tester l'importance des effets principaux et des interactions. Dans ce but, Abramovich *et. al.* [1] proposent de considérer les tests d'hypothèses fonctionnels suivants :

$$H_0 : \quad \mu(t) \equiv 0, \quad t \in [0, 1], \text{ pas de tendance globale} \quad (4.4.6)$$

$$H_0 : \quad \gamma_i(t) \equiv 0, \quad \forall i = 1, \dots, r; \quad t \in [0, 1]. \text{ pas de différence de forme.} \quad (4.4.7)$$

Aucune forme paramétrique n'est spécifiée pour μ et γ_i pour les hypothèses alternatives. L'ensemble d'hypothèses alternatives est construit de manière à ce qu'il soit suffisamment séparé des hypothèses nulles (4.4.6) et (4.4.7) au sens d'une distance ρ dans $L^2([0, 1])$. Afin de tester une large classe d'alternatives possibles, il est supposé que les m_i (et donc aussi μ et γ_i) appartiennent une boule de Besov de rayon $C > 0$

sur $[0, 1]$, $B_{p,q}^s(C)$ où $s > 0$ et $1 \leq p, q \leq \infty$. Pour certains choix des paramètres s, p et q les espaces de Besov incluent de nombreux espaces fonctionnels (e.g. Hölder, Sobolev, BV). Nous renvoyons à Meyer [66] pour une définition rigoureuse et une étude détaillée de ces espaces. Les hypothèses alternatives suivantes ont alors été considérées :

$$\begin{aligned} H_1 : \quad & \mu \in F(\rho), \\ H_1 : \quad & \gamma_i(t) \in F(\rho), \text{ pour au moins un } i = 1, \dots, r. \end{aligned}$$

où $F(\rho) = \{f \in B_{p,q}^s(C) : \|f\|_2 \geq \rho\}$ et $\|\cdot\|_2$ est la norme usuelle sur $L^2([0, 1])$.

La moyenne des r réalisations dans le modèle de FANOVA (4.4.4)-(4.4.5) conduit à (étant donné les conditions d'identifiabilité (4.4.6) et (4.4.6) pour les composants de m_i) :

$$\begin{aligned} d\bar{Y}(t) = \frac{1}{r} \sum_{i=1}^r dY_i(t) &= \left(\frac{1}{r} \sum_{i=1}^r m_i(t) \right) dt + \epsilon d\bar{W}(t) \\ &= (m_0 + \mu(t))dt + \epsilon d\bar{W}(t); \quad t \in [0, 1], \end{aligned} \quad (4.4.8)$$

où \bar{W} est la moyenne des r processus de Wiener indépendants sur $[0, 1]$. L'équation (4.4.8) implique également que :

$$d(Y_i - \bar{Y})(t) = a_i + \gamma_i(t)dt + \epsilon d(W_i - \bar{W})(t), \quad i = 1, \dots, r; \quad t \in [0, 1]. \quad (4.4.9)$$

Ainsi, le modèle (4.4.8) et chacun des r modèles donnés par l'équation (4.4.9) peuvent être écrits sous la forme générale :

$$dZ(t) = f(t)dt + \eta dW(t), \quad t \in [0, 1], \quad (4.4.10)$$

où $Z(t) = \bar{Y}(t)$, $f(t) = m_0 + mu(t)$, $\eta = \epsilon/\sqrt{r}$ et $Z(t) = (Y_i - \bar{Y})(t)$, $f(t) = a_i + \gamma_i(t)$, $\eta = \epsilon\sqrt{(r-1)/r}$ pour les modèles (4.4.8) et (4.4.9) respectivement. Dans les deux cas, sous les hypothèses nulles (4.4.6) et (4.4.7), f est une fonction constante. Les auteurs dans [1] ont donc cherché à développer des procédures pour tester les hypothèses :

$$H_0 : f \equiv \int_0^1 f(t)dt \text{ contre } H_1 : \left(f - \int_0^1 f(t)dt \right) \in F(\rho),$$

pour le modèle général (4.4.10). Ces tests sont basés sur des décompositions en ondelettes de la fonction f . Le problème du test de l'hypothèse nulle $H_0 : f \equiv \int_0^1 f(t)dt$ peut être transformé en un problème de test d'un signal nul dans un modèle Gaussien "signal plus bruit" lorsque l'on s'intéresse aux coefficients d'ondelettes. Nous ne présentons pas la méthodologie développée dans [1] en détail. En particulier, nous ne donnons pas les formules explicites des statistiques de test utilisées. Les propriétés asymptotiques de ces tests sont également étudiées en détail dans Abramovich *et. al.*

[1] et nous renvoyons le lecteur à cet article pour une définition précise de l’optimalité de ces tests. Nous mentionnerons juste que ces tests font intervenir deux entiers $j(s)$ et j_η qui définissent les niveaux de résolution à utiliser dans la décomposition en ondelettes des données observées pour calculer les statistiques de tests. En pratique, ces entiers doivent être estimés. Toutefois, plutôt qu’essayer de définir une méthode automatique pour calculer $j(s)$ et j_η , les auteurs ont effectué quelques simulations numériques pour étudier les performances de ces tests, et en ont déduit les valeurs appropriées de $j(s)$ et j_η en fonction de la taille de l’échantillon n . Nous finissons cette courte présentation du modèle de FANOVA développé dans [1], en précisant que dans la plupart des applications, la variance du bruit σ^2 est inconnue et que son estimation est cruciale pour le succès des tests d’hypothèses fonctionnels proposés. Rappelons que nous avons expliqué dans le chapitre 3 que dans un problème de débruitage par ondelettes, σ pouvait être estimé en divisant la médiane de la valeur absolue des coefficients d’ondelettes au niveau de résolution de plus fin, par 0.6745 (estimateur robuste MAD, voir Donoho et Johnstone [30]). Dans le modèle (4.4.4), σ est donc obtenu en moyennant les r estimation robustes (MAD) du bruit qui correspondent à chacune des courbes individuelles.

4.3.2 Application à deux exemples réels

4.3.2.1 Données d’orthèses

A titre d’illustration, Abramovich *et. al.* [1] ont appliqué leur modèle de FANOVA aux données d’orthèse, mais **sans recalage préalable**. Etant donné que le but de cette étude est de comprendre comment un individu se comporte en fonction d’une perturbation extérieure, il est nécessaire de pouvoir quantifier les différences entre les moyennes des courbes individuelles selon les conditions expérimentales. Le modèle suivant est proposé dans [1] pour les données d’orthèse :

$$dY_{ijk} = m_{ij}(t) dt + \epsilon dW_{ijk}(t), \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad k = 1, \dots, K; \quad t \in [0, 1],$$

avec

$$m_{ij}(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t) + \beta_j + \delta_j(t), \quad i = 1, \dots, I; \quad j = 1, \dots, J; \quad t \in [0, 1],$$

où i est l’indice de condition, j est l’indice du sujet, k est l’indice de répétition et t est une variable temporelle. Les sujets dans le modèle ci-dessus sont naturellement considérés comme des effets de blocks. En effet, les sujets ne se comportent pas de la même façon pas, mais les chercheurs qui ont mené cette étude ne sont pas intéressés par la quantification de ces différences. Traiter les sujets de la même façon permet d’être plus précis dans l’évaluation des conditions expérimentales puisque la variabilité due aux effets de blocks est exclue (dans cet exemple, il y a $I = 4$ conditions). Ainsi, si l’on suppose les conditions d’identifiabilité standards (du type (4.4.6) et (4.4.6)), il est possible d’utiliser les tests d’hypothèses fonctionnels

développés dans [1] pour le modèle “moyenné” (sur les 7 sujets et les 10 essais répétés) :

$$d\bar{Y}_{i..} = m_i(t) dt + \eta dW_{i..}(t), \quad i = 1, \dots, I; \quad t \in [0, 1],$$

avec

$$m_i(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t), \quad i = 1, \dots, I; \quad t \in [0, 1],$$

où $\eta = \epsilon/\sqrt{JK}$. Etant donné que le nombre de points d'échantillonnages est de 256 (i.e. 7 niveaux de résolution dans la décomposition en ondelettes de ces signaux), les auteurs dans [1] recommandent de prendre $j(s) = 4$ et $j_\eta = 6$ (voir [1] pour une définition précise de ces entiers et comment les choisir en pratique). Comme suggéré dans [1], nous avons pris l'ondelette à support compact *Coiflet 18-tap filter* pour l'analyse de ces signaux, et avons utilisé la moyenne des estimations de l'écart-type du bruit via MAD pour chaque condition, comme estimateur de ϵ/\sqrt{JK} . Les hypothèses nulles suivantes ont été testées dans [1] (contre des alternatives du type $H_1 : \mu \in F(\rho)$ pour un $\rho > 0$) :

1. $H_0 : \mu \equiv 0$.
2. $H_0 : \gamma_i \equiv 0 \quad i = 1, \dots, 4$.
3. $H_0 : \gamma_1 - \gamma_2 \equiv 0$: les effets des conditions “Contrôle” et “Orthèse” sont les mêmes.
4. $H_0 : \gamma_3 - \gamma_4 \equiv 0$: les effets des conditions “Ressort 1” et “Ressort 2” sont les mêmes.

Comme indiqué précédemment, l'analyse des données d'orthèse effectuées dans [1] a été faite sans recalage préalable des courbes. Afin d'illustrer l'intérêt de l'alignement de signaux, nous avons comparé les résultats obtenus dans [1] avec ceux obtenus en recalant au préalable les courbes. Dans la section précédente, nous avons précisé comment les extrema des courbes des données d'orthèse pouvaient être alignés pour chaque sujet et chaque condition. Toutefois, pour pouvoir être alignés, les signaux ont été débruités ; donc, afin de pouvoir utiliser un modèle de FANOVA sur les données recalées, il est nécessaire de **re-bruiter** les courbes après alignement. En effet, en l'absence de bruit, toutes les courbes sont différentes et il n'y a pas d'inférence possible. Pour chaque courbe, le niveau de bruit a été estimé par MAD (en utilisant l'ondelette Symmlet 8), puis une séquence de bruit Gaussien de même variance a été ajoutée à chaque courbe estimée et recalée. Les signaux alignés puis re-bruités de cette façon sont donnés dans la Figure 4.21. Les estimateurs empiriques de la moyenne globale et des effets de chaque traitement sont donnés dans la Figure 4.22 pour les données non-recalées et les données recalées/re-bruitées. On peut constater que l'alignement des données modifie localement la forme des courbes de la moyenne générale et des effets de chaque traitement. Pour les données non-recalées et les données recalées/re-bruitées, les hypothèses nulles $H_0 : \mu \equiv 0$ et $H_0 : \gamma_i \equiv 0 \quad i = 1, \dots, 4$ sont clairement rejetées avec une p -value inférieure à 10^{-3} dans les deux cas. Pour les données non-recalées, les p -values des tests $H_0 : \gamma_1 - \gamma_2 \equiv 0$ et

$H_0 : \gamma_3 - \gamma_4 \equiv 0$ sont respectivement 0.097 et 0.487, alors que pour les données recalées/re-bruitées, les p -values des tests $H_0 : \gamma_1 - \gamma_2 \equiv 0$ et $H_0 : \gamma_3 - \gamma_4 \equiv 0$ sont respectivement 0.009 et 0.011. Au seuil $\alpha = 5\%$ et pour les données non-recalées, on conclue donc que sous les conditions “Contrôle” et “Orthèse”, ou sous les conditions “Ressort 1” et “Ressort 2”, les sujets se comportent de manière similaire. Ceci indiquerait donc que les individus ajustent leur posture de manière similaire sous l’effet de perturbations de même nature. Toutefois, si l’on recalcule au préalable les essais pour chaque sujet et pour chaque condition, on conclue au contraire (au seuil $\alpha = 5\%$), que les sujets se comportent différemment sous les conditions “Contrôle” et “Orthèse”, ou sous les conditions “Ressort 1” et “Ressort 2”.

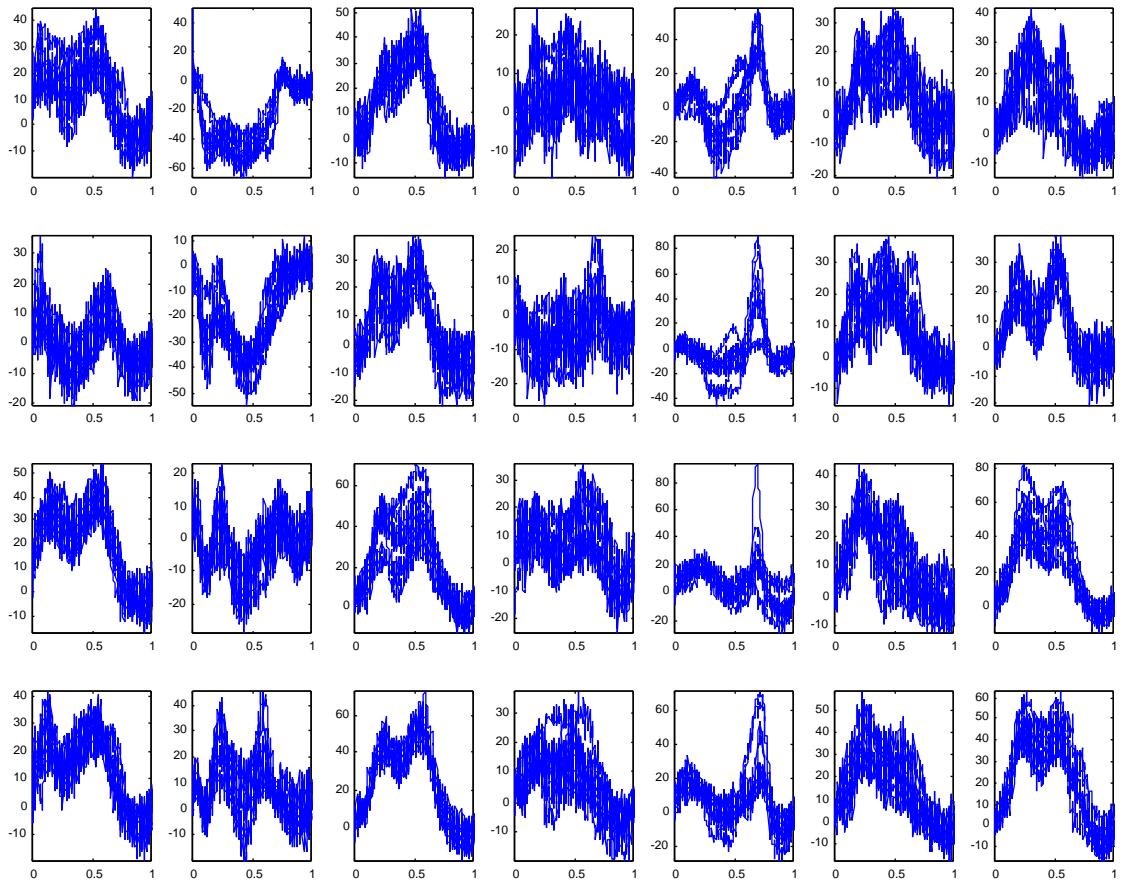


FIG. 4.21: Ensemble des données d’orthèse débruitées, recalées à partir de leurs extrêmes, puis rebruitées avec le même niveau de bruit que pour les données originales : les lignes correspondent aux quatre conditions expérimentales et les colonnes aux sept sujets

4.3.2.2 Données de la ganglioside

Dans l’étude des données de la ganglioside, on souhaite savoir si la forme des profils de densité varie avec l’âge en fonction de la région du cerveau, ce qui peut

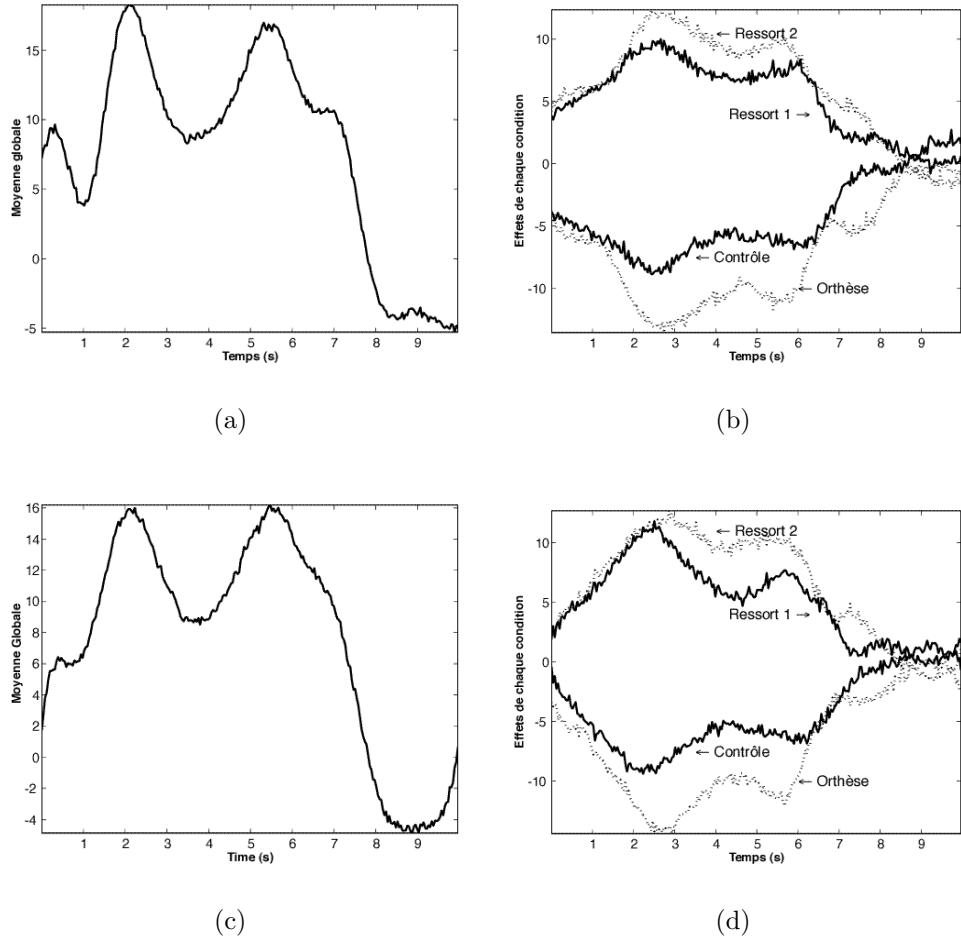


FIG. 4.22: Données d'orthèse : (a) Estimation empirique de la moyenne globale $m_0 + \mu(t)$ et (b) estimateurs empiriques des effets de chaque conditions (i.e $\alpha_i + \gamma_i(t)$) sans recalage préalable des données. (c) Estimation empirique de la moyenne globale $m_0 + \mu(t)$ et (d) estimateurs empiriques des effets de chaque conditions (i.e $\alpha_i + \gamma_i(t)$) avec recalage/re-bruitage des données.

s'effectuer en quantifiant les différences entre les moyennes des profils de densité donnés dans la Figure 4.20. Pour chaque région du cerveau, nous avons utilisé le même modèle de FANOVA que celui proposé pour les données d'orthèse en prenant $I = 2$ (jeunes ou vieux rats), $J = 5$ (il y a 5 jeunes rats et 5 vieux rats) et $K = 1$ (pas de répétitions d'expérience par sujet). Une fois alignées, les données ont été rebrutées à partir des niveaux de bruit originaux comme expliqué précédemment. Pour chaque région, nous avons estimé la moyenne globale et les effets de l'âge (voir Figure 4.23). Afin de tester si la distribution de la ganglioside varie avec l'âge, il est intéressant de tester les hypothèses nulles $H_0 : \gamma_1 - \gamma_2 \equiv 0$ pour chaque zone du cerveau. En prenant $j(s) = 4$ et $j_\eta = 5$, les hypothèses sont largement rejetées (p -value inférieure à 10^{-3}) pour les trois régions MD, LC et HY. Ceci était prévisible car le niveau de bruit dans ces données est extrêmement faible. Ainsi, contrairement aux conclusions de Munoz Maldonado *et. al.* [71], notre étude indique que la distribution de la ganglioside varie avec l'âge dans les régions MD, LC et HY. Toutefois, notre modélisation n'est peut être pas adaptée car en inspectant les profils de densités moyens de la Figure 4.20, il semble que la distribution de la ganglioside, chez les jeunes et les vieux rats, soit pratiquement la même pour la région HY, ce qui était une des conclusions de l'analyse statistique effectuée dans [71].

4.4 Discussions

Nous terminons ce chapitre en discutant quelques aspects de notre méthode d'alignement automatique à partir de landmarks que nous n'avons pas abordés jusqu'à présent :

- Dans le chapitre 3, une méthode de détection des singularités d'un signal a été proposée. Nous n'avons pas cherché à définir une méthode de mise en correspondance des singularités éventuelles de deux fonctions devant être alignées, car nous n'avons pas trouvé d'exemples réels de recalage de signaux présentant des singularités. Toutefois, nous pensons qu'il est possible de reprendre la méthodologie développée pour les lignes de zero-crossings, pour adapter l' ϵ -correspondance à l'alignement automatique des singularités de deux signaux (en utilisant en particulier l'intensité structurelle pondérée des lignes de maxima d'ondelettes, car les amplitudes des modes sont alors caractéristiques de l'ordre des singularités d'une fonction).
 - Nous avons montré à l'aide de simulations que notre méthode de recalage automatique conduit à un meilleur alignement des landmarks que la méthode de Wang et Gasser ([91], [92]) quand un nombre suffisant de points caractéristiques peut être estimé. Wang et Gasser ([91], [92]) ont proposé d'incorporer les estimations des dérivées des signaux dans la fonction de coût associée à un problème d'alignement global, afin de mettre en correspondance les extrema communs.
- Toutefois, le recalage par DTW peut conduire à un alignement des extrema qui n'est visuellement pas satisfaisant. En effet, dans les Figures 4.24 et 4.25, nous

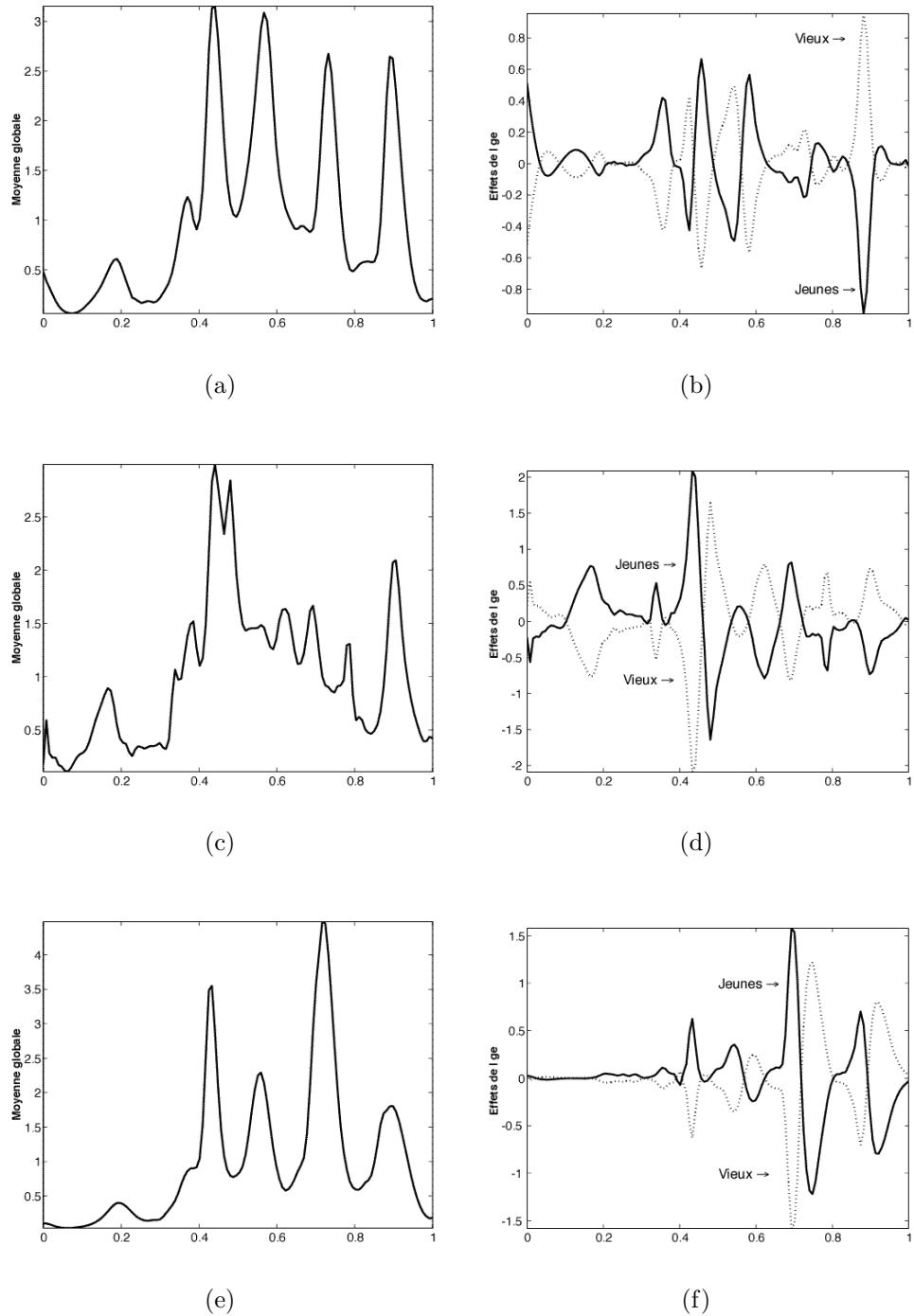


FIG. 4.23: Données de la ganglioside : estimation empirique de la moyenne globale $m_0 + \mu(t)$ pour chaque région : (a) MD, (c) LC, (e) HY. Estimateurs empiriques des effets de chaque conditions (i.e $\alpha_i + \gamma_i(t)$) pour chaque région : (b) MD, (d) LC, (f) HY.

donnons un exemple de recalage où notre méthode d’alignement automatique à partir de landmarks donne un alignement des extrema qui est plus satisfaisant que celui obtenu avec la méthode de Wang et Gasser ([91], [92]). Les courbes des Figures 4.24a et 4.25a sont des enregistrements de sons de poissons que l’on souhaite utiliser pour un problème de discrimination entre différentes espèces ou différents individus (nous dénoterons ces fonctions par “courbe 4”, “courbe 6” et “courbe 7”). Ces données sont présentées en détail dans Wood [93]. Afin de pouvoir classifier les sons de poissons, il est nécessaire de déterminer la forme caractéristique du son émis par un poisson ou une espèce. Il s’agit donc d’un problème qui peut se résoudre par des méthodes de recalage. Dans les Figures 4.24c et 4.25c, nous avons représenté la différence entre l’intensité structurelle des lignes de zero-crossings qui correspondent aux maxima locaux et l’intensité structurelle des lignes de zero-crossings qui correspondent aux minima locaux de chacune des fonctions (i.e $\hat{G}_{z1}^* = \hat{G}_{z1}^+ - \hat{G}_{z1}^-$ et $\hat{G}_{z2}^* = \hat{G}_{z2}^+ - \hat{G}_{z2}^-$, voir section 4.1.3.3). Les “pics positifs” des Figures 4.24c et 4.25c correspondent donc à des maxima significatifs des fonctions, alors que les “pics négatifs” correspondent à des mimina significatifs. L’alignement automatique à partir de landmarks par ϵ -correspondance MinMax des courbes des Figures 4.24a et 4.25a avec débruitage par TI-H est donné dans les Figures 4.24b et 4.25b, et l’alignement par DTW (algorithme de Wang et Gasser avec débruitage par une méthode du noyau) est donné dans les Figures 4.24d et 4.25d. On peut constater que le recalage à partir de landmarks donne un alignement correct des “extrema communs” des courbes, qui correspond à celui que l’on ferait manuellement, alors que le recalage par DTW conduit à un alignement des extrema qui n’est pas satisfaisant. En effet, il est clair que la courbe 4 a deux maxima significatifs aux points $x = 0.76$ et $x = 0.95$ dont les formes suggèrent qu’ils doivent être associés avec les deux maxima significatifs de la courbe 6 aux points $x = 0.63$ et $x = 0.8$ respectivement. Ces maxima sont correctement alignés par notre méthode de recalage à partir de landmarks, alors que l’alignement par DTW associe le maxima de la courbe 4 au point $x = 0.76$ avec le maxima de la courbe 6 au point $x = 0.8$. La même remarque est également vraie pour l’alignement des courbes 4 et 7. L’alignement par DTW associe le maxima de la courbe 4 au point $x = 0.76$ avec le maxima de la courbe 7 au point $x = 0.69$, alors que manuellement on aurait plutôt mis en correspondance le maxima de la courbe 4 au point $x = 0.76$ avec le maxima de la courbe 7 au point $x = 0.55$.

Ainsi, l’alignement par DTW peut conduire à un “alignement erroné” des extrema communs de deux fonctions. De plus, la méthode de Wang et Gasser ([91], [92]) ne garantie pas un alignement correct des points d’infexion, et ne prend pas en compte la possibilité d’aligner des fonctions à partir de leurs singularités éventuelles. L’intérêt de notre méthode est donc de pouvoir spécifiquement déterminer les points caractéristiques qui doivent se correspondre en utilisant l’alignement automatique des extrema des intensités structurelles, ce qui peut se résoudre efficacement par DTW car la forme des pics des intensités structurelles reflète bien la structure des landmarks d’un

signal.

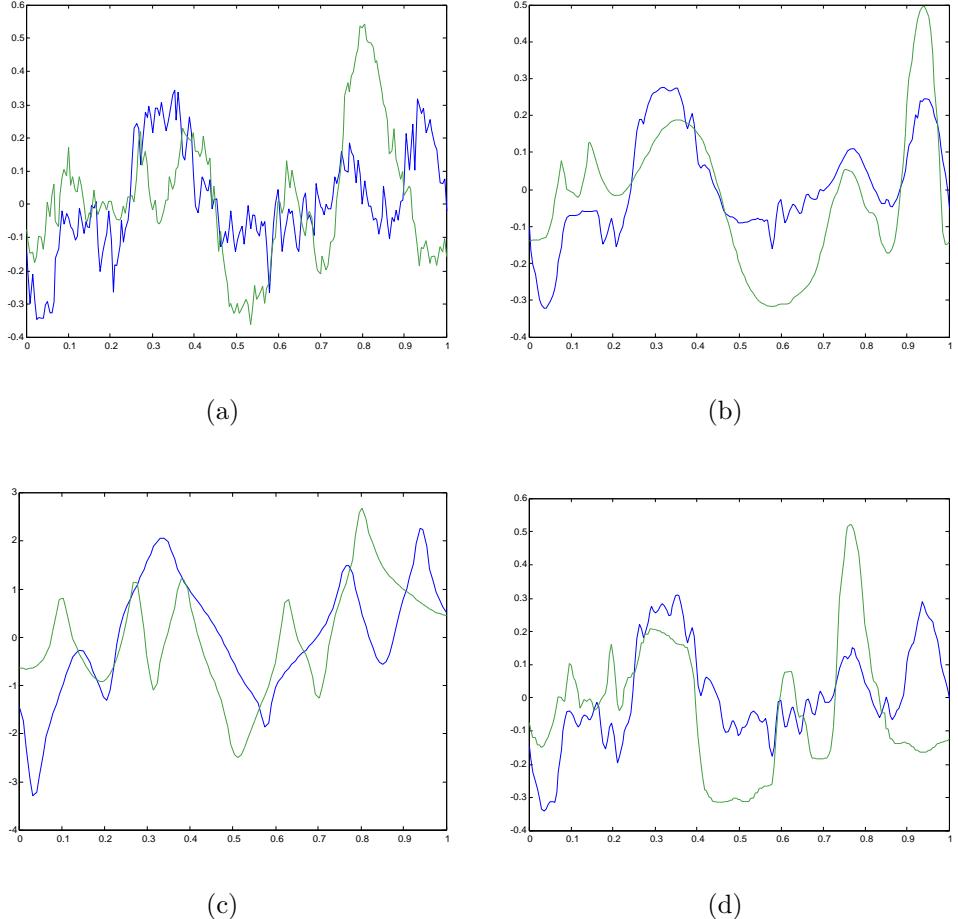


FIG. 4.24: Alignement de sons de poissons, courbe 4 (trait plein) et courbe 6 (trait pointillé) : (a) Courbes bruitées (b) Alignement automatique à partir des extrema, débruitage par TI-H (c) Différence entre l'intensité structurelle des lignes de zero-crossings qui correspondent aux maxima locaux et l'intensité structurelle des lignes de zero-crossings qui correspondent aux minima locaux de chacune des courbes (i.e $\hat{G}_{z4}^* = \hat{G}_{z4}^+ - \hat{G}_{z4}^-$ (trait plein) et $\hat{G}_{z6}^* = \hat{G}_{z6}^+ - \hat{G}_{z6}^-$ (trait pointillé), voir section 4.1.3.3). (d) Alignement par DTW, débruitage par une méthode du noyau

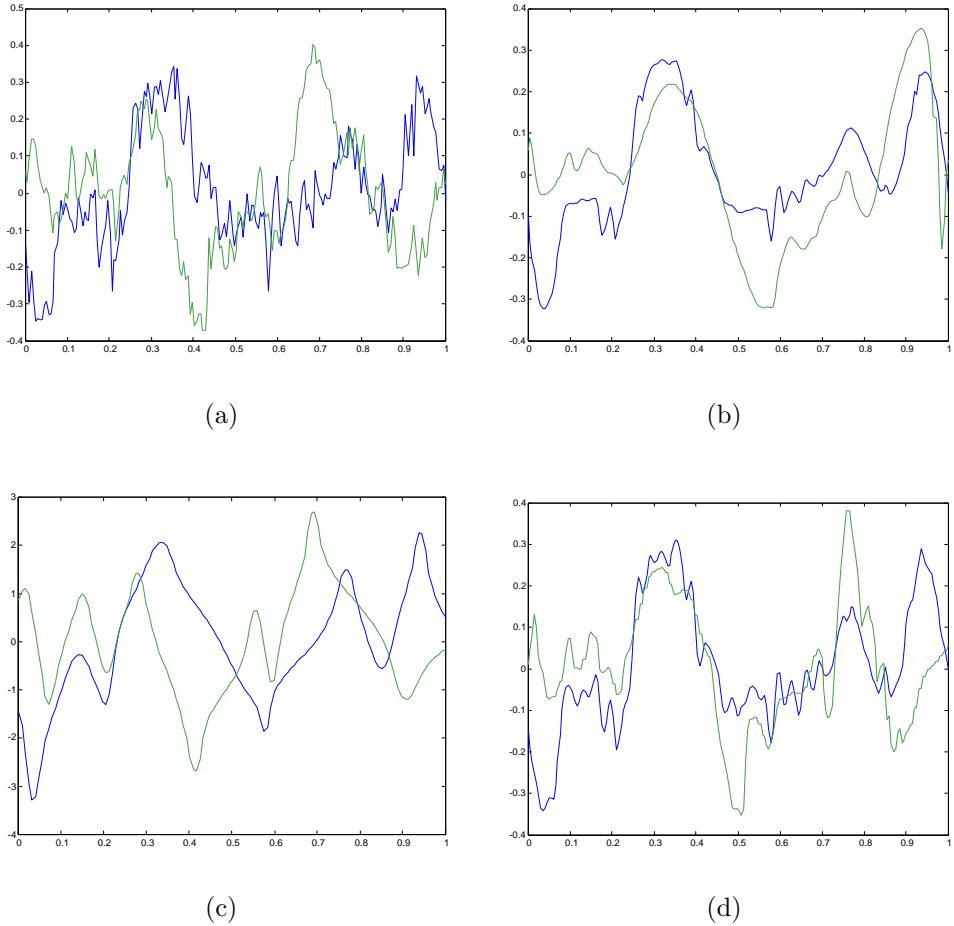


FIG. 4.25: Alignement de sons de poissons, courbe 4 (trait plein) et courbe 7 (trait pointillé) : (a) Courbes bruitées (b) Alignement automatique à partir des extrema, débruitage par TI-H (c) Différence entre l'intensité structurelle des lignes de zero-crossings qui correspondent aux maxima locaux et l'intensité structurelle des lignes de zero-crossings qui correspondent aux minima locaux de chacune des courbes (i.e $\hat{G}_{z4}^* = \hat{G}_{z4}^+ - \hat{G}_{z4}^-$ (trait plein) et $\hat{G}_{z7}^* = \hat{G}_{z7}^+ - \hat{G}_{z7}^-$ (trait pointillé), voir section 4.1.3.3). (d) Alignement par DTW, débruitage par une méthode du noyau

Chapitre 5

Détection de contours dans une image et alignement de signaux 2D

Les contours sont souvent les caractéristiques les plus importantes dans une image. Ainsi, la qualité visuelle de la synchronisation de deux ou plusieurs images dépend principalement de la qualité de l'alignement de leurs “contours communs” après recalage. Dans le contexte de l'alignement global de signaux 2D, des mesures de similarité qui prennent en compte les contours pour déterminer la qualité du recalage d'images, ont été considérées par Bajcsy et Kovacic [4] et Moshfeghi [69] pour aligner des images médicales, et par Li *et. al* [56] et Hsieh *et. al* [46]. Dans ce chapitre, nous proposons une approche différente pour le problème basique de l'alignement de deux images, qui combine la détection de contours et la mise en correspondance automatique de landmarks en 2D. Notre méthode consiste à extraire tout d'abord les “contours significatifs” de deux images devant être recalées, puis à déterminer des ensembles de landmarks qui sont localisés le long de ces contours. Ensuite, nous déterminons une transformation qui met en correspondance les landmarks qui se trouvent sur les “contours communs” des deux images. Cette transformation est finalement appliquée à l'ensemble des pixels de l'une des images pour les synchroniser. Il s'agit donc d'une extension de la méthode décrite dans les chapitres précédents, où les contours des images jouent le rôle des extrema ou des singularités dans un signal 1D.

Avant de présenter cette méthode, il est nécessaire de se demander comment décrire un contour dans une image ? On pourrait définir les contours comme les points où l'intensité d'une image varie brutalement. Ce type de points peut être détecté par l'algorithme de Canny [16] qui calcule le module du gradient d'une image f :

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2} \right).$$

Le gradient d'une image ∇f est parallèle à la direction de changement maximum d'intensité dans la surface définie par f . Un point $y \in \mathbb{R}^2$ appartient donc à un

contour si ∇f est localement maximum au point $x = y$, pour $x = y + \lambda \nabla f$ et $|\lambda|$ suffisamment petit. Ceci signifie que la dérivée partielle de f au point y possède un maximum local dans la direction du gradient, et y correspond alors à un point d'inflexion de f .

Toutefois une telle définition des contours n'est pas satisfaisante. En effet, les textures d'une image sont souvent des zones où l'intensité varie brutalement bien que l'on considère généralement qu'elles ne contiennent pas de contours. Lorsque l'on regarde un ensemble de textures, on peut décider d'inclure dans la définition des contours celui de chaque région texturée, ou bien considérer uniquement la forme globale de l'ensemble des textures comme un contour qui contient une seule zone texturée. Par exemple, dans l'image de Lenna (voir Figure 5.1), doit-on prendre en compte le contour de chaque plume ou bien uniquement le contour de l'ensemble des plumes ? Ainsi, il est difficile de définir une méthode générale de détection de contours qui donnera des résultats satisfaisants dans beaucoup de contextes différents. La distinction entre des contours globaux, qui caractérisent par exemple la forme générale de l'image, et ceux plus locaux qui peuvent être dus à des textures dépend en particulier de l'échelle d'analyse. Les techniques de décomposition multi-échelles d'une image sont donc bien adaptées pour déterminer l'importance de ses contours.

En particulier, la transformée en ondelettes continue, introduite dans le chapitre 3, peut être étendue au cas bidimensionnel. Soit $\theta(u)$ un noyau convolution avec $u = (u_1, u_2) \in \mathbb{R}^2$ (θ peut par exemple être un noyau Gaussien 2D) et définissons les deux ondelettes suivantes, qui sont les dérivées partielles de θ :

$$\psi^1 = -\frac{\partial \theta}{\partial u_1} \text{ et } \psi^2 = -\frac{\partial \theta}{\partial u_2}.$$

Pour $k = 1, 2$ et $s > 0$, posons :

$$\psi_s^k(u_1, u_2) = \frac{1}{s} \psi^k\left(\frac{u_1}{s}, \frac{u_2}{s}\right).$$

Comme décrit dans Mallat ([63], Chapitre 6), la transformée en ondelettes continue d'un signal $f \in L^2(\mathbb{R}^2)$ au point $x = (x_1, x_2)$, dans les deux directions indexées par $k = 1, 2$ est alors définie par :

$$W_s^k(f)(x) = \langle f(\cdot), \psi_s^k(x - \cdot) \rangle,$$

où $\langle \cdot, \cdot \rangle$ est le produit scalaire usuel sur $L^2(\mathbb{R}^2)$. Soit $\theta_s(x) = \frac{1}{s} \theta\left(\frac{x}{s}\right)$ et $\bar{\theta}_s(x) = \theta_s(-x)$. Les composants de la transformée en ondelettes 2D sont alors proportionnels aux coordonnées du gradient de f lissée par $\bar{\theta}_s$:

$$\begin{pmatrix} W_s^1(f) \\ W_s^2(f) \end{pmatrix} = s \begin{pmatrix} \frac{\partial}{\partial x_1}(f \star \bar{\theta}_s)(x) \\ \frac{\partial}{\partial x_2}(f \star \bar{\theta}_s)(x) \end{pmatrix} = s \nabla(f \star \bar{\theta}_s)(x)$$

Le module du gradient de l'image lissée est alors proportionnel au module de la transformée en ondelette de f à l'échelle s et au point x :

$$M_s(f)(x) = \sqrt{|W_s^1(f)(x)|^2 + |W_s^2(f)(x)|^2}.$$

Soit $A_s(f)(x)$ l'angle du vecteur de la transformée en ondelette dans le plan (x_1, x_2) ,

$$\begin{aligned} A_s(f)(x) &= \alpha(x) \text{ si } W_s^1(f)(x) \geq 0, \\ &= \pi - \alpha(x) \text{ si } W_s^1(f)(x) < 0, \end{aligned}$$

avec

$$\alpha(x) = \tan^{-1} \left(\frac{W_s^2(f)(x)}{W_s^1(f)(x)} \right).$$

Le vecteur unitaire $\eta_s(x) = (\cos A_s(f)(x), \sin A_s(f)(x))$ est colinéaire à $\nabla(f \star \bar{\theta}_s)(x)$. Un point d'un contour à l'échelle s est donc un point x_0 tel que $M_s(f)(x_0)$ est localement maximum à $x = x_0$ quand $x = x_0 + \lambda \eta_s(x)$ pour $|\lambda|$ suffisamment petit. Ces points sont également appelés maxima d'ondelettes et l'image lissée $f \star \bar{\theta}_s$ a donc un point d'inflexion au niveau des maxima d'ondelettes. La détection des maxima d'une transformée en ondelettes 2D est donc équivalente à un détecteur multi-échelles de Canny.

Les coefficients $M_s(f)(x)$ de grandes amplitudes et les maxima d'ondelettes se concentrent autour des contours de l'image quand le niveau de résolution s tend vers zéro (voir Mallat [63]). La propagation des grands coefficients d'une transformée en ondelettes 2D permet donc une caractérisation multi-échelles des coins d'une image. Dans la Figure 5.1, nous avons représenté le module de la transformée continue en ondelettes de l'image de Lenna à différentes échelles dyadiques (ces figures correspondent aux exemples donnés dans Mallat [63], Chapitre 6, et ont été obtenues à partir du logiciel libre *LastWave* qui peut être téléchargé à l'URL : <http://www.cmap.polytechnique.fr/users/www.bacry>). On peut constater que les coefficients $M_s(f)(x)$ de grandes amplitudes (en noir dans les figures) correspondent aux contours de Lenna.

Comme pour le cas 1D, les landmarks d'une image peuvent donc être localisés en suivant la propagation des maxima locaux de sa transformée en ondelettes à différentes échelles. Toutefois, de nombreux coefficients d'ondelettes sont nécessaires pour représenter les contours d'une image (voir Figure 5.1), ce qui peut conduire à un nombre de landmarks beaucoup trop important pour être utilisé dans un problème de recalage. Ainsi, nous n'avons pas cherché à étendre au cas d'images, les résultats obtenus dans le chapitre 3 pour la détection de singularité d'un signal 1D.

Afin de mieux décrire les contours d'un signal 2D, nous avons été amené à utiliser de nouvelles représentations multi-échelles spécialement adaptées à la géométrie des images. Dans ces nouvelles représentations, les contours d'une image jouent le rôle

des points de singularités d'un signal 1D. En particulier, les wedgelets et les platelets, introduites respectivement par Donoho [29] et Willet et Nowak [94], conduisent à des approximations constantes ou linéaires par morceaux d'une image à différentes échelles et niveaux d'orientation. Ces nouveaux outils permettent en particulier une approximation des contours d'une image par des polygones à différents niveaux de résolution. Afin de détecter automatiquement les contours d'une image, nous avons adapté les Platelets au cas d'images contaminées par un bruit Gaussien. Nous montrons à l'aide de simulations que ces nouveaux outils sont bien adaptés à la détection de contours pour des images régulières par morceaux qui est le modèle *sous-jacent* de ces techniques. Dans le cas d'images réelles, la détection automatique de contours est plus délicate. Toutefois, nous montrons comment combiner l'approximation d'une image à partir de Platelets/Wedgelets, avec l'algorithme de mise en correspondance automatique de landmarks en 2D développé par Chui et Rangarajan [22]. L'intérêt de cette nouvelle méthode de mise en correspondance des "contours communs" de deux images est illustré par des exemples réels.

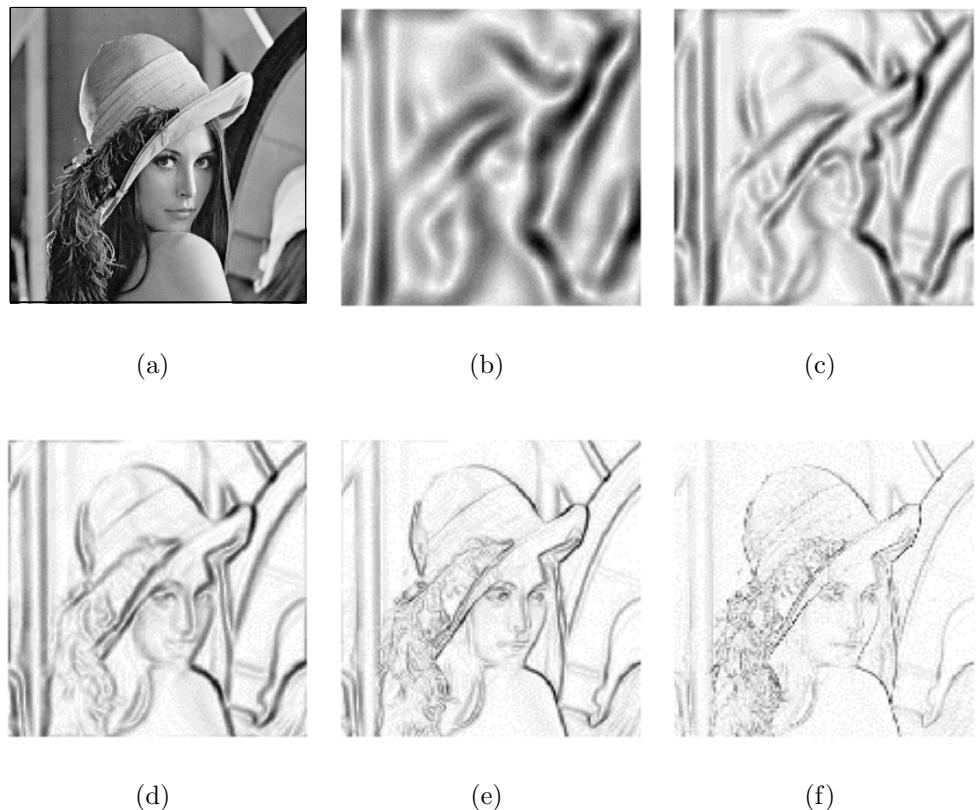


FIG. 5.1: Détection multi-échelles des contours de Lenna : (a) Image de Lenna. Module de la transformée en ondelette de Lenna $M_s(f)(x)$ pour (b) $s = 2^{-3}$ (c) $s = 2^{-4}$ (d) $s = 2^{-5}$ (e) $s = 2^{-6}$ (f) $s = 2^{-7}$. Le noir correspond aux coefficients de grandes amplitudes.

5.1 De nouvelles représentations adaptées aux contours

Les méthodes multi-échelles et en particulier les ondelettes sont des outils bien connus qui ont été largement utilisés, avec succès, en traitement du signal. Ces techniques ont montré leur efficacité pour deux classes de problèmes très importantes en traitement de l'image :

- **Compression** : le but est de réduire le nombre de bits nécessaire pour coder une image en trouvant une représentation creuse de celle-ci, qui est le meilleur compromis entre la taille mémoire de la nouvelle représentation (i.e. le nombre de bits) et la qualité visuelle de l'image reconstruite à partir de cette approximation.
- **Estimation** : lorsqu'une image inconnue est observée avec du bruit, il s'agit de trouver un moyen pour restaurer l'image originale.

Le but de l'analyse harmonique (Computational Harmonic Analysis, CHA), qui comprend les ondelettes et les paquets de cosine (voir e.g. Mallat [63]) et d'autres représentations multi-échelles que nous décrirons brièvement, est de trouver une “représentation optimale” d'un objet et d'obtenir un “algorithme rapide” pour construire cette représentation. L'approche générale en CHA est de décomposer un signal (1D ou 2D) en une collection d'objets organisés de façon hiérarchique depuis des échelles grossières vers des échelles plus fines. Ces représentations permettent une approximation d'un signal à différentes échelles (fréquences) et positions (temps). Les objets aux échelles grossières représentent les structures globales du signal original, alors que les objets aux fines échelles correspondent à des détails plus fins et plus locaux. Le but des méthodes de CHA est alors de sélectionner les objets les plus significatifs dans cette représentation multi-échelles, c'est à dire ceux qui résument le mieux l'information présente dans le signal original, en fonction de l'application désirée (e.g. compression ou estimation).

Les représentations les plus efficaces d'une image sont actuellement obtenues avec des bases d'ondelettes orthonormées. La plupart des algorithmes récents de codage d'images (e.g. le nouveau standard JPEG-2000) sont basés sur des ondelettes pour transformer et compresser les images. Comme nous l'avons précisé au début de ce chapitre, les contours sont les caractéristiques dominantes dans les images. Ils représentent l'information la plus significative pour notre perception visuelle d'une image, et sont donc d'une importance fondamentale pour la qualité de l'estimation et de la compression. Cependant, les ondelettes donnent généralement de mauvais résultats au voisinage des contours d'une image, car de nombreux coefficients sont nécessaires pour représenter un contour, et les algorithmes de codages ne prennent généralement pas en compte la dépendance des coefficients d'ondelettes le long des contours (voir Candes and Donoho [15]). Puisque les ondelettes sont bien adaptées pour caractériser les singularités d'un signal 1D, de nouvelles représentations temps-échelles ont été récemment développées, dans lesquelles les contours d'une image

jouent le rôle des singularités dans un signal 1D :

- **Wedgelets** : cette technique a été développée par Donoho [29]. La décomposition multi-échelles en wedgelets divise une image en blocks dyadiques à différentes échelles. Sur chaque block, l'image est projetée sur des wedgelets qui sont des fonctions *constantes par morceaux* avec des discontinuités linéaires. Les wedgelets sont donc des approximations constantes d'une image à différentes échelles et orientations. Une approximation par wedgelets réalise le meilleur compromis entre l'erreur de projection d'une image sur des régions constantes et la parcimonie de la représentation (i.e. le nombre de wedgelets et de blocks). Une présentation détaillée de cette approche est donnée dans la section suivante.
- **Beamlets** : cette analyse est basée sur un dictionnaire de segments linéaires (appelés beamlets) organisés de façon dyadique à différentes échelles, positions et orientations, et a été proposé par Donoho et Huo [33]. La transformée en beamlets est le calcul des intégrales d'une image le long de ces segments linéaires pour toutes les beamlets du dictionnaire. Arrangé sous forme pyramidale et de façon hiérarchique, l'analyse par beamlets permet d'obtenir un outil performant pour explorer de manière multi-résolution des données contenant des filaments ou des courbes, et les simulations effectuées par Donoho et Huo [33] montrent que cette approche est extrêmement robuste en présence de bruit.
- **Platelets** : introduites par Willet et Nowak [94] pour débruiter des images médicales (obtenues par photon limited), les platelets généralisent les wedgelets. Comme les wedgelets, l'analyse en platelets est basée sur une partition dyadique récursive d'une image en blocks dyadiques, mais produit des approximations *linéaires par morceaux* d'une image. Les platelets sont donc bien adaptées pour des images qui sont composées de régions linéaires séparées par des contours réguliers, alors que les wedgelets sont mieux adaptées pour des images constituées de régions constantes. Les approximations par platelets et wedgelets sont discutées sur de nombreux exemples dans les sections suivantes.

D'autres représentations multi-échelles, que nous ne discuterons pas ici, ont également été récemment proposées pour incorporer un traitement géométrique spécifique des contours :

- **Ridgelets et Curvelets** : ces nouvelles transformées multi-échelles ont été introduites par Candes et Donoho [14], [15]. Ces nouveaux outils donnent des bases orthonormées dont les éléments sont indexés par des paramètres d'échelle, de position et de direction, et conduisent à des représentations des images qui sont meilleures que les ondelettes d'un point de vue théorique. Ces transformations sont basées sur un principe de représentation *anisotropique* qui est assez différent de la représentation isotropique des ondelettes. Cependant, l'implémentation de ces méthodes est assez difficile et leurs avantages sur des images naturelles restent à confirmer.

- **Bandelets** : introduites par Le Pennec et Mallat [54], ces techniques utilisent des ondelettes foveal (voir Chang *et. al.* [18]) pour construire des familles orthonormales pour suivre la géométrie des contours réguliers d'une image. Cette approche donne de meilleurs résultats que les algorithmes à base d'ondelettes pour l'estimation et la compression d'images réelles.
- **ENO-EA (edge-adapted)** : Cohen et Matei [25], [65] ont proposé une approche qui n'utilise pas le produit tensoriel, et qui est adaptée à la géométrie des contours en utilisant les techniques de multirésolution discrètes introduites par Harten [45]. Leur représentation est bien adaptée pour des images qui sont constituées de régions polynomiales par morceaux séparées par des contours réguliers. Dans ce cas, leur approche permet une amélioration significative par rapport aux méthodes d'ondelettes pour la reconstruction des images le long des contours.

Rappelons que le but de notre étude est de construire une déformation qui peut aligner deux objets. Notre idée principale est d'extraire tout d'abord les contours de deux images devant être alignées, et d'utiliser ensuite l'algorithme de Chui et Rangarajan [22] pour trouver automatiquement une transformation qui met ces contours en correspondance. Les wedgelets et les platelets représentent les contours d'une image comme des courbes polygonales, dont les sommets peuvent être alors utilisés comme des landmarks représentant les caractéristiques importantes d'une image. Avant d'illustrer cette approche sur quelques exemples réels et simulés, nous décrivons plus précisément les techniques développées dans [29] et [94].

5.1.1 Détection de contours avec des wedgelets

Considérons une image dyadique qui est un tableau de nombres $n \times n$ avec $n = 2^J$ pour un entier $J \geq 0$. Nous avons choisi de considérer un pixel comme un carré de taille $1/n$ -par- $1/n$ placé sur une grille du carré $[0, 1] \times [0, 1]$ et I comme une fonction du carré unité $[0, 1] \times [0, 1]$ qui est constante par morceaux sur chaque pixel. Nous introduisons maintenant quelques notations et terminologies en suivant le développement proposé dans [29] :

- Un carré dyadique $S(k_1, k_2, j)$ est la collection de points $[k_1/2^j, (k_1 + 1)/2^j] \times [k_2/2^j, (k_2 + 1)/2^j]$ où $0 \leq k_1, k_2 < 2^j$ pour un entier $j \geq 0$. Pour simplifier les notations, nous écrirons parfois seulement S pour désigner un carré dyadique à une échelle donnée.
- Une division par wedgelets est définie par une ligne reliant deux points des côtés d'un carré dyadique. Les points sont choisis à partir d'un ensemble fini de sommets régulièrement espacés d'une distance δ les uns des autres. Dans [29], δ est égal à 2^{-J-K} pour un $K \geq 0$. Comme δ est dyadique, les sommets sont donc positionnés aux coins des pixels (dans tous nos exemples, nous avons pris $\delta = 1/n$) et il y a exactement $M_j = 4 \times 2^K 2^{j-J}$ sommets construits de cette façon sur un carré dyadique S à une échelle donnée j . En partant du coin supérieur droit d'un carré dyadique, les sommets sont marqués dans le

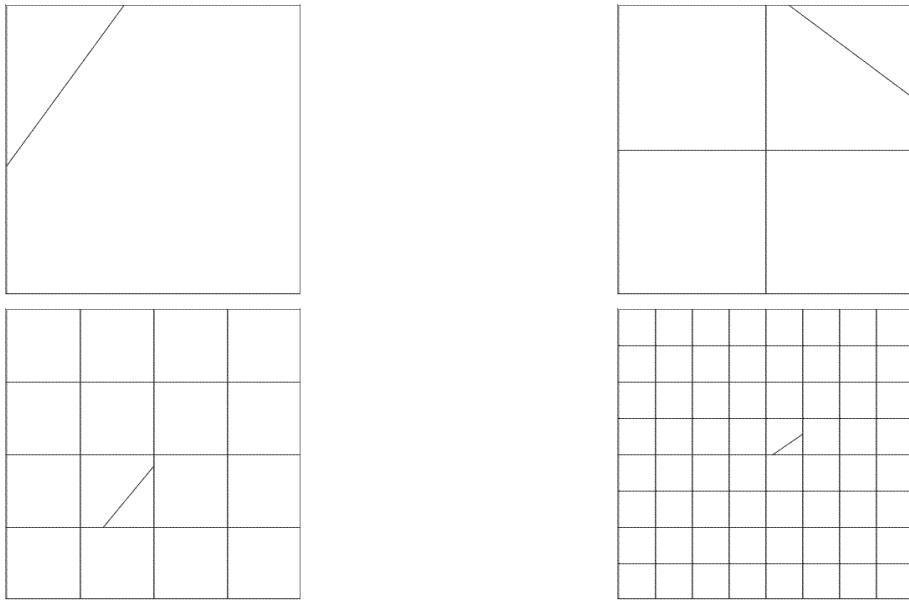


FIG. 5.2: Des wedgelets à différentes échelles.

sens des aiguilles d'une montre, et nous notons par $V(S) = \{v_{i,S}; 1 \leq i \leq M_j\}$ la collection de sommets ainsi obtenue pour un carré dyadique S à l'échelle j . Notons que si $j = J$ et $\delta = 1/n$, les sommets sont simplement les coins du carré. L'ensemble des edgelets $E(S)$ sur un carré dyadique S est la collection de toutes les lignes reliant deux sommets dans $V(S)$, i.e. $E(S) = \{\overline{v_{i,S}v_{i',S}}; 1 \leq i, i' \leq M_j\}$. Un edgelet est dit non-dégénéré s'il ne connecte pas deux sommets appartenant au même côté. L'ensemble des wedgelets sur un carré dyadique S est défini comme $W(S) = \{\mathbb{1}_S\} \cup \{w; w \in E(S) \text{ non-dégénéré}\}$ où $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A dans $[0, 1]^2$. Une wedgelet $w \in W(S), w \neq \mathbb{1}_S$, divise le carré dyadique S en deux régions appelées wedges que nous noterons par S_{wA} et S_{wB} . Pour un n donné et pour $\delta = 1/n$, $W(n)$ désigne la collection de toutes les wedgelets appartenant à un carré dyadique $S(k_1, k_2, j)$ pour $0 \leq k_1, k_2 < 2^j$ et $0 \leq j \leq J$. Un exemple de wedgelets à différentes échelles est donné dans la Figure 5.2.

- Soit $\tilde{Ave}\{I|A\}$ la valeur moyenne de l'image I sur un ensemble A dans $[0, 1]^2$. La décomposition en wedgelets d'une image est le vecteur de toutes les projections :

$$\begin{aligned}\tilde{Ave}\{I|w\} &= Ave\{I|S\}\mathbb{1}_S, \text{ si } w = \mathbb{1}_S \\ \tilde{Ave}\{I|w\} &= Ave\{I|S_{wA}\}\mathbb{1}_{S_{wA}} + Ave\{I|S_{wB}\}\mathbb{1}_{S_{wB}}, \text{ si } w \neq \mathbb{1}_S\end{aligned}$$

pour tout $w \in W(S) = W(S(k_1, k_2, j))$, $0 \leq k_1, k_2 < 2^j$ et $0 \leq j \leq J$. Notons qu'à l'échelle J , $S(k_1, k_2, J)$ correspond à un pixel et que $\tilde{Ave}\{I|w\}$ est donc égale à la valeur de I sur le pixel indexé par (k_1, k_2) si $w = \mathbb{1}_{S(k_1, k_2, J)}$.

Nous décrivons ensuite quelques terminologies d'héritage pour les carrés dyadiques de $[0, 1]^2$ à des échelles voisines. Notons que chaque carré dyadique $S(k_1, k_2, j)$

(pour $0 \leq j \leq J - 1$) se divise en quatre carrés dyadiques plus petits de même taille. Ces quatre carrés sont appelés les “enfants” de $S(k_1, k_2, j)$ et seront notés $\{ch(S^i(k_1, k_2, j))\}_{i=1}^4$ alors que $S(k_1, k_2, j)$ est appelé leur “parent”. Nous écrirons parfois $S' \in ch(S)$ pour désigner l’un des enfants S' d’un carré dyadique S . L’opération qui réunit quatre enfants en un parent commun est appelée une “quad-fusion”, alors que l’opération de diviser un carré parent en ses quatre enfants est appelé une “quad-division”, ce qui conduit à la définition suivante :

Définition 5.1 *Une partition récursive dyadique (RDP) est toute partition de $[0, 1]^2$ qui peut être construite en appliquant les règles suivantes :*

- La partition triviale $P = \{[0, 1]^2\}$ est une RDP.
- Si $P = \{S_1, \dots, S_m\}$ (où $S_i, i = 1, \dots, m$ désignent des carrés dyadiques) est une RDP, alors la partition obtenue en appliquant une “quad-division” à l’un des carrés dans P est encore une RDP.

L’ensemble de tous les carrés dyadiques depuis les échelles grossières vers les échelles les plus fines $P^* = \{S(k_1, k_2, j), 0 \leq k_1, k_2 < 2^j; 0 \leq j \leq J\}$ est parfois appelé une RDP complète, et peut être associée à un graphe de type “quad-arbre”. Cette arbre peut être élagué de manière adaptative pour construire une RDP incomplète, dont les sommets terminaux sont des carrés dyadiques de tailles variées, à différentes positions dans l’espace. Une partition récursive dyadique décorée par des wedgelets (WD-RDP) est une RDP dans laquelle chaque sommet terminal S d’une RDP a été décoré par exactement une wedgelet dans $W(S)$. Les régions terminales d’une WD-RDP sont donc soit un carré dyadique, soit deux wedges. Si P est une RDP, nous utiliserons la notation $S \in P$ pour désigner un carré dyadique S qui correspond à l’un des sommets terminaux de P .

Notons par $I(S)$ la restriction de l’image I à un carré dyadique S , et définissons pour chaque wedgelets $w \in W(S)$ la somme quadratique des résidus pénalisée (PRSS) :

$$\begin{aligned} PRSS(S, w) &= \|I(S) - \tilde{Ave}\{I|w\}\|^2 + \lambda^2, \text{ si } w = \mathbb{1}_S \\ PRSS(S, w) &= \|I(S) - \tilde{Ave}\{I|w\}\|^2 + 2\lambda^2, \text{ si } w \neq \mathbb{1}_S, \end{aligned}$$

où λ un paramètre de pénalisation. Le terme de pénalisation reflète la complexité du modèle : seulement un coefficient est requis pour l’approximation constante de I sur un carré dyadique, alors que deux coefficients sont nécessaires pour une division par une wedgelet. Ainsi, pour chaque carré dyadique S , définissons :

$$PRSS'(S) = \min_{w \in W(S)} (PRSS(S, w)). \quad (5.5.1)$$

Pour une WD-RDP P , définissons la complexité de la somme quadratique des résidus pénalisée :

$$J(P) = \sum_{S \in P} PRSS'(S).$$

Soit $WD - RDP([0, 1]^2)$ la collection de toutes les WD-RDP de $[0, 1]^2$. Une approximation par wedgelets d'une image est alors obtenue en trouvant la WD-RDP optimale \hat{P} :

$$\hat{P} = \operatorname{argmin}_{P \in WD - RDP([0, 1]^2)} J(P).$$

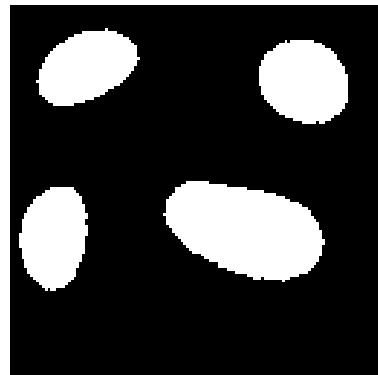
Dans [29], il est montré que ce problème d'optimisation peut être résolu en utilisant l'algorithme d'élagage d'arbre suivant :

- Construire l'arbre-quad complet P^* .
- **Initialisation** : Marquer chaque carré dyadique S de $[0, 1]^2$ par $w_S = \operatorname{argmin}_{w \in W(S)} (PRSS(S, w))$ et poser $j = J$.
- **Loop** : Pour chaque carré dyadique S à l'échelle j :
 - **Héritage** : Si $j < J$ poser $d_S = \sum_{S' \in ch(S)} PRSS'(S')$ ou poser $d_S = \infty$ si $j = J$ (étant donné qu'il n'y a pas d'enfants).
 - **Minimisation** : Calculer $c_S = \min(d_S, PRSS'(S))$. Si le minimum est donné par $PRSS'(S)$ marquer le sommet S : **Sommet terminal décoré par w_S** sinon marquer le sommet S : **Interior**.
 - **Elagage** : Si S est marqué **Terminal**, supprimer de l'arbre-quad courant les quatre enfants de S et les sous-arbres dont les racines se trouvent éventuellement au niveau des enfants.
- Poser $j = j - 1$. Si $j \geq 0$ aller à **Loop**.

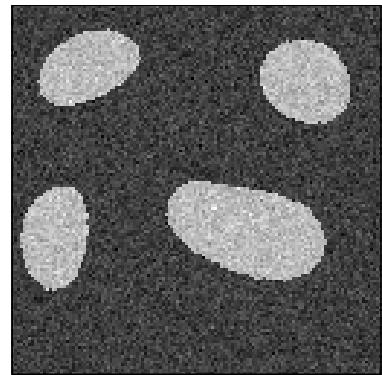
Notons que si le terme de régularisation n'est pas ajouté (i.e. $\lambda = 0$), la WD-RDP optimale est la RDP complète P^* . Cet algorithme peut être utilisé sur des données non-bruitées pour une forme de compression (voir [33]). Supposons que l'image I est contaminée par un bruit Gaussien de variance σ^2 et que nous observons un tableau n -par- n de données bruitées Y :

$$Y(k_1, k_2) = I(k_1, k_2) + \sigma \epsilon_{k_1, k_2}, 1 \leq k_1, k_2 \leq n,$$

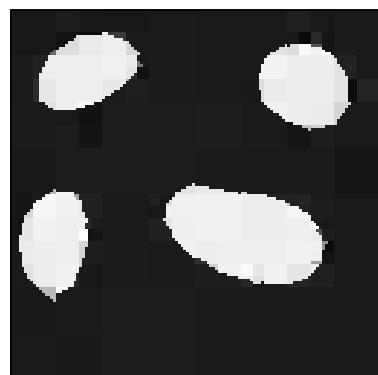
où (k_1, k_2) désigne la position d'un pixel sur la grille unité de $[0, 1]$ et les ϵ_{k_1, k_2} sont des variables de loi normale i.i.d. de moyenne nulle et de variance 1. Pour obtenir une estimation par wedgelets de l'image I , le paramètre de régularisation λ dans (5.5.1) doit être calibré en fonction du bruit. Dans [33], il est expliqué qu'un choix recommandé pour λ est $2\sigma\sqrt{\log(n)}$ (ceci est basé sur des arguments théoriques développés dans [29]). Un exemple numérique d'approximation d'une image bruitée constituée de régions constantes par morceaux est donné dans la Figure 5.3. Pour représenter une WD-RDP, nous avons seulement tracé les sommets terminaux qui forment une partition de $[0, 1]$. La qualité de l'estimation est très bonne et produit des résultats très satisfaisants au niveau des frontières. Notons que les décorations dans la WD-RDP sont placées le long des contours de l'image. Cependant, les frontières données par l'approximation par wedgelets sont des fragments linéaires déconnectés, car aucune contrainte directionnelle ou de continuité n'est imposée. Quelques possibilités pour incorporer des contraintes géométriques dans l'estimation par wedgelets sont discutées dans la section 5.3.



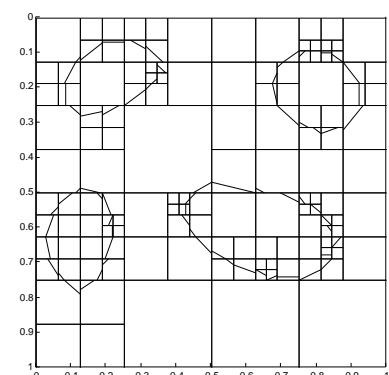
(a)



(b)



(c)



(d)

FIG. 5.3: Approximation par wedgelets d'une image bruitée : (a) Image originale (b) Image plus Bruit Gaussien (c) Approximation par wedgelets (d) WD-RDP associée à l'approximation par wedgelets

5.1.2 Détection de contours avec des platelets

Les platelets généralisent les wedgelets. Elles ont été introduites par Willet et Nowak [94] pour l'estimation d'images médicales photon-limited contaminées par un bruit Poissonien. Dans la décomposition en platelets, les approximations constantes par morceaux de la décomposition en wedgelets sont remplacées par un modèle linéaire par morceaux. Afin d'être plus précis, la décomposition en platelets d'une image n -par- n $I(x, y)$, est le vecteur des projections sur tous les carrés dyadiques $S = S(k_1, k_2, j)$, $0 \leq k_1, k_2 < 2^j$ et $0 \leq j \leq J$:

$$\begin{aligned}\tilde{Lin}\{I|w\} &= (D_S x + E_S y + F_S) \mathbb{1}_S, \text{ si } w = \mathbb{1}_S \\ \tilde{Lin}\{I|w\} &= (D_{S_{wA}} x + E_{S_{wA}} y + F_{S_{wA}}) \mathbb{1}_{S_{wA}} + \\ &\quad (D_{S_{wB}} x + E_{S_{wB}} y + F_{S_{wB}}) \mathbb{1}_{S_{wB}}, \text{ si } w \neq \mathbb{1}_S\end{aligned}$$

pour tout $w \in W(S)$, où les coefficients des modèles linéaires sont choisis de façon à minimiser l'erreur quadratique de l'approximation linéaire de l'image restreinte au carré S ou à une wedge. Les platelets sont donc bien adaptées pour approximer des images constituées de régions lisses séparées par des contours réguliers, et dans ce cas l'erreur d'approximation au m -ème terme est meilleure que celle obtenue avec des bases d'ondelettes, de Fourier ou des wedgelets (voir [94]). L'estimation par platelets est également basée sur un schéma de partition dyadique récursive et sur des estimateurs de maximum de vraisemblance pénalisée. Dans [94], la pénalisation reflète la complexité croissante des modèles linéaires par morceaux par rapport aux approximations constantes : trois coefficients sont nécessaires pour l'approximation par platelets sur un carré dyadique et six coefficients sont nécessaire pour une division par wedgelets (à comparer avec un coefficient pour une approximation constante sur un carré dyadique et deux coefficients pour une division par wedgelets). Dans [94], l'analyse d'images par platelets est proposée pour des données distribuées selon une loi de Poisson, mais nous pouvons l'adapter à des images contaminées par un bruit Gaussien. Dans le cas Gaussien, le critère de maximum de vraisemblance utilisé dans [94] est alors équivalent à un critère des moindres carrés. Rappelons que $I(S)$ désigne la restriction de l'image I au carré dyadique S , et définissons pour chaque wedgelets $w \in W(S)$ la somme quadratique des résidus pénalisée (PRSS) suivante :

$$\begin{aligned}PRSS_0(S, w) &= \|I(S) - \tilde{Ave}\{I|w\}\|^2 + \beta_0 \lambda^2, \text{ si } w = \mathbb{1}_S \\ PRSS_1(S, w) &= \|I(S) - \tilde{Ave}\{I|w\}\|^2 + \beta_1 \lambda^2, \text{ si } w \neq \mathbb{1}_S, \\ PRSS_2(S, w) &= \|I(S) - \tilde{Lin}\{I|w\}\|^2 + \beta_2 \lambda^2, \text{ si } w = \mathbb{1}_S \\ PRSS_3(S, w) &= \|I(S) - \tilde{Lin}\{I|w\}\|^2 + \beta_3 \lambda^2, \text{ si } w \neq \mathbb{1}_S,\end{aligned}$$

où λ est un paramètre de pénalisation et $\beta_i, i = 0, \dots, 3$ sont des constantes qui reflètent la complexité du modèle. En suivant les considérations dans [94], nous avons pris $\beta_0 = 1, \beta_1 = 2, \beta_2 = 3$ et $\beta_3 = 6$. Ensuite, pour chaque carré dyadique S ,

définissons :

$$\begin{aligned}
PRSS_1^*(S) &= \min_{w \in W(S)} (PRSS_1(S, w)), \\
w_S &= \operatorname{argmin}_{w \in W(S)} (PRSS_1(S, w)), \\
PRSS_3^*(S) &= \min_{w \in W(S)} (PRSS_3(S, w)), \\
w_{PS} &= \operatorname{argmin}_{w \in W(S)} (PRSS_3(S, w)), \\
PRSS_{min}(S) &= \min [PRSS_0(S, \mathbf{1}_S), PRSS_1^*(S), PRSS_2(S, \mathbf{1}_S), PRSS_3^*(S)], \\
\theta_{min}(S) &= \operatorname{argmin} [PRSS_0(S, \mathbf{1}_S), PRSS_1^*(S), PRSS_2(S, \mathbf{1}_S), PRSS_3^*(S)],
\end{aligned}$$

où $\theta_{min}(S) = i$ si le minimum est atteint par $PRSS_i^*(S)$, $i = 0, \dots, 3$. Pour une WD-RDP P , définissons la complexité de la somme quadratique des résidus pénalisée :

$$J_2(P) = \sum_{S \in P} PRSS_{min}(S).$$

Soit $WD - RDP([0, 1]^2)$ la collection de toutes les WD-RDP de $[0, 1]^2$. Une approximation par platelets d'une image est alors obtenue en trouvant la WD-RDP optimale \hat{P} :

$$\hat{P} = \operatorname{argmin}_{P \in WD - RDP([0, 1]^2)} J_2(P).$$

Ce problème d'optimisation peut être résolu en utilisant l'algorithme d'élagage d'arbre suivant :

- Construire l'arbre-quad complet P^* .
- **Initialisation** : $j = J$.
- **Loop** : Pour chaque carré dyadique S à l'échelle j :
 - **Héritage** : Si $j < J$ poser $d_S = \sum_{S' \in ch(S)} PRSS_{min}(S')$ ou bien poser $d_S = \infty$ si $j = J$ (puisque il n'y a pas d'enfants).
 - **Minimisation** : Calculer $c_S = \min(d_S, PRSS_{min}(S))$. Si le minimum est atteint par $PRSS_{min}(S)$ marquer le sommet S comme suit :
 - si $\theta_{min}(S) = 0$, Sommet terminal décoré par $\mathbf{1}_S$: approximation constante sur le carré dyadique S .
 - si $\theta_{min}(S) = 1$, Sommet terminal décoré par w_S : approximation constante sur deux wedges.
 - si $\theta_{min}(S) = 2$, Sommet terminal décoré par $\mathbf{1}_S$: approximation linéaire sur le carré dyadique S .
 - si $\theta_{min}(S) = 3$, Sommet terminal décoré par w_{PS} : approximation linéaire sur deux wedges.
 - sinon marquer le sommet S : Intérieur.
- **Elagage** : Si S est marqué Terminal, supprimer de l'arbre-quad courant les quatre enfants de S et les sous-arbres dont les racines se trouvent éventuellement au niveau des enfants.

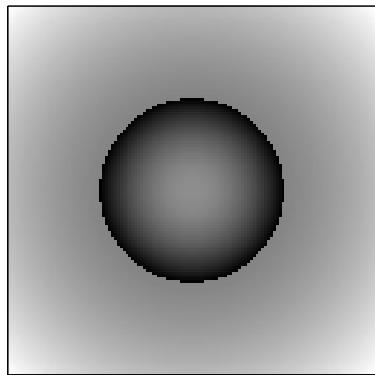
- Poser $j = j - 1$. Si $j \geq 0$ aller à Loop.

Si nous observons un tableau n -par- n Y qui consiste en une image I contaminée par un bruit Gaussien de moyenne nulle et de variance σ^2 , nous proposons de poser $\lambda = 2\sigma\sqrt{\log(n)}$ pour estimer I , comme dans le cas de l'approximation par wedgelets. Nous pensons qu'il est possible de quantifier les performances théoriques de l'estimation par platelets en suivant le développement proposé dans [29], mais ce problème est seulement considéré dans une discussion sur des perspectives de recherche à la fin de ce chapitre. L'exemple numérique donné dans la Figure 5.4 illustre la supériorité de l'estimation par platelets sur le débruitage par wedgelets pour des images constituées de régions lisses séparées par des frontières régulières. L'image de la Figure 5.4a est une “boule quadratique”. Elle est utilisée dans [94] pour montrer la supériorité des platelets sur les wedgelets. La qualité visuelle de l'estimation par platelets est bien meilleure, et les contours détectés par la WD-RDP associée à l'estimation par platelets correspondent aux frontières de la boule. La WD-RDP obtenue avec l'estimation par wedgelets est beaucoup plus bruitée : les décorations ne sont pas cohérentes globalement, et de nombreuses wedgelets ne correspondent pas à un contour de l'image.

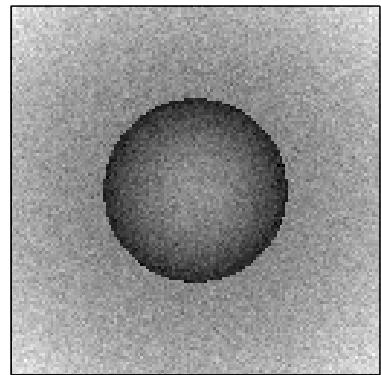
5.2 Combiner la détection de contours et l'alignement de landmarks

5.2.1 Alignement de wedgelets

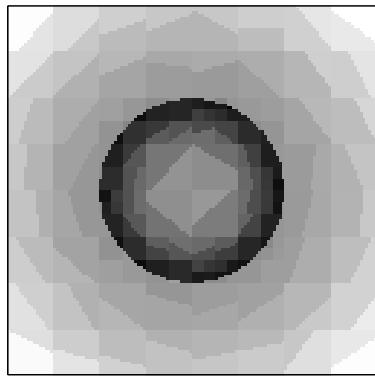
Dans le chapitre 2, nous avons présenté une méthode de mise en correspondance automatique de deux ensembles de landmarks en 2D, qui peuvent éventuellement contenir des landmarks erronés. Dans la section précédente, nous avons également décrit une approche multi-échelles pour approximer les contours d'une image observée avec un bruit Gaussien. Nous proposons de combiner ces deux techniques pour déterminer une transformation non-rigide entre deux images. Supposons que les contours de deux images I_1 et I_2 ont été estimés par une approximation par wedgelets ou platelets. Soit $W_1 = \{w_{1,1}, \dots, w_{1,N}\}$ et $W_2 = \{w_{2,1}, \dots, w_{2,K}\}$ les deux ensembles de wedgelets représentant les contours de I_1 et I_2 respectivement. Pour illustrer notre approche, nous considérons le problème de recalage “simple” de la Figure 5.5 où un losange (Figure 5.5a) a été déformé par une transformation non-rigide (Figure 5.5b). Les WD-RDP dans les Figures 5.5cd ont été obtenues par une approximation par wedgelets avec le même paramètre de régularisation, et l'on peut constater que les wedgelets correspondent effectivement aux frontières dans les deux images. Pour aligner les deux images I_1 et I_2 , nous proposons de calculer une déformation qui trouve le “meilleur alignement” entre les segments linéaires dans W_1 et W_2 . Evidemment, une telle approche nécessite de définir tout d'abord une “distance” pour mesurer la qualité de l'alignement de deux wedgelets. De plus, si l'on veut aligner les segments linéaires de la Figure 5.5c avec ceux de la Figure 5.5d, l'on doit être capable de “diviser une wedgelet” à une échelle grossière pour l'associer avec



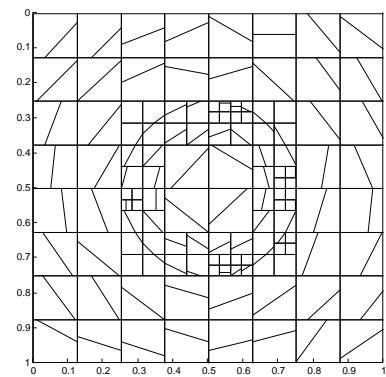
(a)



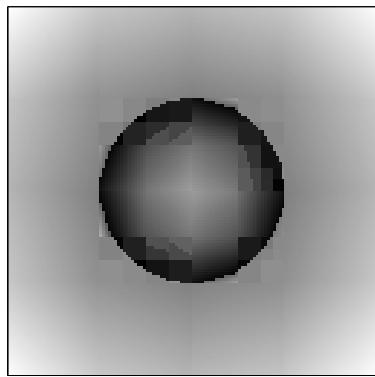
(b)



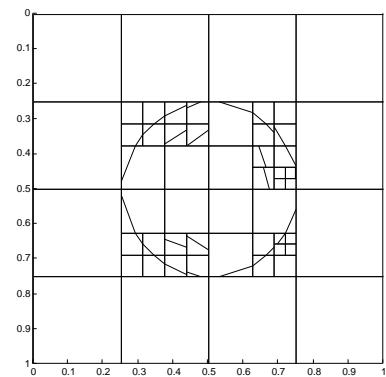
(c)



(d)



(e)



(f)

FIG. 5.4: Estimation par platelets contre estimation par wedgelets : (a) Image originale (b) Image plus bruit Gaussien (c) Estimation par wedgelets (d) WD-RDP optimale associée à l'estimation par wedgelets (e) Estimation par platelets (f) WD-RDP optimale associée à l'estimation par platelets.

différentes wedgelets à une échelle plus fine, ou bien de “réunir” plusieurs wedgelets à une échelle fine en une seule wedgelets à une échelle plus grossière. Cependant, jusqu'à présent, nous n'avons pas réussi à incorporer ces opérations dans un processus de recalage. Ainsi, plutôt que d'effectuer le recalage sur les wedgelets directement, nous proposons d'aligner les deux ensembles de **landmarks définis par les sommets des wedgelets** dans W_1 et W_2 en utilisant l'algorithme TPS-RPM de Chui and Rangajaran [22]. Evidement, cette approche peut fonctionner si les sommets des wedgelets forment des ensembles de points qui reflètent la forme des contours dans les deux images. Cependant, en regardant la Figure 5.5, il paraît clair que cette approche ne peut pas fonctionner, puisque nous avons obtenu seulement 8 sommets pour le losange qui ne forment pas une “description satisfaisante” de ses contours. Pour remédier à ce problème, nous avons choisi de considérer uniquement les wedgelets à une échelle donnée suffisamment fine, de sorte que les sommets de ces wedgelets décrivent de manière suffisamment précise la forme des contours. Il ne s'agit pas d'une solution idéale, mais elle est très pratique et donne de très bons résultats comme nous allons le montrer dans les exemples qui suivent.

Pour être plus précis, supposons que l'image I_2 doive être alignée avec I_1 . Nous proposons finalement l'algorithme suivant :

- Soit j_0 une échelle suffisamment fine.
- Calculer la décomposition en platelets de I_1 et I_2 pour chaque carré dyadique à l'échelle j_0 .
- Soit λ_1 et λ_2 deux paramètres de régularisation associés à I_1 et I_2 respectivement. Pour chaque image et pour chaque carré dyadique S à l'échelle j_0 , calculer $\theta_{min}(S) = \operatorname{argmin} [PRSS_0(S, \mathbf{1}_S), PRSS_1^*(S), PRSS_2(S, \mathbf{1}_S), PRSS_3^*(S)]$. Si $\theta_{min}(S) = 1$ alors S est décoré par w_S et si $\theta_{min}(S) = 3$ alors S est décoré par w_{PS} (pour l'image I_i , les sommes quadratiques des résidus pénalisées sont calculées avec $\lambda_i, i = 1, 2$).
- Soit $W_1^{j_0}$ et $W_2^{j_0}$ deux ensembles de wedgelets obtenus de cette façon pour I_1 et I_2 , et soit V_1 et V_2 les ensembles de landmarks qui correspondent aux sommets des wedgelets dans $W_1^{j_0}$ et $W_2^{j_0}$ respectivement.
- Calculer la transformation u qui aligne l'ensemble de points V_1 sur V_2 avec l'algorithme TPS-RPM.
- Une version déformée de I_2 est $I_2 \circ u$. Si (k_1, k_2) désigne la position d'un pixel sur une grille du carré unité $[0, 1]^2$, alors typiquement $u(k_1, k_2)$ ne correspond pas à un pixel de cette grille. Ainsi, comme expliqué dans le chapitre 2, $I_2(u(k_1, k_2))$ est obtenu par interpolation bilinéaire de I_2 . Si $u(k_1, k_2)$ est en dehors de $[0, 1]^2$, nous posons $I_2(u(k_1, k_2))$ égal à zéro (ce qui correspond à des pixels noirs dans les images).

Les images des Figure 5.6 sont de taille 128×128 pixels. Pour aligner l'image de la Figure 5.6b avec le losange de la Figure 5.6a, nous avons choisi $j_0 = 5$ et les wedgelets tracées dans la Figure 5.6cd ont été extraites avec le même paramètre de régularisation pour les deux images. Notons qu'à cette échelle, les sommets des

wedgelets donnent une bien meilleure approximation des contours des deux images par rapport à l'estimation obtenue dans la Figure 5.5cd. Le résultat de l'alignement des sommets des wedgelets de la Figure 5.6c avec ceux de la Figure 5.6d est donné dans la Figure 5.5f. Le résultat est très satisfaisant car la forme de l'image déformée est très proche de la forme d'un losange, sauf que les côtés du losange déformé dans la Figure 5.5e ne sont pas des lignes parfaitement droites. Un exemple plus sophistiqué de recalage est montré dans la Figure 5.7. Les images de la Figure 5.7cd sont de taille 128×128 pixels et sont observées avec un bruit Gaussien de moyenne nulle et de variance σ^2 (les images originales sont dessinées dans la Figure 5.7ab). Nous avons à nouveau choisi $j_0 = 5$, et les wedgelets tracées dans la Figure 5.7ef ont été estimées avec $\lambda = 2\sigma\sqrt{\log(n)}$. Notons que la forme de ces wedgelets correspond précisément à la forme des contours des images des Figures 5.7ab. L'alignement des sommets des wedgelets de la Figure 5.7e avec ceux de la Figure 5.7f est donné dans la Figure 5.7h. Pour mesurer la qualité de l'alignement, nous avons déformé l'image non-bruitée de la Figure 5.7b. L'image recalée est donnée dans la Figure 5.7g. Le résultat est excellent car nous obtenons bien "quatre cercles" qui sont presque exactement alignés avec les cercles originaux de la Figure 5.7a.

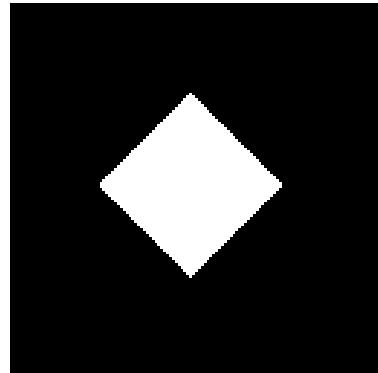
5.2.2 Application à des images réelles

Nous allons maintenant montrer comment notre approche peut s'appliquer à deux exemples réels de recalage d'images que nous avons présentés dans l'introduction et le chapitre 2 : alignement de visages et de poissons. Les visages des Figures 5.8-5.9 sont de taille 112×92 pixels, alors que les images de poissons des Figures 5.10 sont de taille 300×500 pixels. Ainsi, ces images ne sont pas dyadiques et l'algorithme de décomposition en platelets donné précédemment doit être légèrement modifié. Supposons que nous observons un tableau n_1 -par- n_2 qui représente une image I , et que nous voulons effectuer sa décomposition en platelets à une échelle donnée où la longueur d'un côté d'un block dyadique S est sh_0 (quand $n_1 = n_2 = 2^J$, alors $sh_0 = 2^{J-j_0}$ à l'échelle j_0). L'image I est alors "étendue" de la manière suivante :

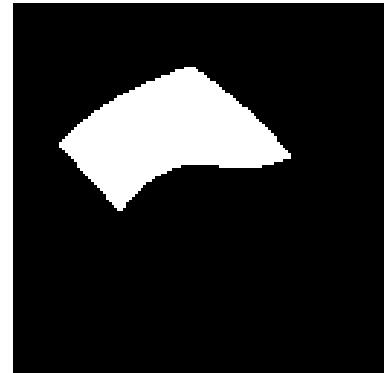
- soit $I_{ext} = I$, $l_2 = \lfloor n_2/sh_0 \rfloor sh_0 - n_2$ et $l_1 = \lfloor n_1/sh_0 \rfloor sh_0 - n_1$ ($\lfloor x \rfloor$ désigne la partie entière de x).
- si $l_2 > 0$, alors ajouter l_2 colonnes à I_{ext} qui sont égales à la n_2 -ème colonne de I .
- si $l_1 > 0$, alors ajouter l_1 lignes à I_{ext} qui sont égales à la n_1 -ème ligne de I_{ext} .

Après ces opérations, I_{ext} est alors une image qui contient $\frac{n_1+l_1}{sh_0} \times \frac{n_2+l_2}{sh_0}$ carrés dyadiques dont le côté est de longueur sh_0 . Alors, l'approximation par platelets de I est obtenue en calculant la PRSS minimale sur tous les carrés dyadiques S dont le côté est de longueur sh_0 , qui sont dans I_{ext} , et en décorant éventuellement ces carrés dyadiques avec des wedgelets.

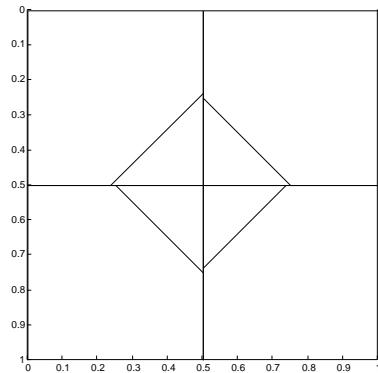
Après avoir choisi une longueur de côté appropriée sh_0 , le point délicat est le choix d'un paramètre de régularisation qui permette d'extraire les wedgelets "les



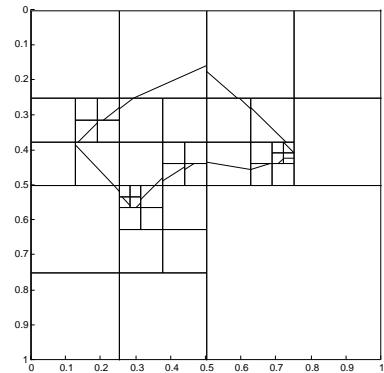
(a)



(b)

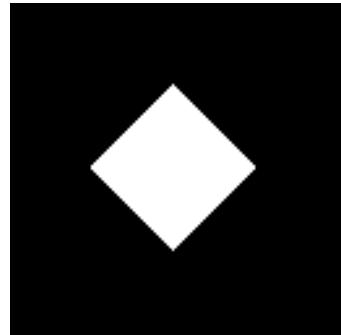


(c)



(d)

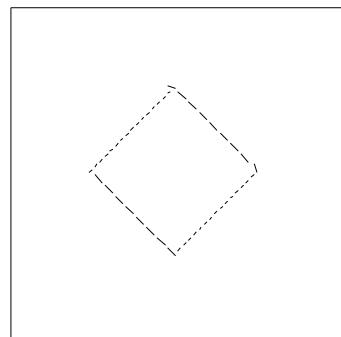
FIG. 5.5: Décomposition en wedgelets d'un losange et sa déformation par une transformation non-rigide : (a) Losange original (b) Transformation non-rigide du losange. WD-RDP optimale associée à l'approximation par wedgelets du losange (c) et du losange déformé (d) avec le même paramètre de régularisation.



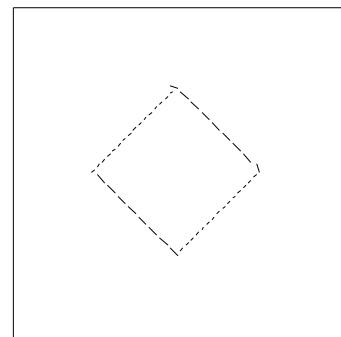
(a)



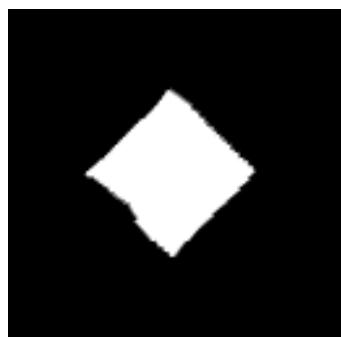
(b)



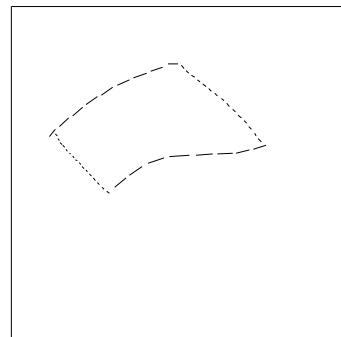
(c)



(d)



(e)



(f)

FIG. 5.6: Résultat du recalage du losange déformé : (a) Losange original (taille : 128×128). (b) Déformation non-rigide du losange (taille : 128×128). Wedgelets optimales associées à l'approximation par platelets du losange (c) et du losange déformé (d) à l'échelle $j_0 = 5$ avec le même paramètre de régularisation. (e) Version recalée du losange déformé. (f) Alignement des sommets des wedgelets du losange avec ceux du losange déformé.

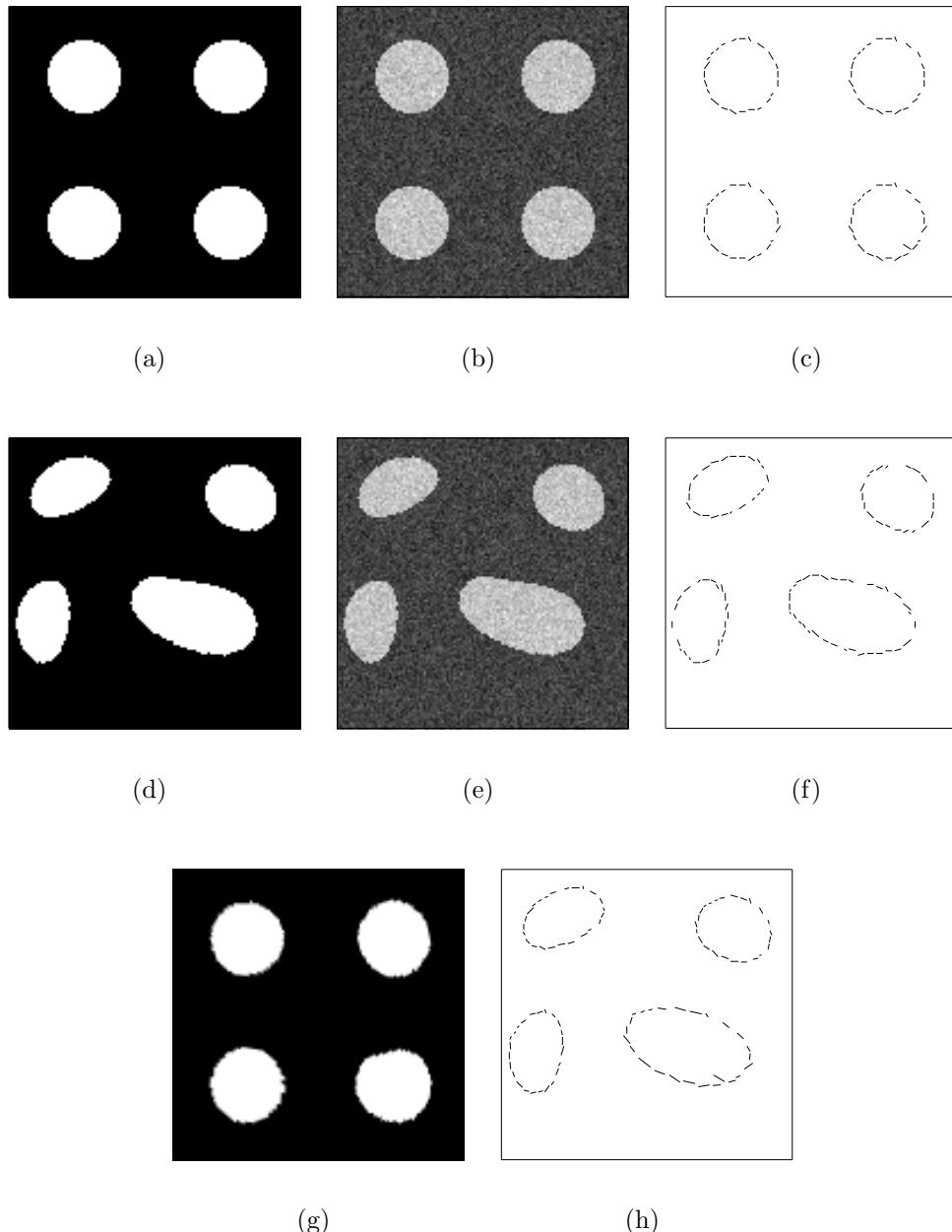


FIG. 5.7: Résultat du recalage de 4 cercles déformés : (a) Cercles originaux (taille : 128×128). (d) Déformation non-rigide des cercles (taille : 128×128). (b), (e) Images originales observées avec un bruit Gaussien. Wedgelets optimales associées à l'approximation par platelets des cercles (c) et des cercles déformés (f) observés avec du bruit à l'échelle $j_0 = 5$ avec $\lambda = 2\sigma\sqrt{(\log(n))}$. (g) Version recalée des cercles déformés. (h) Alignement des sommets des wedgelets des cercles avec ceux des cercles déformés.

plus significatives” à cette échelle. En suivant nos discussions précédentes, nous devrions prendre $\lambda = K\sigma\sqrt{\log(N)}$ où σ est le niveau de bruit estimé dans l’image I , K une constante appropriée et N est le nombre de pixels dans l’image I . Cependant, nous n’avons pas réussi à obtenir de résultats satisfaisants de cette façon car le nombre de wedgelets sélectionnées est beaucoup trop important pour pouvoir utiliser l’algorithme TPS-RPM. Il serait toutefois très intéressant de pouvoir déterminer automatiquement ce paramètre. Les exemples donnés dans ce chapitre sont basés sur des choix empiriques pour λ , mais illustrent l’efficacité de notre approche en pratique.

Pour aligner le visage de la Figure 5.8b avec celui de la Figure 5.8a, nous avons pris $sh_0 = 4$ et $\lambda = Var(I)/4$ pour extraire les wedgelets les plus significatives dans les carrés dyadiques de taille sh_0 , où $Var(I)$ dénote la variance des pixels de l’image I (il s’agit d’une mesure de l’“énergie” d’une image). Pour aligner le visage de la Figure 5.9b avec celui de la Figure 5.9a, nous avons choisi $sh_0 = 4$ et $\lambda = Var(I)/8$. Finalement, pour aligner le haddock avec le whiting (voir Figure 5.10) nous avons pris $sh_0 = 16$ et $\lambda = Var(I)/6$. Pour mesurer la qualité de l’alignement, nous avons également représenté la différence pixel-par-pixel (les valeurs nulles correspondent au niveau de gris moyen) entre les images avant et après recalage. Les résultats pour l’alignement des visages sont très satisfaisants étant donné que les principales caractéristiques d’un visage (e.g. yeux, bouche, oreilles) sont correctement alignés ce qui n’était pas le cas avant le recalage. Ceci peut être expliqué par le choix de $sh_0 = 4$ qui donne une description précise des contours des différents visages. Pour l’alignement des poissons, le résultat est encore très bon. Avec $sh_0 = 16$, nous avons obtenu une estimation globale de la forme de chaque poisson, mais certains détails plus fins n’ont pas été pris en compte (e.g. les yeux), et certaines parties des deux images ne sont pas correctement alignées après recalage. Ceci pourrait être amélioré en diminuant la longueur des carrés dyadiques sh_0 , mais ceci augmente très fortement le nombre de landmarks devant être alignés et l’algorithme TPS-RPM devient extrêmement lent. Cet exemple illustre le fait que considérer des wedgelets à des échelles différentes est d’une importance cruciale pour améliorer la qualité du recalage.

5.3 Discussions

Nous avons présenté une approche basée sur les contours pour le recalage d’images qui combine l’algorithme TPS-RPM développé dans [22] avec l’estimation des contours d’une image à partir de son approximation en platelets. Cette approche donne des résultats satisfaisants à la fois sur des exemples simulés et sur des images réelles cependant tout au long de ce chapitre, nous avons noté quelques problèmes qu’il est nécessaire de discuter :

- L’algorithme TPS-RPM ne garantie pas que la solution obtenue est bijective. Des plissements sont possibles et comme précisé dans [22], l’algorithme TPS-

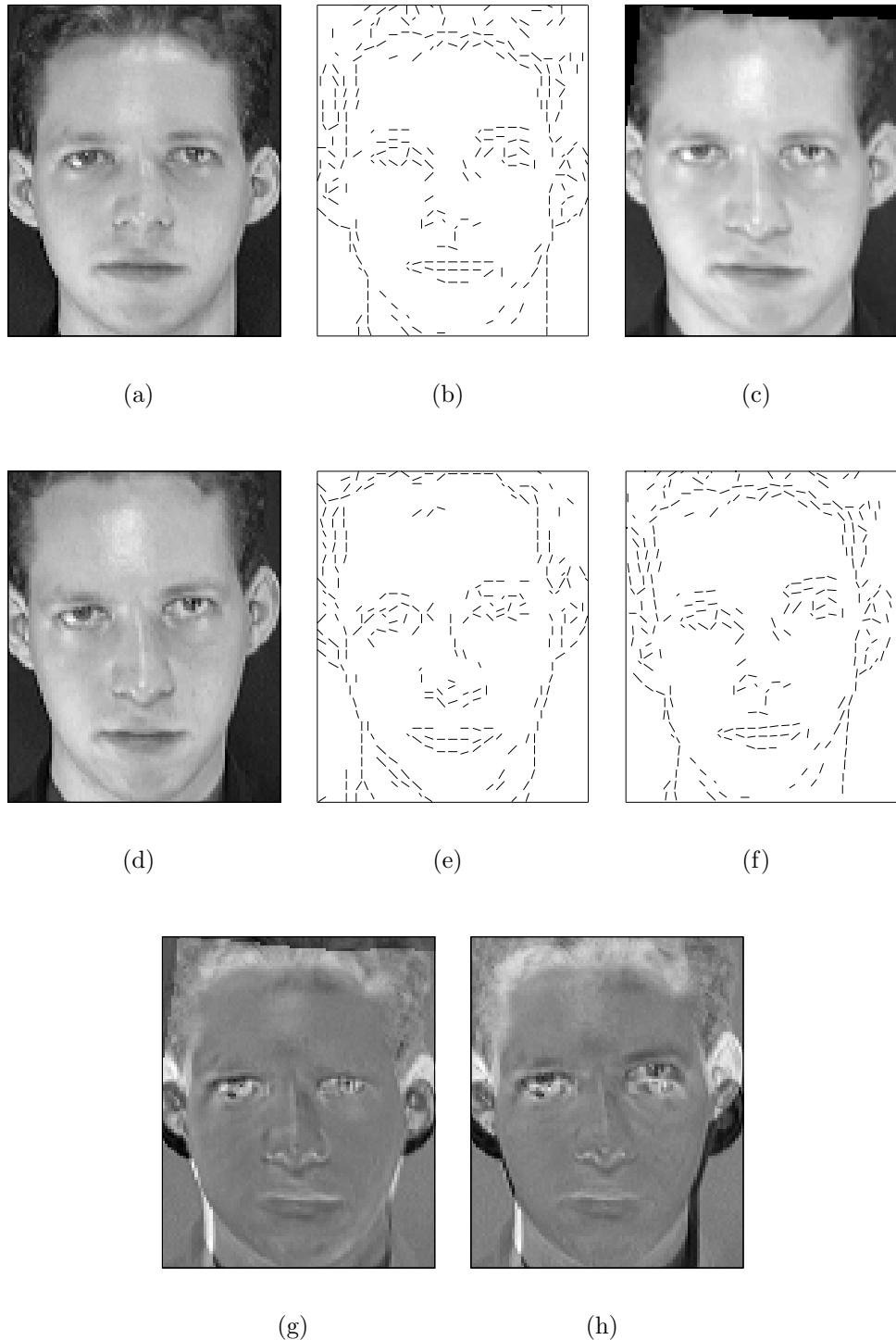


FIG. 5.8: Résultat du recalage du visage 13 avec le visage 11 : (a) Visage 11 (taille : 112×92). (d) Visage 13 (taille : 112×92). Wedgelets optimales associées à l'approximation par platelets du visage 11 (b) et du visage 13 (e) avec $sh_0 = 4$ et $\lambda = Var(I)/4$. (c) Version recalée du visage 13 (f) Alignement des sommets des wedgelets du visage 11 avec ceux du visage 13. Différences pixel-par-pixel entre les deux visages avant (h) et après (g) le recalage.

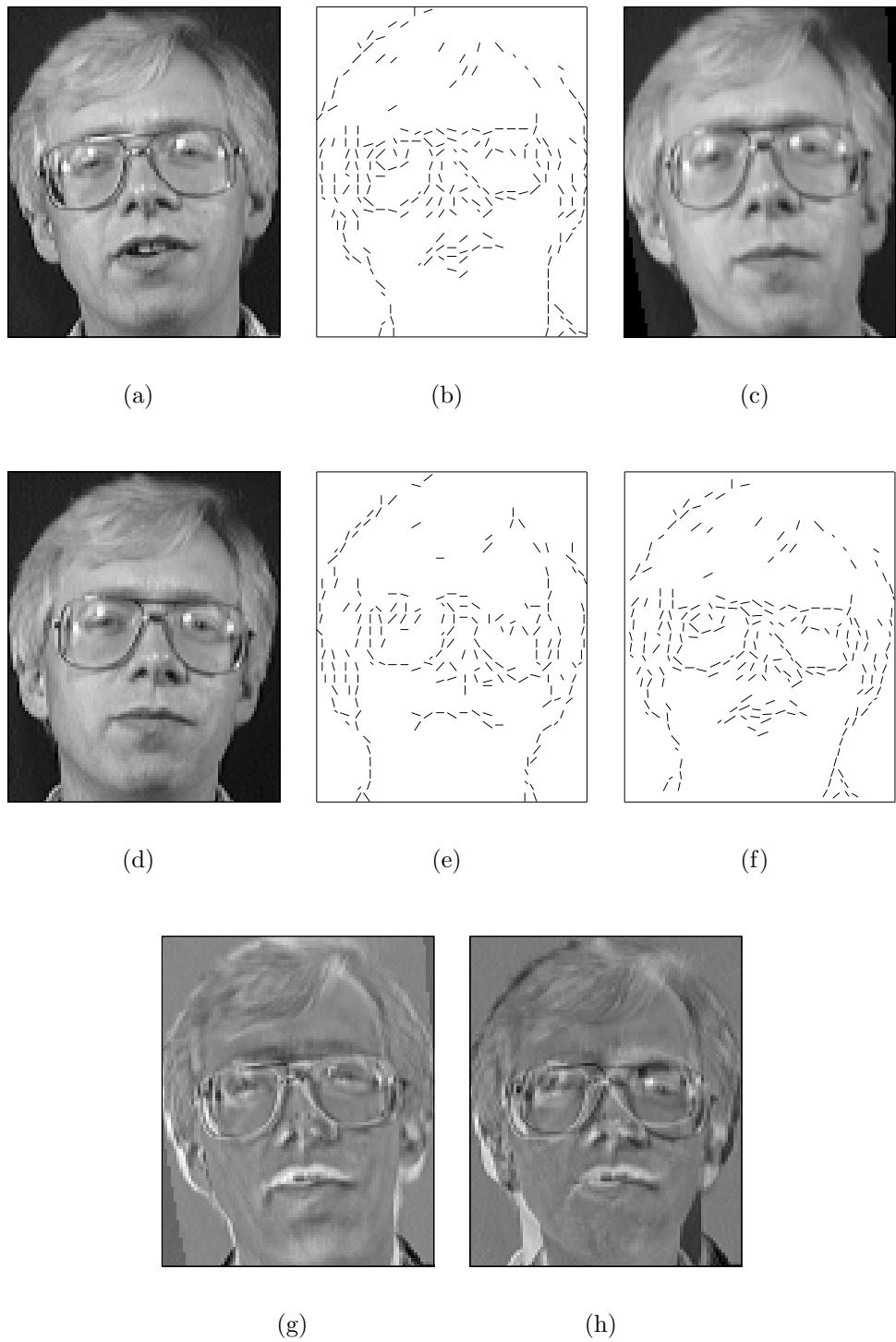


FIG. 5.9: Résultat du recalage du visage 22 avec le visage 21 : (a) Visage 21 (taille : 112×92). (d) Visage 22 (taille : 112×92). Wedgelets optimales associées à l'approximation par platelets du visage 21 (b) et du visage 22 (e) avec $sh_0 = 4$ et $\lambda = Var(I)/8$. (c) Version recalée du visage 22 (f) Alignement des sommets des wedgelets du visage 21 avec ceux du visage 22. Différences pixel-par-pixel entre les deux visages avant (h) et après (g) le recalage.

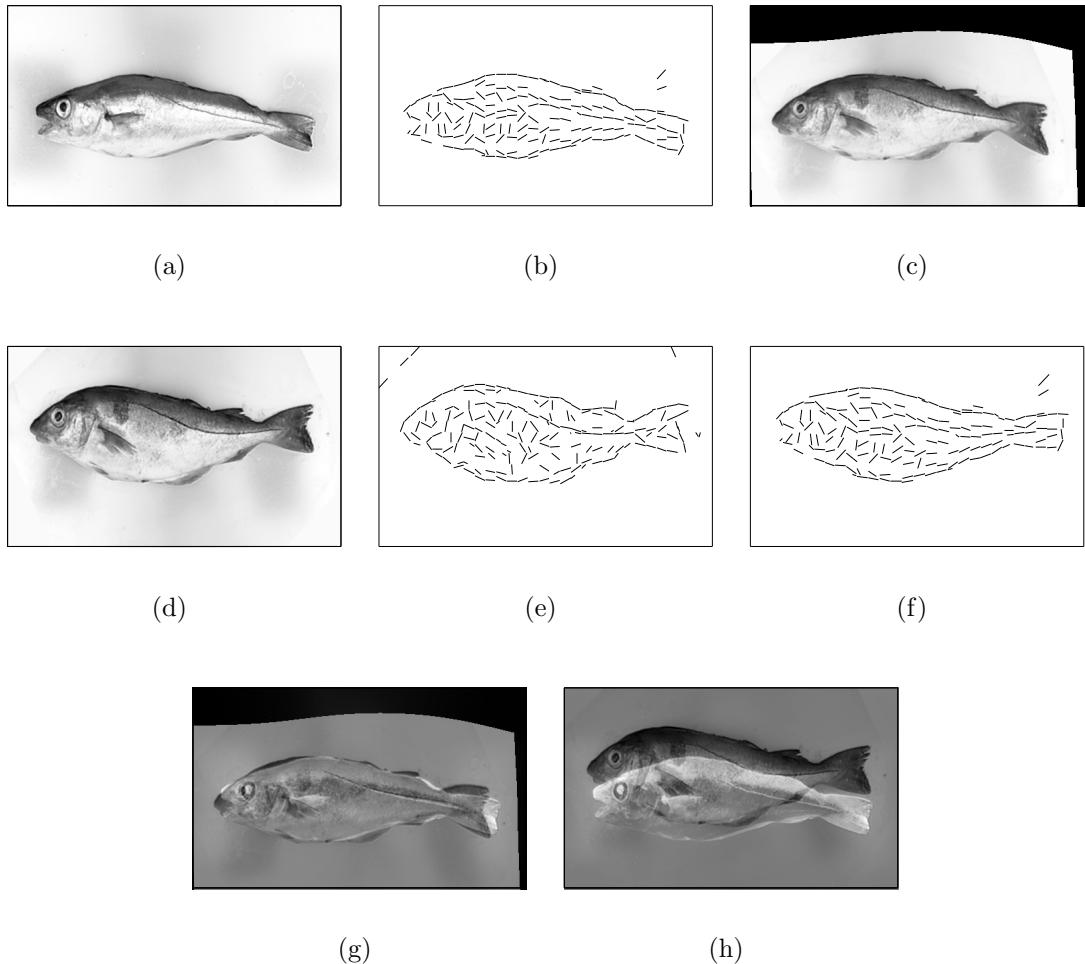


FIG. 5.10: Résultat du recalage des poissons : (a) Whiting (taille : 300×500). (d) Haddock (taille : 300×500). Wedgelets optimales associées à l'approximation par platelets du whiting (b) et du haddock 13 (e) avec $sh_0 = 16$ and $\lambda = Var(I)/6$. (c) Version recalée du haddock (f) Alignement des sommets des wedgelets du whiting avec ceux du haddock. Différences pixel-par-pixel entre les deux poissons avant (h) et après (g) le recalage.

RPM peut conduire à des transformations irréalistes pour de larges déformations entre deux objets. Une solution possible pour obtenir des difféomorphismes serait d'incorporer la méthodologie de Camion et Younes [17] dans le processus de *Deterministic Annealing* proposé dans in [22].

- Nous avons choisi d'associer les sommets des wedgelets extraits de deux images à la même échelle de sorte que la forme de ces sommets corresponde à celle des contours des deux images. Evidement, il ne s'agit pas d'une solution idéale. Tout d'abord, nous ne savons pas comment sélectionner automatiquement une échelle appropriée j_0 , et de plus l'échelle choisie peut ne pas être suffisamment fine pour correctement représenter certains contours des images (e.g. les yeux des poissons dans la Figure 5.10). Une meilleure solution serait de considérer des wedgelets à différentes échelles et de trouver un moyen d'associer directement plusieurs wedgelets (en “divisant” une wedgelet ou “réunissant” plusieurs wedgelets en une seule) plutôt que d'aligner seulement leurs sommets.
- Pour des images constituées de régions lisses séparées par des frontières régulières, la décomposition en platelets est bien adaptée pour détecter les contours, même si les images sont bruitées. Cependant pour des images plus sophistiquées (comme un visage ou un poisson), il n'est pas évident de fixer le paramètre de régularisation λ de façon à extraire les wedgelets les plus significatives, et en particulier celles qui sont importantes pour le processus de recalage. Il serait donc intéressant de pouvoir étudier théoriquement les propriétés d'estimation des platelets.
- Il a été souligné que les wedgelets à différentes échelles et positions ne se comportent pas de manière cohérente car elles forment des segments linéaires discontinus pour approximer les contours d'une image. Des contraintes géométriques devraient donc être incluses dans la complexité de la somme quadratique des résidus pénalisée associée à une WD-RDP, pour assurer que les wedgelets estimées forment des structures géométriques naturelles. Une tentative pour incorporer de telles contraintes géométriques a été récemment proposée par Romberg *et. al.* [81].
- Nous n'avons pas choisi d'étendre les résultats de détection des maxima d'ondelettes obtenus pour le cas 1D au cas d'images car nous pensons que le nombre de landmarks alors détectés est trop important (de nombreux coefficients d'ondelettes sont nécessaires pour représenter un contour) pour être utilisé dans un problème de recalage. Toutefois, il serait intéressant d'essayer d'utiliser l'une des nombreuses techniques de détection/définition de contours qui existent dans la littérature (voir Ziou et Tabbone [97] pour une revue de ces techniques) pour sélectionner uniquement, le long des contours, les landmarks qui sont utiles pour une étape de recalage (par exemple en détectant les jonctions entre deux contours données par les zero-crossings du Laplacien d'une image lissée par une Gaussienne, voir [97] pour plus de détails).
- Dans ce chapitre, nous n'avons pas discuté l'alignement de plusieurs images. Nous mentionnerons juste qu'il existe également une méthode automatique

développée par Chui *et. al.* [24] pour aligner plusieurs ensembles de landmarks non-marqués, et nous pensons que leur méthodologie pourrait être adaptée au problème du recalage de plusieurs images à partir de landmarks estimés par des méthodes multi-échelles.

Chapitre 6

Conclusion et perspectives

Dans cette thèse, nous avons choisi d'étudier le problème du recalage de structures 1D ou 2D à partir de l'alignement de certains points caractéristiques d'un ensemble de signaux. Nous avons montré que la synchronisation d'un ensemble de courbes ou d'images est une étape de pré-traitement des données qui est souvent nécessaire avant d'appliquer tout autre procédure d'inférence statistique. En particulier, la qualité de l'estimation de la forme caractéristique d'un processus observé chez différents individus dépend fortement de la qualité de l'alignement des signaux observés.

Deux étapes principales interviennent dans le recalage de fonctions à partir de leurs landmarks :

1. La définition et l'estimation des landmarks de chacune des fonctions.
2. Le calcul de transformations qui mettent en correspondance les landmarks communs à un ensemble de signaux.

Pour des signaux 1D ou 2D, des méthodes temps-échelles ont été utilisées pour estimer les points caractéristiques d'un signal, éventuellement observé avec un bruit Gaussien. Les méthodes multi-échelles permettent de caractériser l'importance des landmarks d'un signal en décomposant celui-ci à des niveaux plus ou moins fins d'approximation. Nous avons ensuite cherché à développer des méthodes automatiques pour mettre en correspondance les landmarks significatifs d'un ensemble de signaux.

6.1 Alignement de courbes

Pour le cas 1D, nous avons proposé une approche nonparamétrique pour estimer les lignes de maxima d'ondelettes et de zero-crossings d'une fonction, et avons introduit l'intensité structurelle qui permet d'associer à un signal une densité de probabilité dont les modes caractérisent l'importance des landmarks de celui-ci. De nombreuses simulations et des exemples réels ont permis de montrer que cette approche est bien adaptée à la détermination des landmarks significatifs d'un signal (en

particulier de ses extrema, points d’infexion ou singularités éventuels). Toutefois, certains aspects de cette procédure restent encore à étudier :

- *Propriétés asymptotiques des estimateurs :*

- Nous avons étudié le comportement asymptotique des maxima locaux des intensités structurelles en supposant qu’un nombre fixé (p ou q , voir section 3.6) de lignes de maxima d’ondelettes ou de zero-crossings ont été estimées. Il serait intéressant d’étudier les propriétés asymptotiques de ces estimateurs quand p et q dépendent du nombre d’observations n .
- L’estimation des landmarks d’un signal a été envisagée dans le cas d’un modèle de bruit blanc. Nous pensons que pour le problème de la détection de landmarks, il serait possible de comparer des résultats de minimalité asymptotique entre les modèles de bruit blanc et de régression non-paramétrique en suivant les idées de Donoho et Johnstone [30].
- *Masse minimale m_0* : A l’aide de simulations, nous avons montré que les modes importants des intensités structurelles correspondent aux landmarks significatifs d’un signal, et qu’il est possible de construire une mesure de modalité pour caractériser l’importance de ces modes. Le problème de la détermination du seuil de masse minimale est directement lié au problème de l’estimation du nombre de modes significatifs dans une densité de probabilité. Nous pensons que les approches du type “Excess mass” (voir Polonik [73], Cheng et Hall [21], Fisher et Marron [38]) utilisées pour tester l’existence de plusieurs modes dans une densité, pourraient être adaptées à la détermination du nombre de modes significatifs dans une intensité structurelle.
- *Procédure de bagging* : il n’est a priori pas évident de trouver une méthode automatique pour sélectionner le niveau de bruit σ_*^2 qui devrait être injecté dans les données initiales. Pour étudier de façon précise les propriétés de la procédure de bagging, il faut tout d’abord définir une mesure de l’erreur de l’estimation des zero-crossings et des maxima d’ondelettes d’un signal. Un exemple de distance entre deux fonctions basée sur la position de leurs zéro-crossings peut être trouvé dans Mallat [62]. Celle-ci pourrait certainement être utilisée pour étudier la procédure de bagging d’un point de vue théorique.

Dans de nombreux problèmes de recalage, il est souvent souhaitable d’aligner les extrema ou les points d’infexion communs à un ensemble de courbes. Afin de déterminer automatiquement les paires de landmarks qui doivent être associées pour l’alignement de deux fonctions, une nouvelle méthode, l’ ϵ -correspondance, a été proposée. Cette approche s’appuie à la fois sur le fait que la forme des pics des intensités structurelles reflète le comportement temps-échelles des landmarks et sur l’alignement global de fonctions à partir d’un algorithme de DTW. Nous avons montré à l’aide de simulations et d’exemples réels, qu’il s’agissait d’une technique efficace pour détecter les landmarks communs à un ensemble de courbes, et pour déterminer les déformations nécessaires pour l’alignement de ces dernières. Des exemples provenant du domaine biomédical nous ont permis de juger des performances de cette méthode et en particulier de la qualité de l’alignement des landmarks après recalage. Nous

avons également illustré l'intérêt de l'alignement de signaux à l'aide de la méthode de FANOVA par ondelettes récemment développée par Abramovich *et. al.* [1]. *Sans recalage* initial des données, le modèle de FANOVA utilisé suggère que les individus se comportent de manière similaire pour des conditions expérimentales proches, alors qu'*avec recalage* préalable des données on conclue que les individus se comportent de manière différente sous des conditions expérimentales similaires.

Toutefois, nous n'avons pas réussi à étudier de manière théorique la consistance de l' ϵ -correspondance, et donc l'étude asymptotique des estimateurs des fonctions de déformation. Il peut paraître un peu compliqué d'estimer tout d'abord les landmarks d'un ensemble de fonctions à l'aide des modes des intensités structurelles, de recaler ensuite ces dernières par une méthode d'alignement global, de déterminer les landmarks communs à cet ensemble de fonctions, pour finalement calculer des déformations qui alignent les données initiales. En effet, l'algorithme de Wang et Gasser [91], [92] est une méthode efficace d'alignement automatique de signaux, et pourrait donc être utilisé directement sur les fonctions originales. Toutefois, nous avons montré à l'aide de simulations que notre méthode d'alignement automatique à partir de landmarks conduisait à un meilleur alignement des points caractéristiques de deux fonctions. De plus, sur quelques exemples réels (recalage de sons de poissons, voir chapitre 4), notre méthode donne un alignement des signaux qui correspond à celui que l'on aurait fait naturellement, ce qui n'est pas le cas pour le recalage obtenu avec l'algorithme de Wang et Gasser [91] qui conduit à un alignement des extrema communs qui ne semble pas correct.

6.2 Alignement d'images

Dans la seconde partie de cette thèse, nous avons cherché à étendre la méthodologie développée pour le recalage de courbes au cas de l'alignement d'images. Etant donné que les contours sont des caractéristiques essentielles des images, qui permettent en particulier de juger visuellement la qualité de l'alignement, il nous a paru naturel d'essayer de calculer une transformation qui mette en correspondance les landmarks qui se trouvent sur les “contours communs” de deux images.

Plutôt que d'étendre au cas 2D les résultats obtenus dans le cas 1D pour la détection des lignes de zero-crossings ou de maxima d'ondelettes, nous avons cherché à utiliser de nouvelles représentations multi-échelles adaptées à la géométrie des images. Nous avons étendu l'approximation par Platelets d'une image, au cas de données observées avec un bruit Gaussien. La décomposition en Wedgelets/Platelets est bien adaptée au cas d'images composées de régions lisses séparées par des frontières régulières. Pour le cas d'images réelles, il est plus difficile de détecter automatiquement les contours significatifs. Afin de pouvoir utiliser l'algorithme TPS-RPM de mise en correspondance automatique de landmarks en 2D développé par Chui et Rangarajan [22], nous avons seulement considéré des décompositions en

Platelets à un niveau de résolution donné suffisamment fin, de sorte que les sommets des wedgelets constituent une approximation satisfaisante des contours d'une image. Il ne s'agit bien sûr pas d'une solution optimale, car nous n'avons pas réussi à incorporer dans un processus de recalage la possibilité d'associer une wedgelet à un niveau de résolution grossier avec plusieurs wedgelets à des échelles plus fines (ou l'opération inverse). Toutefois, nous avons montré l'applicabilité de cette méthode sur des problèmes réels nécessitant le recalage d'images.

Lorsque le nombre de landmarks à mettre en correspondance est trop important, l'algorithme TPS-RPM devient extrêmement lent. Ceci est principalement du au coût du calcul de la déformation Spline qui peut être de $O(N^3)$ opérations (où N est le nombre de paires de landmarks à associer). Comme souligné par Chui et Rangarajan [22], l'algorithme TPS-RPM, basé sur la notion de correspondance un-à-un, est bien adapté aux cas d'ensemble de landmarks parcimonieux (contenant peu de points). Chui *et. al.* [23] ont proposé une extension de l'algorithme TPS-RPM qui permet d'associer deux ensembles denses de landmarks. Nous pensons qu'il pourrait être intéressant d'appliquer cet algorithme au cas du recalage d'images lorsque de nombreux landmarks sont nécessaires pour décrire les contours significatifs de celles-ci.

Par ailleurs, nous avons choisi de ne pas utiliser la transformée en ondelettes 2D car de nombreux coefficients sont nécessaires pour décrire les contours d'une image. Toutefois, s'il est possible d'obtenir un algorithme qui permet de mettre en correspondance deux ensembles denses de landmarks (qui contiennent beaucoup de points), il pourrait être judicieux d'étendre au cas 2D la notion d'intensité structurelle et les résultats de détection des lignes de zero-crossings et de maxima d'ondelettes pour estimer les contours significatifs d'une image.

Bibliographie

- [1] Abramovich F., Antoniadis, A. Sapatinas, T. et Vidakovic, B. (2002). Optimal Testing in Functional Analysis of Variance. *To appear in JASA*.
- [2] Amit, Y. Grenader, U. et Piccioni, M. (1991). Structural image restoration through deformable templates, *JASA*. **86**, 376-387.
- [3] Antoniadis, A. Bigot, J. et Sapatinas, T. (2001). Wavelet estimators in nonparametric regression : a comparative simulation study. *Journal of Statistical Software*, Vol. 6, Issue 6, 1–83.
- [4] Bajcsy, R. et Kovacic, S. (1989). Multiresolution elastic matching, *Comput. Vis. Graph. Image Process.* **46**, 1-21.
- [5] Besl, P.J. et McKay, N.D. (1992). A method for registration of 3-D shapes, *IEEE Trans. Patt. Anal. Mach . Intell.* **14** (2), 239-256.
- [6] Bigot, J. (2002). A scale-space approach to landmark detection with wavelets. Technical Report TR0246, PAI (Interuniversity Attraction Pole network), <http://www.stat.ucl.ac.be/lapdp/tr2002/TR0246.pdf>.
- [7] Billingsley, P. (1968). Convergence of Probability Measures, *Wiley, New York*.
- [8] Bookstein, L.F. (1991). Morphometric tools for landmark data ; geometry and biology, *Cambridge University press*.
- [9] Breiman, L. (1996). Bagging Predictors, *Machine Learning* **24** 123–140.
- [10] Brown, L.G. (1992). A survey of image registration techniques, *ACM Compt. Surveys*. 24(4), 325-376.
- [11] Brown, L.D. et Low, M.G. (1996). Asymptotic equivalence of nonparametric regression and white noise, *Ann. Statist.* **3** No.6 , 2384–2398.
- [12] Buckheit, J.B. Chen, S. Donoho, D.L. Johnstone, I.M. et Scargle, J. (1995). About WaveLab. *Technical Report, Department of Statistics, Stanford University, USA*. Available <http://www-stat.stanford.edu/wavelab>.
- [13] Cahourt, V. Martin, L. et Amarantini, D. (2002). Static optimal estimation of joint accelerations for inverse dynamic problem solution, *Journal of Biomechanics*, **35**, 1507-1513.
- [14] Candes, E.J et Donoho, D.L. (1999). Ridgelets : the key to high-dimensional intermittency ?, *Phil. Trans. R. Soc. Lond. A..* 357 :2495-2509.

- [15] Candes, E.J et Donoho, D.L. (2000). Curvelets : a surprisingly effective non-adaptive representation of objects with edges, *Curves and Surfaces, L.L. Schumaker et. al. (eds), Vanderbilt University Press, Nashville, TN.*
- [16] Canny, J.F. (1986). A computational Approach to Edge Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **8**(6) :679-698.
- [17] Camion, V. et Younes, L. (2001). Geodesic Interpolating Splines, *Energy Minimization Methods for Computer Vision and Pattern Recognition (EMM-CVPR)*, New York. 513-527.
- [18] Chang, E. Mallat, S. et Yap, C. (2000). Wavelet-based foveation, *Applied and Computational Harmonic Analysis*. vol. 9, 312-335.
- [19] Chaudhuri, P. et Marron, J S. (1999). SiZer for exploration of structures in curves, *J. Am. Statist. Ass.* To appear.
- [20] Chaudhuri, P. et Marron, J S. (2000). Scale space view of curve estimation, *Ann. Statist.* **28** No.2 , 408–428.
- [21] Cheng, M.Y. et Hall, P. (1998). Calibrating the excess mass and dip tets of modality, *J.R.Statist.Soc. B*. **60** No.3, 579–589.
- [22] Chui, H. et Rangajaran, A. (2000). A new point matching algorithm for non-rigid registration, *EEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Vol. 2, 44-51.
- [23] Chui, H. Win, L. Dukan, J. Schultz, R. et Rangajaran, A. (2002). A unified non-rigid feature registration method for brain mapping, *Medical Image Analysis* (in press).
- [24] Chui, H. Zhang, J. et Rangajaran, A. (2002). Unsupervised learning of an atlas from unlabeled point-sets, *IEEE Trans. PAMI* (submitted).
- [25] Cohen, A. et Matei, B. (2001). Compact representations of images by edge adapted multiscale transforms, *IEEE ICIP conference, Tessaloniki*.
- [26] Coifman, R.R. et Donoho, D.L. (1995). Translation-invariant de-noising. In *Wavelets and Statistics*, Antoniadis, A. and Oppenheim, G. (Eds.), Lect. Notes Statist., **103**, pp. 125–150, New York : Springer-Verlag.
- [27] Cross, A.D.J. et Hancock, E.R. (1998). Graph matching with a dual-step EM algorithm, *IEEE Trans. Patt. Anal. Mach . Intell..* **20** (11), 1236-1253.
- [28] Daubechies, I. (1992). Ten Lectures on Wavelets, *Philadelphia : SIAM*.
- [29] Donoho, D.L. (1999). Wedgelets : Nearly minimax estimation of edges, *Annals of Statistics*. **27**(3) :859-897.
- [30] Donoho, D.L. et Johnstone, I.M. (1994). Ideal spatial adaptation by wavelet shrinkage, *Biometrika* **81**, 425–455.
- [31] Donoho, D.L. et Johnstone, I.M. (1998). Minimax estimation via wavelet shrinkage, *Ann. Statist.* **26**, 879–921.
- [32] Donoho, D.L. et Johnstone, I.M. (1999). Asymptotic minimality of wavelet estimators with sampled data, *Stat. Sinica* **9**, 1–32.

- [33] Donoho, D.L. et Huo, X. (2001). Beamlets and multiscale image analysis, Springer, *Lecture Notes in Computational Science and Engineering : Multiscale and Multiresolution Methods*.
- [34] Dryden, I.L. (1996). General shape and registration analysis, *Technical Report STAT 96/03, Department of Statistics, University of Leeds*.
- [35] Dryden, I.L. et Mardia, K.V. (1998). Statistical shape analysis, *A primer*.
- [36] Fan, J. et Lin, S K. (1998). Test of significance when data are curves, *JASA* **91**, 2, 674–688.
- [37] Faraway, J.J. (1997). Regression analysis for a functional response, *Technometrics* **39**, 254–261.
- [38] Fisher, N I. et Marron, J S. (2001). Mode testing via the excess mass estimate, *Biometrika* **88**, 2, 499–517.
- [39] Gasser, T. et Kneip, A. (1995). Searching for Structure in Curve Samples, *J. Am. Statist. Ass.* **90** No.432, 1179–1188.
- [40] Gasser, T. Kneip, A. et Kohler, W. (1991). A flexible and fast method for automatic smoothing, *J. Am. Statist. Ass.* **86** , 643–652.
- [41] Gasser, T. et Muller, H. (1984). Estimating regression functions and their derivatives by the kernel method, *Scand. J. Statist.* **11** 171–185.
- [42] Glasbey, C.A. et Mardia, K.V. (1998). A review of image warping methods, *J. Appl. Statist.* **25**, 155-171.
- [43] Glasbey, C.A. et Mardia, K.V. (2001). A penalized approach to image warping, *J. R. Statist. Soc. B.* **63**, Part 3, 465-514 (with discussions).
- [44] Gu, C. (2002). Smoothing spline ANOVA models. New York : Springer-Verlag.
- [45] Harten A. (1993). Discrete multiresolution analysis and generalized wavelets, *Journal of Applied Numerical Mathematics*. **12** :153-193.
- [46] Hsieh J.W. Liao H.Y.M. Fan K.C. Ko M.T. et Hung Y.P. (1997). Image Registration Using a New Edge-Based Approach, *Computer Vision and Image Understanding*. (67), No.2 112-130.
- [47] Huang, J.Z. (1998). Projection estimation in multiple regression with application to functional ANOVA models. *Ann. Statist.* **26** 242–272.
- [48] Jaffard, S. (1999). Mathematical Tools for Multifractal Signal Processing, *Signal Processing for Multimedia*, J.S Byrnes(Ed.), IOS Press, 111–128.
- [49] Jennrich, R.I. (1969). Asymptotic properties of non-linear least squares estimators, *Ann. Math. Statist.* **40** 633–643.
- [50] Kendall, D.G. (1984). Shape manifolds, procrustean metrics and projective spaces, *Bull. London Math. Soc* **16** :81-121.
- [51] Kneip, A. et Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves, *Ann. Statist.* **20** No.3 , 1266–1305.
- [52] Knight, K. (1999). Epi-convergence in distribution and stochastic equi-semicontinuity. *Preprint*.

- [53] Leadbetter, M R. , Lindgren, G. et Rootzén, H. (1983). Extremes and Related Properties of Random Sequences and Processes, *Springer-Verlag*.
- [54] Lepennec, E. et Mallat, S. (2001). Image compression with geometrical wavelets, *IEEE-ICP Conference, Tessaloniki*.
- [55] Lester, H. et Arridge, S.R. (1999). A survey of hierarchical non-linear medical image registration, *Pattern Recognition*. **32**, 129-149.
- [56] Li H. Manjunath B.S. et Mitra S.K. (1995). A contour-based approach to multisensor image registration, *IEEE Trans. Image Processing*. vol.4, 320-334.
- [57] Lin, Y. (2000). Tensor product space ANOVA models. *Ann. Statist.* **28** 734–755.
- [58] Lindeberg, T. (1994). Scale Space Theory in Computer Vision, *Kluwer, Boston*.
- [59] Mammen, E. Marron, J.S. Turlach, B. Wand, M.P. (2001). A general projection framework for constrained smoothing, *Statistical Science* 16(3) : 232-248.
- [60] Mammen, E. Thomas-Agnan, C. (1999) Smoothing splines and shape restrictions. *Scand. J. Statist.* **26**, 239 - 252.
Smoothing splines and shape restrictions. *Scand. J. Statist.* 26, 239 - 252 (with C. Thomas-Agnan)
- [61] Mallat, S. (1989). A theory for multiresolution signal decomposition : the wavelet representation, *IEEE Trans. Patttn. Anal. Mach. Intell.* **11** 674–693.
- [62] Mallat, S. (1991). Zero-Crossings of a Wavelet Transform, *IEEE Trans. Inform. Theory* **37** No.4, 1019–1033.
- [63] Mallat, S. (1998). A Wavelet Tour of Signal Processing, *Academic Press*.
- [64] Mallat, S. et Hwang, W.L. (1992). Singularity Detection and Processing with Wavelets, *IEEE Trans. Inform. Theory* **38** No.2, 617–643.
- [65] Matei, B. (2002). Méthodes multiéchelles non-linéaires et applications au traitement d’images, *Thèse de l’université Paris VI*.
- [66] Meyer, Y. (1992). Wavelets and Operators, *Cambridge : Cambridge University Press*.
- [67] Minnotte , M.C. et Scott, D.W. (1993). The mode tree : a tool for visualization of nonparametric density features *Journal of Computational and Graphical Statistics* **2** 51–68.
- [68] Minnotte , M.C. , Marchette , D.J. and Wegman, E.J. (1998). The bumpy road to the mode forest *Journal of Computational and Graphical Statistics* **7** 239–251.
- [69] Moshfeghi, M. (1991). Elastic matching of multimodality medical images, *Comput. Vis. Graph. Image Process.* **53**, 271-282.
- [70] Müller, H.G. (1985). Kernel estimators of zeros and of location and size of extrema of regression functions, *Scand. J. Statist* **12** 221–232.

- [71] Munoz Maldonado, Y. Staniswalis, J.G. Irwin, L.N. and Byers, D. (2002). A similarity analysis of curves, *The Canadian Journal of Statistics* **30** No.3, ???–???.
- [72] Piccioni M. , Scarlatti S. et Trouvé, A. (1998) A variational problem arising from speech recognition, *SIAM Journ. Appl. Math.*, V.58. No.3, pp. 753–771.
- [73] Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters – an excess mass approach, *Ann. Statist.* **23** 855–881.
- [74] Raimondo, M. (1998). Minimax estimation of sharp change points, *Ann. Statist.* **26**, No. 4, 1379–1397.
- [75] Ramsay, J.O. (1998) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B*, 60, 365-375.
- [76] Ramsay, J.O. et Li, X. (1998) Curve registration. *Journal of the Royal Statistical Society, Series B*, 60, 351-363.
- [77] Ramsay, J.O. et Silverman, B.W. (1997). Functional data analysis, *New York : Springer Verlag*.
- [78] Raviv, Y. et Intrator, N. (1996). Bootstrapping with Noise : An Effective Regularization Technique, *Connection Science, Special issue on Combining Estimator* **8**, 356–372.
- [79] Revuz, D. et Yor, M. (1991). Continuous Martingales and Brownian Motion, *Springer-Verlag, Berlin*.
- [80] Rockafellar R.T. et Wets R. (1998) Variational Analysis, *Springer-Verlag*.
- [81] Romberg J. Wakin M. et Baraniuk R. (2002). Multiscale Wedgelet Image Analysis : Fast Decompositions and Modeling, *IEEE International Conference on Image Processing, September*.
- [82] Sakoe, H. et Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech Signal Process.* **26**, 43-49.
- [83] Stone, C.J. Hansen, M. Kooperberg, C. et Truong, Y. (1997). Polynomial splines and their tensor products in extended linear modeling (with discussion), *Ann. Statist.* **25**, 1371-1470.
- [84] Trouve, A. (1998). Diffeomorphism groups and pattern matching in image analysis, *Int. J. of Comp. Vis.* **28**, 213-221.
- [85] Trouve, A. et Younes, L. (2000). Diffeomorphic matching problems in one dimension : designing and minimizing matching functionals, *Proceedings ECCV 2000*.
- [86] Unser, M. , Aldroubi, A. et Eden, M. (1992). On the Asymptotic Convergence of B-Spline Wavelets to Gabor Functions, *IEEE Trans. Inform. Theory* **38** No.2, 864–872.
- [87] Wahba, G. (1990). Spline models for observational data, *SIAM, Philadelphia PA*.

- [88] Wahba, G. Wang, Y. Gu, C. Klein, R. et Klein, B. (1995). Smoothing spline ANOVA for exponential families with application to the Wisconsin epidemiological study for diabetic retinopathy, *Ann. Statist.* **23**, 1865–1895.
- [89] Wang, Y. (1995). Jump and Sharp Cusp Detection by Wavelets, *Biometrika* **82** 385–397.
- [90] Wang, Y.P. et Lee, S.L. (1998). Scale-Space Derived From B-Splines, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **20** No.10, 1040–1055.
- [91] Wang, K. et Gasser, T. (1997). Alignment of curves by dynamic time warping, *Ann. Statist.* **25** No.3, 1251–1276.
- [92] Wang, K. et Gasser, T. (1999). Synchronizing sample curves nonparametrically, *Ann. Statist.* **27** No.2, 439–460.
- [93] Wood, M. (2002). Discriminant Analysis using Wavelet Derived Features, *PhD Thesis, University of Aberdeen*.
- [94] Willet, R.M. et Nowak, R.D. (2001). Platelets : A Multiscale Approach for Recovering Edges and Surfaces in Photon Limited Medical Imaging, *Tech. Rep. TREE0105, Rice University*.
- [95] Younes, L. (2000). Deformations, Warping and Object Comparison, *Tutorial* <http://www.cmla.ens-cachan.fr/~younes>.
- [96] Yuille, A.L. et Poggio, T.A. (1986). Scaling Theorems for Zero Crossings, *IEEE Trans. on Pattern Analysis and Machine Intelligence* **8** No.1, 15–25.
- [97] Ziou, D. et Tabbone, S. (1998). Edge Detection Techniques - An Overview, *International Journal of Pattern Recognition and Image Analysis* **8** No.4, 537–559.