


Numéro National de Thèse : 2016LYSEN031

**THESE DE DOCTORAT DE L'UNIVERSITE DE LYON**
opérée par :
**l'Ecole Normale Supérieure de Lyon**

**Ecole Doctorale** N°52
**Physique et Astrophysique de Lyon (PHAST)**

**Spécialité de doctorat :** Traitement du signal et des images
**Discipline :** Physique

Soutenue publiquement le 11 Octobre 2016 par :

**Jordan FRÉCON**

---

# Méthodes d'optimisation pour l'analyse de processus invariants d'échelle

---

Devant le jury composé de :

| | | |
|---|---|---|
| M. Patrice Abry | Directeur de recherche, ENS de Lyon | Directeur |
| M. Pierre-Olivier Amblard | Directeur de recherche, GIPSA-lab | Rapporteur |
| M. Jean-François Aujol | Professeur, IMB | Examinateur |
| M. Franck Picard | Directeur de recherche, LBBE | Rapporteur |
| Mme. Nelly Pustelnik | Chargée de recherche, ENS de Lyon | Encadrante |
| Mme. Béatrice Vedel | Maître de conférences, LMBA | Examinatrice |

# Résumé

L'invariance d'échelle repose sur l'intuition que les dynamiques temporelles ne sont pas gouvernées par une (ou quelques) échelle(s) caratéristique(s). Cette propriété est massivement utilisée dans la modélisation et l'analyse de données univariées issues d'applications réelles. Son utilisation pratique se heurte pourtant à deux difficultés dans les applications modernes : les propriétés d'invariance d'échelle ne sont plus nécessairement homogènes en temps ou espace ; le caractère multivarié des données rend fortement non linéaires et non convexes les fonctionnelles à minimiser pour l'estimation des paramètres d'invariance d'échelle.

La première originalité de ce travail est d'envisager l'étude de l'invariance d'échelle inhomogène comme un problème conjoint de détection/segmentation et estimation et d'en proposer une formulation par minimisation de fonctionnelles vectorielles, construites autour de pénalisation par variation totale, afin d'estimer à la fois les frontières délimitant les changements et les propriétés d'invariance d'échelle de chaque région. La construction d'un algorithme de débruitage par variation totale vectorielle à la volée est proposée.

La seconde originalité réside dans la conception d'une procédure de minimisation de fonctionnelle non convexe type « branch and bound » pour l'identification complète de l'extension bivariée, du mouvement brownien fractionnaire, considéré comme référence pour la modélisation de l'invariance d'échelle univariée. Cette procédure est mise en oeuvre en pratique sur des données de trafic Internet dans le contexte de la détection d'anomalies.

Dans un troisième temps, nous proposons des contributions spécifiques au débruitage par variation totale : modèle poissonnien d'attache aux données en relation avec un problème de détection d'états pour la fluorescence intermittente ; sélection automatique du paramètre de régularisation.

# Table des matières

**Introduction** **9**

**I Etat de l'art** **13**

**1 Analyse multifractale homogène** **15**
1.1 Invariance d'échelle et autosimilarité . . . . . 16
1.2 Analyse multifractale . . . . . 17
1.2.1 Exposant de Hölder . . . . . 17
1.2.2 Spectre multifractal . . . . . 18
1.3 Formalisme multifractal . . . . . 19
1.3.1 Fonction de partition et fonction d'échelle . . . . . 20
1.3.2 Processus monofractal et coefficients d'ondelettes . . . . . 21
1.3.3 Processus multifractal et coefficients dominants . . . . . 22
1.4 Questions ouvertes . . . . . 25

**2 Variation totale** **27**
2.1 Introduction aux problèmes $\ell_0$-TV et $\ell_1$-TV . . . . . 29
2.1.1 Problème non convexe $\ell_0$-TV . . . . . 29
2.1.2 Problème convexe $\ell_1$-TV . . . . . 30
2.1.3 Schémas itératifs d'optimisation convexe . . . . . 32
2.2 Résolution du problème $\ell_1$-TV . . . . . 34
2.2.1 Passage au dual . . . . . 35
2.2.2 Augmentation de la dimensionnalité . . . . . 37
2.2.3 Lagrangien augmenté . . . . . 38
2.2.4 Algorithme primal-dual . . . . . 39
2.2.5 Comparaison des algorithmes . . . . . 40
2.2.5.1 Choix des paramètres des algorithmes . . . . . 40
2.2.5.2 Coût de calcul . . . . . 42
2.3 Evaluation des performances d'estimation . . . . . 42
2.3.1 Rappels sur l'indice de Jaccard . . . . . 44
2.3.2 Application à la détection de discontinuité . . . . . 45
2.3.3 Choix du filtre passe-bas gaussien . . . . . 46
2.4 Extension bidimensionnelle . . . . . 47
2.4.1 Problème non convexe $\ell_0$-TV 2D . . . . . 48
2.4.2 Problème convexe $\ell_1$-TV 2D . . . . . 49
2.4.3 Problème convexe $\ell_0$-TV 2D à $Q$ niveaux . . . . . 49
2.5 Questions ouvertes . . . . . 51

**II Estimation constante par morceaux vectorielle 53**

**3 Analyse multifractale inhomogène 55**
3.1 Modèle de processus multifractals par morceaux . . . 56
3.1.1 Processus monofractal par morceaux . . . 56
3.1.2 Processus multifractal par morceaux . . . 57
3.2 Formalisme multifractal local . . . 57
3.3 Estimation vectorielle par approche TV (1D) . . . 58
3.3.1 Estimation vectorielle . . . 58
3.3.1.1 Présentation et résolution du problème $\ell_{1,2}$-TV . . . 58
3.3.1.2 Apport de l'estimation vectorielle . . . 61
3.3.2 Quantification des performances . . . 62
3.4 Segmentation vectorielle par approche TV (2D) . . . 66
3.4.1 Segmentation vectorielle . . . 66
3.4.2 Quantification des performances . . . 68
3.5 Conclusion et perspectives . . . 71

**4 Estimation $\ell_{1,2}$-TV à la volée 73**
4.1 Nature non locale du problème $\ell_{1,2}$-TV . . . 74
4.2 Conditions de Karush-Kuhn-Tucker . . . 76
4.2.1 Formulation duale . . . 76
4.2.2 Réécriture des conditions de KKT . . . 77
4.2.3 Solution approchée . . . 78
4.3 Solution algorithmique à la volée . . . 79
4.3.1 Cas idéal où $\widehat{\mathbf{z}}$ est connu . . . 79
4.3.1.1 Bornes inférieures et supérieures . . . 79
4.3.1.2 Règles de mise à jour & Règle 1 . . . 80
4.3.1.3 Prolongation du signal & Règle 2 . . . 81
4.3.1.4 Estimer la position de la discontinuité $n_{\text{rupt}}$ . . . 82
4.3.1.5 Commencer un nouveau segment . . . 83
4.3.2 Estimation du vecteur auxiliaire $\widehat{\mathbf{z}}$ . . . 83
4.3.2.1 Estimateur constant par morceaux de $\widehat{\mathbf{z}}$ . . . 84
4.3.2.2 Estimer les positions candidates de la discontinuité $n_{\text{rupt}}^{(q)}$ . . . 84
4.3.2.3 Estimer la position de la discontinuité $n_{\text{rupt}}$. . . . 85
4.3.2.4 Commencer un nouveau segment . . . 85
4.4 Résultats . . . 86
4.4.1 Configuration expérimentale . . . 86
4.4.2 Conception de $\mathcal{Q}$ . . . 86
4.4.3 Performances hors ligne . . . 88
4.4.4 Performances en ligne . . . 89
4.5 Conclusion et perspectives . . . 91

**III Autosimilarité multivariée 93**

**5 Régression non linéaire pour l'analyse d'OfBm 95**
5.1 Autosimilarité multivariée et OfBm . . . 97
5.1.1 Autosimilarité multivariée . . . 97
5.1.2 Mouvement Brownien opérateur fractionnaire . . . 97
5.1.3 Modèle paramétrique et indéterminations . . . 99
5.2 Analyse en ondelettes des Biv-OfBm . . . 99
5.2.1 Modèle paramétrique des Biv-OfBm . . . 99
5.2.2 Spectre en ondelettes . . . 101
5.2.2.1 Mélange de lois de puissance . . . 101
5.2.2.2 Indétermination supplémentaire . . . 102
5.2.2.3 Spectre en ondelettes empirique . . . 102
5.3 Régression non linéaire pour l'identification d'OfBm . . . 102
5.3.1 Formulation variationnelle du problème d'identification d'OfBm . . . 102
5.3.2 Algorithme séparation-évaluation pour les problèmes de minimisation globale . . . 104
5.3.3 Algorithme séparation-évaluation pour l'identification de Biv-OfBm . 105
5.3.3.1 Relaxation convexe . . . 105
5.3.3.2 Encadrement via l'arithmétique par intervalles . . . 106
5.3.3.3 Algorithme séparation-évaluation . . . 107
5.4 Performances d'estimation : étude théorique asymptotique . . . 107
5.4.1 Normalité asymptotique du spectre d'ondelettes . . . 108
5.4.2 Consistance de l'estimateur $\widehat{\Theta}_N^M$ . . . 109
5.4.3 Normalité asymptotique de l'estimateur $\widehat{\Theta}_N^M$ . . . 110
5.5 Performances d'estimation : étude empirique . . . 111
5.5.1 Configurations expérimentales . . . 111
5.5.2 Performances d'estimation . . . 111
5.5.3 Coût de calcul . . . 113
5.5.4 Nombre d'échantillons vs. précision. . . . 115
5.5.5 Normalité asymptotique de $\widehat{\Theta}$ . . . 115
5.6 Conclusion et perspectives . . . 116

**6 Détection d'anomalies dans le trafic internet 119**
6.1 Identifiation paramétrique de Biv-OfGns . . . 121
6.1.1 Bruit Gaussien opérateur fractionnaire . . . 121
6.1.2 Intégration fractionnaire et analyse en ondelettes . . . 121
6.1.3 Spectre en ondelettes paramétrique de Biv-OfGn . . . 121
6.1.4 Procédure d'identification . . . 122
6.1.5 Performances . . . 123
6.2 Application aux données internet . . . 123
6.2.1 Données MAWI et projections aléatoires . . . 123
6.2.2 Rôle du paramètre d'intégration fractionnaire . . . 124
6.2.3 Résultats . . . 125
6.3 Conclusion et perspectives . . . 128

**IV Développements méthodologiques autour de la TV** **131**

**7 Processus de Poisson homogènes par morceaux** **133**
7.1 Présentation du problème . . . 134
7.1.1 Modèle . . . 134
7.1.2 Impact de l'aggrégation . . . 135
7.2 Statistique Poissonienne vs. Gaussienne . . . 136
7.2.1 Divergence de Kullback-Leibler . . . 136
7.2.2 Transformée d'Anscombe . . . 137
7.3 Méthodologies . . . 139
7.3.1 Méthode du seuil . . . 139
7.3.2 Méthodes d'estimation par approche TV . . . 140
7.3.2.1 Estimées par approche TV . . . 140
7.3.2.2 Réestimation à 2 niveaux . . . 141
7.3.3 Méthode de segmentation par approche TV . . . 142
7.4 Comparaison des différentes méthodes . . . 142
7.4.1 Configuration expérimentale . . . 142
7.4.2 Quantification des performances . . . 144
7.5 Conclusion et perspectives . . . 146

**8 Estimation du paramètre de régularisation $\ell_0$-TV** **149**
8.1 Paramétrisation du problème . . . 150
8.1.1 Paramétrisation des discontinuités $\boldsymbol{r}$ . . . 150
8.1.2 Paramétrisation des valeurs des segments $\boldsymbol{\mu}$ . . . 150
8.1.3 Reformulation du problème . . . 151
8.2 Dérivation Bayésienne de la pénalisation $\phi$ . . . 151
8.2.1 Modèle hiérarchique Bayésien . . . 151
8.2.2 Critère MAP joint . . . 152
8.3 Solution algorithmique . . . 154
8.4 Sélection automatique de $\chi$ : Illustration et performances . . . 155
8.4.1 Configuration expérimentale et choix des hyperparamètres . . . 155
8.4.2 Illustration du principe de sélection automatique de $\chi$. . . . 155
8.4.3 Quantification des performances . . . 157
8.4.4 Comparaison contre de l'inférence bayésienne . . . 158
8.4.5 Robustesse de l'estimateur proposé . . . 159
8.5 Conclusion et perspectives . . . 160

**Conclusion** **163**

**A Définitions utiles en optimisation convexe** **169**

**B Algorithme de segmentation vectorielle couplée** **171**
B.1 Enoncé du problème . . . 171
B.2 Solution algorithmique . . . 172
B.3 Quantification des performances . . . 174

**C Démonstrations** **175**
C.1 Démonstration du Théorème 5.2 . . . . . 175
C.2 Démonstration du Théorème 5.3 . . . . . 176
C.3 Discussion à propos de la Remarque 5.5 . . . . . 177

**D Estimateurs bayésiens** **179**

**Bibliographie** **179**

# Introduction

**Invariance d'échelle : contexte.** L'invariance d'échelle est une propriété omniprésente dans une grande variété d'applications réelles et offre une description parcimonieuse de nombreux phénomènes naturels (e.g., turbulence [95], rythme cardiaque [71]) et humains (e.g., trafic internet [83, 2], art [119]). Le paradigme de l'invariance d'échelle repose sur l'hypothèse que la dynamique temporelle des données n'est pas déterminée par une ou quelques échelles caractéristiques, mais par un large continuum d'échelles. L'objectif de l'analyse d'échelle est alors de modéliser, d'estimer et de valider la relation entre les différentes échelles.

Un modèle d'invariance d'échelle couramment utilisé est celui de l'autosimilarité [184] énonçant qu'un processus $X$ ne peut pas être distingué de ses copies dilatées, i.e.,

$$X(t) = a^H X(t/a) \quad , \forall a > 0 \tag{1}$$

L'information clef sur la dynamique de $X$ réside alors dans un seul paramètre $H$, appelé exposant de Hurst. Dans ce cas, pour rendre compte du fait qu'un seul exposant de loi d'échelle décrit entièrement $X$, le processus est dit *monofractal*. Une méthode efficace en pratique pour estimer $H$ repose sur l'utilisation de quantités multi-échelles $T(a,t)$ dépendant à la fois de la position $t$ et de l'échelle $a$ (e.g., accroissements, coefficients d'ondelettes, coefficients d'ondelettes dominants) [128, 213]. L'autosimilarité de $X$ se traduit ainsi par un comportement en loi de puissance en fonction de l'échelle d'analyse $a$ de la fonction de partition suivante

$$S_q(a) = \mathbb{E}|T(a,\cdot)|^q \underset{a\to 0}{\sim} a^{qH}. \tag{2}$$

Plus récemment, des modèles plus riches ont été développés, i.e., processus *multifractals*, où la quantité $H$ est remplacée par un ensemble de quantités locales $h(t)$, dites exposants de Hölder [114]. Il advient alors que

$$S_q(a) = \mathbb{E}|T(a,\cdot)|^q \underset{a\to 0}{\sim} a^{\zeta(q)}, \tag{3}$$

où $\zeta$ est une fonction concave non linéaire étroitement liée à la répartition géométrique des $(h(t))_{t\in\mathbb{R}}$. En pratique, l'information sur la dynamique de $X$ est décrite par le vecteur $(c_1,\ldots,c_p)$ des coefficients du développement polynomial de $\zeta$ en 0, i.e., $\zeta(q) = c_1 q + c_2 \frac{q^2}{2} + \ldots + c_p \frac{q^p}{p!}$ [13]. L'estimation de $\zeta$ à partir de (3) est une procédure d'estimation appelée *formalisme multifractal*.

**Invariance d'échelle : limitations et questions ouvertes.** Dans sa formulation actuelle, le formalisme multifractal suppose *a priori* que les propriétés d'invariance d'échelle de l'objet étudié sont homogènes en $t$, c'est-à-dire, que $H$ et $(c_1,\ldots,c_p)$ ne varient pas au cours du temps. Cependant, en pratique, $H$ et $(c_1,\ldots,c_p)$ peuvent varier. De plus, dans nombre d'applications, l'information peut davantage être contenue dans le changement des propriétés multifractales que dans les valeurs des paramètres qui les caractérisent. Ce phénomène est notamment visible dans le cas où l'estimation de l'exposant de l'Hölder $h$ du rythme cardiaque foetal au cours

du temps permettrait de surveiller en temps réel l'état de santé du foetus [190]. La même question se pose également dans le cadre bidimensionnel, par exemple pour distinguer sur une image deux textures, e.g., nuage et neige, chacune caractérisée par sa propre collection de coefficients $(c_1, \ldots, c_p)$ [181].

Une autre limitation de l'actuel formalisme multifractal est due au fait que la plupart des contextes d'applications modernes implique l'enregistrement de séries temporelles multivariées qui nécessitent alors d'être analysées conjointement. Cependant, leur analyse multirésolution se heurte à une difficulté majeure : une extension multivariée de la fonction de structure $S_q(a)$ pour $M$ signaux monofractals implique *a priori* le mélange de plusieurs lois de puissance faisant intervenir différents exposants $H_1 \leq \ldots \leq H_M$ [4]. De surcroît, $S_q(a)$ est dominée par la loi de puissance ayant l'exposant le plus élevé, limitant ainsi l'application des méthodes univariées à la seule estimation de $H_M$. En dépit de son importance théorique et pratique, la construction d'un paradigme d'estimation multivarié complet est toujours un problème ouvert dans la littérature.

**Optimisation : contexte.** L'optimisation est présente dans différents contextes applicatifs de traitement du signal et de l'image allant de l'estimation de données manquantes [32, 14, 7, 8] à la reconstruction, en passant par l'apprentissage de dictionnaires, de classifieurs ou de modèles.

La formulation générique en optimisation consiste à savoir résoudre

$$\min_{\boldsymbol{x} \in \mathcal{X}} f(\boldsymbol{x}) \tag{4}$$

où $\mathcal{X}$ est l'espace de recherche et $f$ le critère permettant d'évaluer chaque élément $\boldsymbol{x} \in \mathcal{X}$. Selon la nature de $f$, nous pouvons classifier les problèmes en deux catégories : les problèmes où $f$ est convexe, qui disposent d'un cadre théorique bien défini et dont les algorithmes associés permettent de trouver efficacement le minimum, et les problèmes où $f$ est non convexe, qui sont sensiblement plus difficiles en raison de l'existence potentielle de multiples minima locaux, et dont la résolution constitue un axe majeur de la recherche actuelle.

D'un côté, l'optimisation convexe a connu un nouvel essor ces dernières années avec le développement de méthodes proximales pour résoudre des problèmes où $f$ est non lisse (i.e., non différentiable sur tout son domaine) [49, 38, 50]. En pratique, la non-différentialité de $f$ peut être due à l'utilisation de normes qui favorisent la parcimonie (e.g., norme $\ell_1$).

De l'autre côté, la plupart des problèmes d'optimisation non convexe sont difficiles à résoudre en temps polynomial. Certains problèmes peuvent être résolus exactement et efficacement par des algorithmes de programmation dynamique (cf., e.g., [217, 216, 192, 93, 193]). Lorsque le problème implique une pluralité d'inconnues de diverses natures, les algorithmes de séparation–évaluation (≪ branch and bound ≫) [120, 44, 160] permettent d'obtenir un encadrement de la solution globale avec toutefois un temps de calcul plus élevé. Autrement, une solution locale peut être obtenue efficacement par mises à jour alternées des variables.

**Optimisation : limitations et questions ouvertes.** En pratique et selon l'application, le critère $f$ est généralement choisi sous la forme

$$f(\boldsymbol{x}) = \sum_k \lambda_k f_k(\boldsymbol{x}) \tag{5}$$

où les fonctions $f_k$ peuvent modéliser des mesures d'attache aux données $\boldsymbol{y}$ ou peuvent imposer des conditions de régularité à la solution (e.g, parcimonieuse, constante par morceaux, linéaire par morceaux). Les hyperparamètres $\lambda_k$ sont des scalaires permettant de relativiser la contribution des différents $f_k$.

Les hyperparamètres $\lambda_k$ sont souvent choisis *a priori* mais leurs choix sont cruciaux étant donné qu'ils impactent fortement la qualité de la solution. On retrouve par ailleurs, dans la littérature de sélection de paramètres, différents critères pour leur estimation (cf. e.g., validation croisée, périodogramme cumulé normalisé, courbe-L [103], risque généralisé de Stein [59]). Il a récemment été proposé d'inclure leur estimation au sein d'algorithmes de Monte-Carlo par chaînes de Markov [165] mais la solution s'avère trop coûteuse en temps de calcul pour des problèmes de grande dimension. Ainsi, l'estimation rapide et non supervisée des hyperparamètres $(\lambda_k)_k$ représente encore un défi de taille.

Généralement, un coût de calcul élevé est nécessaire pour résoudre le problème (4), en raison de la nature itérative des algorithmes mis en oeuvre. Cependant, dans de nombreux cas pratiques, connaître la solution au fur et à mesure de l'acquisition des données $\boldsymbol{y}$ en temps réel pourrait s'avérer décisif [97, 190]. La solution pourrait être calculée sur une fenêtre glissante mais cette technique engendre le double inconvénient d'introduire une échelle de temps arbitraire et de ne pas permettre la reconstruction de la solution sur la totalité de l'enregistrement. Afin de contourner ces limitations, les algorithmes dits *à la volée* peuvent être utilisés, lorsque les données sont scalaires, i.e., $\mathcal{X} = \mathbb{R}^N$, en introduisant une fenêtre de taille variable au-delà de laquelle la solution ne sera pas modifiée [52]. Toutefois, leur extension au cadre vectoriel demeure difficile [58].

**Contributions.** Le travail effectué dans cette thèse consiste à formaliser l'analyse des processus invariants d'échelle multivariés et non homogènes par la résolution de problèmes d'optimisation.

La Partie I dresse un état de l'art concernant les deux outils mathématiques déclinés tout au long de ce travail. Le premier est celui de l'invariance d'échelle homogène dont la description fait l'objet du Chapitre 1. Nous y présenterons le mouvement Brownien fractionnaire (fBm) comme processus autosimilaire univarié de référence ainsi que le formalisme multifractal permettant d'estimer dans la pratique la quantité scalaire/vectorielle caractérisant les propriétés mono/multi-fractales de processus univariés. Le second outil consiste à modéliser un signal ou une image comme un ensemble de régions de supports disjoints, chacune caractérisée par des propriétés statistiques (e.g. propriétés d'invariance d'échelle) différentes. Le Chapitre 2 établit alors une revue de l'état de l'art des approches variationnelles reposants sur l'utilisation de la variation totale (TV) pour détecter les discontinuités, quand leurs nombres et positions sont inconnus, comme le sont les propriétés sur chaque région.

Le problème d'estimation de discontinuités dans des signaux ou images vectorielles est ensuite abordé dans la Partie II. Nous envisageons d'employer les approches TV au Chapitre 3 combinées à une version locale du formalisme multifractal afin de déterminer les changements de propriétés de processus multifractals homogènes par morceaux. Dans les différents cas considérés, nous discuterons de l'apport des approches TV vectorielles. Ainsi, motivé par l'existence d'un algorithme à la volée dans le cas scalaire, le Chapitre 4 élabore un algorithme à la volée permettant de détecter les discontinuités communes à toutes les composantes de séries temporelles vectorielles.

La Partie III se concentre sur l'étude de l'extension multivariée du modèle d'autosimilarité. L'identification paramétrique complète des mouvements Browniens opérateurs fractionnaires (OfBm), définis comme l'extension multivariée du fBm, est obtenue par la résolution d'un problème d'optimisation non convexe que nous détaillons au Chapitre 5. Un algorithme de séparation-évaluation est également développé et validé pour l'analyse d'OfBms bivariés (Biv-OfBm). La méthode est ensuite évaluée et testée au Chapitre 6 sur des données réelles issues du trafic internet modélisées par les incréments d'un Biv-OfBm dont la valeur des paramètres permet d'identifier la présence d'anomalies dans le trafic.

Dans la Partie IV, nous rapportons divers développements méthodologiques autour de la TV afin de répondre à plusieurs questions soulevées tout au long de ce manuscrit. Notamment, un cas pratique est abordé au Chapitre 7 motivé par le phénomène de fluorescence intermittente où l'intensité d'émission d'un processus poissonien alterne aléatoirement entre deux états. La question visant à analyser la distribution des durées de ces deux états est ici revisitée par les approches TV pour détecter en amont les discontinuités d'intensité. Cependant, les performances d'estimation des méthodes TV dépendent fortement du réglage d'un paramètre de régularisation *a priori* inconnu. C'est dans ce contexte que se place le Chapitre 8 qui propose l'estimation rapide et non supervisée de ce paramètre via la conception d'une procédure variationnelle bayésienne.

# Première partie

# Etat de l'art

# Chapitre 1

# Analyse multifractale homogène

**Sommaire**


**1.1 Invariance d'échelle et autosimilarité . . . . . 16**
**1.2 Analyse multifractale . . . . . 17**
1.2.1 Exposant de Hölder . . . . . 17
1.2.2 Spectre multifractal . . . . . 18
**1.3 Formalisme multifractal . . . . . 19**
1.3.1 Fonction de partition et fonction d'échelle . . . . . 20
1.3.2 Processus monofractal et coefficients d'ondelettes . . . . . 21
1.3.3 Processus multifractal et coefficients dominants . . . . . 22
**1.4 Questions ouvertes . . . . . 25**


**Invariance d'échelle.** L'invariance d'échelle est aujourd'hui reconnue comme une propriété omniprésente dans une grande variété d'applications réelles de différentes natures (cf., e.g. [144] en turbulence, [181] en géophysique, [118] en imagerie médicale et [145] en finance). Le paradigme de l'invariance d'échelle est basé sur l'hypothèse que la dynamique temporelle ou spatiale des données n'est pas déterminée par une ou quelques échelles caractéristiques, mais par une large gamme d'échelles. Par conséquent, l'analyse des données ne doit pas reposer sur l'utilisation de plusieurs échelles caractéristiques mais doit au contraire permettre d'identifier, d'estimer et de valider les mécanismes reliant les échelles entre elles.

**Fractales.** Les fractales constituent l'exemple phare de l'invariance d'échelle [142, 143, 146]. Ce sont des objets mathématiques déterministes autosimilaires desquels il émerge à chaque échelle une structure identique au tout, à un facteur de dilation près. Elles sont conçues comme le point fixe d'opérations relatant les échelles de l'objet entre elles. L'adjectif fractal a été introduit par Mandelbrot pour décrire leur complexité géométrique qui ne pouvait pas être caractérisée par une dimension entière. La dimension fractale fournit alors une description globale de leur complexité et de leur irrégularité.

**Analyse multifractale.** La notion d'autosimilarité peut être étendue au cadre stochastique [184] où la dimension fractale est alors directement liée à l'exposant de Hurst qui mesure la régularité du processus (voir Section 1.1).

Le modèle multifractal se présente comme une généralisation de l'autosimilarité où la régularité peut fluctuer ponctuellement. Les processus multifractals ne sont alors plus décrits par un seul exposant mais par tout le spectre des dimensions fractales des ensembles des points de même régularité, appelé spectre multifractal, et décrit à la Section 1.2.

L'invariance d'échelle se manifeste également par le comportement en lois de puissance de quantités multi-résolution. Une méthode efficace en pratique, appelée formalisme multifractal, vise à estimer l'exposant de cette loi de puissance qui est intimement liée au spectre multifractal. De nombreux travaux promeuvent l'utilisation des coefficients d'ondelettes comme quantité multi-résolution [5, 13, 114]. Plus récemment, il a été proposé l'utilisation des coefficients d'ondelettes dominants [128, 213] afin de reconstruire l'entièreté du spectre multifractal. La présentation de ce formalisme fait l'objet de la Section 1.3.

## 1.1 Invariance d'échelle et autosimilarité

Les processus stochastiques autosimilaires constituent le cadre mathématique fondamental pour la modélisation des phénomènes d'invariance d'échelle. L'autosimilarité énonce qu'un signal $X$ ne peut pas être distingué de ses copies dilatées, i.e.,

**Définition 1.1** (Autosimilarité [184])**.** *Un processus $\{X(t)\}_{t\in\mathbb{R}^d}$ est dit autosimilaire s'il vérifie*

$$\{X(t)\}_{t\in\mathbb{R}^d} \overset{d.d.f.}{=} \{a^H X(t/a)\}_{t\in\mathbb{R}^d} \quad (\forall a>0), \tag{1.1}$$

*où $\overset{d.d.f.}{=}$ représente l'égalité entre les distributions à dimension finie. L'information clef sur la dynamique d'invariance d'échelle est résumé par un seul paramètre $0<H<1$, appelé exposant d'autosimilarité ou exposant de Hurst.*

En particulier, le mouvement brownien fractionnaire, i.e., le seul processus gaussien, autosimilaire, à accroissements stationnaires, a largement été utilisé comme le processus de référence pour modéliser les propriétés d'échelle de signaux univariés issus du monde réel. Nous rappelons ci-dessous sa définition et nous nous intéresserons à son extension multivariée au Chapitre 5.

**Définition 1.2** (Mouvement brownien fractionnaire (fBm) [146, 5])**.** *Le mouvement brownien fractionnaire $\{X(t)\}_{t\in\mathbb{R}}$ d'exposant $0<H<1$ est le processus gaussien de moyenne nulle tel que*

$$\begin{cases} X(0)=0, \\ X(t+\delta)-X(t) \overset{d.d.f.}{=} \mathcal{N}(0,\sigma|\delta|^H), \quad (\forall t\in\mathbb{R})\ (\forall \delta\in\mathbb{R}). \end{cases} \tag{1.2}$$

*Il apparaît donc que $\{X(t) \overset{d.d.f.}{=} \mathcal{N}(0,\sigma|t|^H)\}_{t\in\mathbb{R}}$ est un processus gaussien, autosimilaire, à accroissements stationnaires entièrement caractérisé par sa fonction de covariance*

$$(\forall t\in\mathbb{R})\ (\forall s\in\mathbb{R}) \quad \mathbb{E}X(t)X(s)=\frac{\sigma^2}{2}\left(|t|^{2H}+|s|^{2H}-|t-s|^{2H}\right) \tag{1.3}$$

*où $\sigma^2=\mathbb{E}X(1)^2$.*

Nous remarquons que la Définition 1.2 correspond à celle du mouvement brownien lorsque $H=1/2$. Tout comme le processus des accroissements d'un mouvement brownien est un bruit blanc gaussien, nous définissons le bruit gaussien fractionnaire dont nous ferons usage dans le cadre multivarié au Chapitre 6.

**Définition 1.3** (Bruit gaussien fractionnaire (fGn))**.** *Nous appelons bruit gaussien fractionnaire, le processus des accroissements $\{X^\delta(t)\equiv X(t+1)-X(t)\}_{t\in\mathbb{R}}$ où $X$ est un mouvement brownien fractionnaire.*

**Remarque 1.1.** *Nous pouvons montrer que la corrélation entre les accroissements successifs d'un fBm* $\{X\}_{t\in\mathbb{R}}$*, i.e.,*

$$(\forall t \in \mathbb{R}) \quad \mathbb{E}X^{\delta}(t)X^{\delta}(t+1) = \sigma^2\left(2^{2H-1}-1\right) \tag{1.4}$$

*est positive si* $H > 1/2$*, négative si* $H < 1/2$ *et nulle si* $H = 1/2$*.*

Deux exemples de fBm unidimensionnels sont illustrés sur la Figure 1.1 pour $H = 0.25$ (gauche) et $H = 0.75$ (droite). On observe que la réalisation pour $H = 0.25$ (voir Figure 1.1, gauche) apparaît visuellement plus irrégulière que celle pour $H = 0.75$ (voir Figure 1.1, droite). En effet, puisque les réalisations d'un fBm sont des courbes fractales dont la dimension de Hausdorff vaut $2-H$, le comportement de $X$ est d'autant plus irrégulier que $H$ est proche de 0. De plus, nous remarquons que la variance de $X$ est d'autant plus grande que $H \to 1$ (cf. Définition 1.2 et Remarque 1.1).

Dans le cas bidimensionnel, nous aurions employé l'adjectif « rugueux » pour désigner des textures invariantes d'échelle. Il apparaît alors que la notion de régularité (et de rugosité) est intrinsèquement liée aux propriétés d'invariance d'échelle. Ce lien fait l'objet de la section suivante.

## 1.2 Analyse multifractale

Cependant, la condition d'autosimilarité (1.1) est très restrictive. Une condition plus souple est la suivante

$$\{X(t)\}_{t\in\mathbb{R}^d} \overset{d.d.f.}{=} \{\omega_a X(t/a)\}_{t\in\mathbb{R}^d} \quad (\forall a > 0), \tag{1.5}$$

où $\omega_a = a^{h(t)}$ est une variable aléatoire dépendant de l'échelle $a$. La condition (1.5) peut être interprétée comme une extension de la condition d'autosimilarité (1.1) où $H$ est remplacé par une quantité stochastique $h$ caractérisant la régularité locale et dont la description fait l'objet de cette section.

### 1.2.1 Exposant de Hölder

Il existe diverses façons de mesurer la régularité locale d'un processus autour d'un point donné. Dans cette section, nous nous concentrerons sur l'exposant de régularité ponctuelle de Hölder dont nous rappelons la définition. Le lecteur peut se référer à [185, 114] et les références qui y sont mentionnées pour une étude détaillée.

**Définition 1.4** (Classe $\mathcal{C}^{\alpha}$)**.** *Soit un processus continu* $\{X(t)\}_{t\in\mathbb{R}^d}$*. Pour tout* $t \in \mathbb{R}^d$*,* $X \in \mathcal{C}^{\alpha}(t)$ *s'il existe une constante* $C > 0$ *et un polynome de degré plus petit que la partie entière de* $\alpha$ *tels que, pour tout* $s$ *dans le voisinage de* $t$*,*

$$|X(t) - P(s-t)| \leq C|s-t|^{\alpha}. \tag{1.6}$$

**Définition 1.5** (Exposant de Hölder [114])**.** *On appelle régularité ou* exposant de Hölder *en un point* $t \in \mathbb{R}^d$ *de* $X$*, la quantité*

$$h(t) = \sup\{\alpha \geq 0 \mid X \in \mathcal{C}^{\alpha}(t)\}. \tag{1.7}$$

L'exposant de Hölder est défini en chaque point $t \in \mathbb{R}^d$ et décrit les fluctuations de régularité (ou singularité) locale de $X$. Pour certains processus, $h$ peut varier de façon continue voire discontinue d'un point à l'autre. Dans ce dernier cas, le calcul numérique de $h$ est instable. Nous préférons à l'information locale $h$, une description globale décrite dans la section suivante.

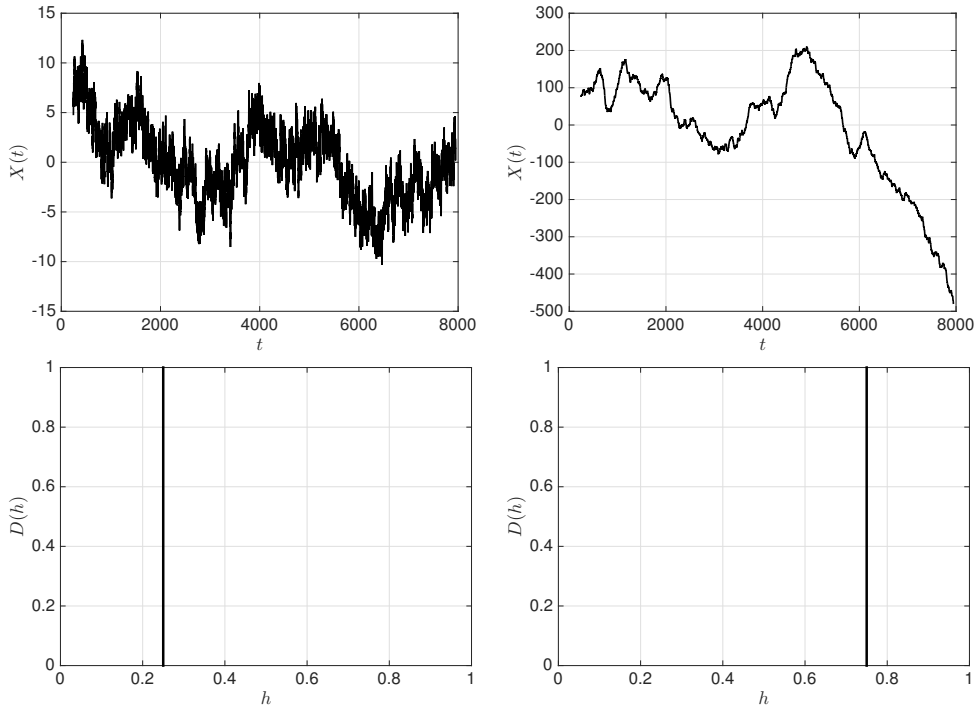



FIGURE 1.1 – **Exemples de processus monofractals unidimensionnels**. Haut : Deux réalisations de mouvements browniens fractionnaires. Bas : spectre multifractal correspondant. Deux valeurs de $H$ sont examinées : à gauche, $H = 0.25$ et à droite $H = 0.75$.

### 1.2.2 Spectre multifractal

Afin d'analyser la collection des $\{h(t)\}_{t\in\mathbb{R}^d}$, on s'intéresse au spectre multifractal $\mathcal{D}$, également appelé spectre des singularités, qui fournit une description globale de la répartition géométrique des $\{h(t)\}_{t\in\mathbb{R}^d}$.

**Définition 1.6** (Spectre multifractal [142, 80],)**.** *Soit un processus $\{X(t)\}_{t\in\mathbb{R}^d}$ ainsi que $\{h(t)\}_{t\in\mathbb{R}^d}$ l'ensemble de ses exposants de Hölder. Soit l'ensemble iso-Hölder $E(h) = \{t \in \mathbb{R}^d \mid h(t) = h\}$. On appelle spectre multifractal $\mathcal{D}$ la collection des dimensions de Hausdorff de $(E(h))_{h\geq 0}$ i.e.,*

$$\mathcal{D}(h) = \dim E(h). \tag{1.8}$$

Lorsque l'exposant de Hölder $h$ vaut $H$ en tout point, le processus est dit *monofractal* et le support du spectre multifractal est réduit à un singleton. En particulier, les processus autosimilaires (voir Définition 1.1) forment une famille de processus monofractals (voir *a contrario* les processus de Lévy [113]). Pour les fBm (voir Définition 1.2), le spectre multifractal prend la forme suivante :

**Exemple 1.1** (Spectre multifractal d'un fBm)**.** *Soit un fBm $\{X(t)\}_{t\in\mathbb{R}}$ caractérisé par l'exposant de Hurst $0 < H < 1$, alors son spectre multifractal $\mathcal{D}$ s'écrit :*

$$(\forall h \geq 0) \quad \mathcal{D}(h) = \begin{cases} 1 & , \text{ si } h = H, \\ -\infty & , \text{ sinon.} \end{cases} \tag{1.9}$$

*Deux réalisations de fBm unidimensionnels sont illustrées par la Figure 1.1 pour $H = 0.25$ (gauche) et $H = 0.75$ (droite). Les spectres multifractals correspondants, également représentés, sont tout deux des diracs en $H$.*

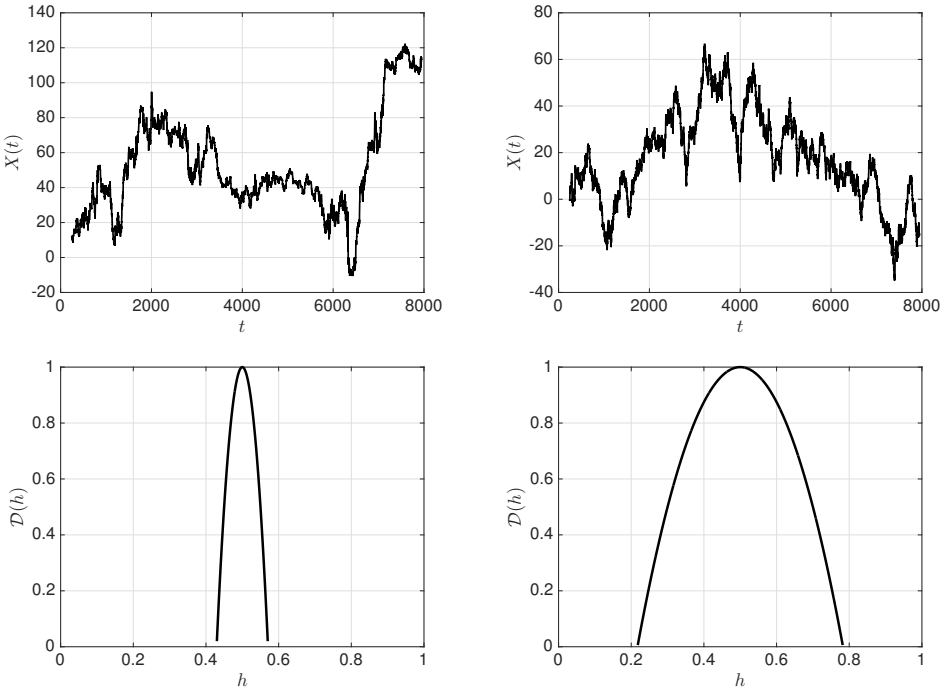



FIGURE 1.2 – **Exemples de processus multifractals unidimensionnels**. Haut : Deux réalisations de marches aléatoires multifractales. Bas : spectre multifractal correspondant. Pour $c_1 = 0.5$, deux valeurs de $c_2$ sont examinées : à gauche, $c_2 = -0.0025$ et à droite $c_2 = -0.04$.

Alors que les processus monofractals sont caractérisés par un seul exposant $H$, les exposants de Hölder $h$ pour des processus *multifractals* peuvent varier (de manière continue ou discontinue) d'un point à un autre, i.e., le support du spectre multifractal correspondant n'est plus réduit à un singleton. En outre, les processus satisfaisant (1.5) forment une famille de processus multifractals. En particulier, nous nous intéresserons dans le Chapitre 3 aux marches aléatoires multifractales, qui ont la particularité d'avoir un spectre multifractal parabolique.

**Exemple 1.2** (Spectre multifractal d'une marche aléatoire multifractale (MRW) [179, 3])**.** *Soit une marche multifractale aléatoire de dimension d caractérisée par les paramètres $\kappa$ et $\lambda$. Alors son spectre multifractal est une parabole d'équation :*

$$(\forall h \geq 0) \quad \mathcal{D}(h) = d + \frac{(h - c_1)^2}{2c_2}, \tag{1.10}$$

*où $c_1 = \kappa + \lambda^2/2$ et $c_2 = -\lambda^2$. Deux exemples de MRW unidimensionnels sont illustrés par la Figure 1.2 pour $(c_1, c_2) = (0.5, -0.0025)$ (gauche) et $(c_1, c_2) = (0.5, -0.04)$ (droite).*

## 1.3 Formalisme multifractal

En pratique, il est difficile voire impossible d'estimer le spectre $\mathcal{D}$ par la mesure de $h$ (cf., e.g., [134] pour une revue de diverses méthodes). Dans cette section, nous présentons une méthode d'estimation efficace basée sur le formalisme multifractal.

### 1.3.1 Fonction de partition et fonction d'échelle

Le formalisme multifractal repose sur l'utilisation de coefficients multirésolution de $X$, notés $T(a,t)$, dépendant à la fois de l'échelle $a$ et de la position $t$. Ces coefficients pouvant être par exemple les accroissements de $X$, i.e., $T(a,t) = X(a+t) - X(t)$, les coefficients d'ondelettes [139] voire les coefficients d'ondelettes dominants [114, 213]. Une propriété nécessaire, est que ces coefficients doivent reproduire localement l'exposant de Hölder dans la limite des petites échelles [112, 115, 139], i.e.,

$$|T(a,t)| \underset{a\to 0}{\sim} a^{h(t)}. \tag{1.11}$$

Plutôt que mesurer ponctuellement $h$ à chaque position $t$, le formalisme multifractal a recours à une approche statistique reposant sur l'utilisation de la fonction de partition suivante décrivant les propriétés statistiques de $X$.

**Définition 1.7** (Fonction de partition)**.** *Soit $T(a,t)$ des coefficients multirésolution dépendant de l'échelle $a > 0$ et de la position $t \in \mathbb{R}^d$. On appelle fonction de partition à l'ordre $q \in \mathbb{Z}^*$, la quantité définie par*

$$S_q(a) = \mathbb{E}|T(a,\cdot)|^q = \mathbb{E}\mathrm{e}^{q \ln T(a,\cdot)} \tag{1.12}$$

L'invariance d'échelle d'un processus se traduit par le comportement en loi de puissance suivant de sa fonction de partition :

**Définition 1.8** (Fonction d'échelle)**.** *Si $X$ est invariant d'échelle, alors sa fonction de partition possède le comportement en loi de puissance suivant*

$$S_q(a) \underset{a\to 0}{\sim} a^{\zeta(q)} \tag{1.13}$$

*où la fonction $q \mapsto \zeta(q)$ est appelée* fonction d'échelle. *En particulier, pour une certaine gamme $a \in [a_{\min}, a_{\max}]$ et $q \in [q_{\min}, q_{\max}]$, on peut considérer en première approximation qu'il existe $F$ tel que*

$$S_q(a) = F(q)a^{\zeta(q)}. \tag{1.14}$$

La forme concave de la fonction d'échelle $\zeta$, observée expérimentalement en [95] peut être expliquée à partir des fluctuations de la régularité locale :

**Proposition 1.1** (Lien entre fonction d'échelle et spectre multifractal)**.** *Soit un processus caractérisé par le spectre multifractal $\mathcal{D}$ et la fonction d'échelle $\zeta$. Alors $\mathcal{D}$ et $\zeta$ sont liés par la transformation de Legendre suivante :*

$$\zeta(q) = \inf_h \left(qh - \mathcal{D}(h) + d\right). \tag{1.15}$$

*Donc $\zeta$ est concave (voir Définition 2.3 et Remarque 2.6). Par conséquent, à partir de $\zeta$, le spectre multifractal peut être obtenu par*

$$\mathcal{D}(h) \leq \inf_{q\neq 0} \left(qh - \zeta(q) + d\right), \tag{1.16}$$

*où l'égalité est obtenue dans* (1.16) *pour certains processus.*

**Remarque 1.2.** *Le résultat de la Proposition 1.1 peut être compris par l'heuristique suivante. Pour tout $t \in \mathbb{R}^d$, il y a environ $|a|^{-\mathcal{D}(h_0)}$ boîtes de volume $|a|^d$ pour lesquelles $h(t) \equiv h_0$ et par conséquent $|T(a,t)| \sim |a|^{h_0}$. Dès lors,*

$$S_q(a) \approx \int_{\mathbb{R}^d} |T(a,t)|^q \mathrm{d}t \sim \int_h |a|^{qh}|a|^{-\mathcal{D}(h)}|a|^d \mathrm{d}h = \int_h |a|^{qh-\mathcal{D}(h)+d}\mathrm{d}h \tag{1.17}$$

*Lorsque $a \to 0$, la contribution dominante dans* (1.17) *est donnée par*

$$\zeta(q) = \inf_h \left(qh - \mathcal{D}(h) + d\right). \tag{1.18}$$

Selon la nature mono/multi-fractal d'un processus, nous pouvons distinguer deux catégories de fonctions d'échelles : pour les processus monofractals, $\zeta$ varie linéairement avec $q$ alors que pour les processus multifractals, $\zeta$ est une fonction non linéaire concave.

En pratique, on utilise des relations comme (1.16) permettant de relier des quantités, par exemple la fonction d'échelle $\zeta$, au spectre multifracal $\mathcal{D}$. Les relations de ce type sont appelées *formalisme multifractal*. Dans la suite, nous détaillons comment appliquer ce formalisme en pratique à partir des coefficients d'ondelettes dominants.

### 1.3.2 Processus monofractal et coefficients d'ondelettes

Pour les processus monofractals où $\mathcal{D}$ est de la forme (1.9), on peut montrer par l'intermédiaire de la Proposition 1.1 que

$$q \mapsto \zeta(q) = qH \tag{1.19}$$

est une fonction linéaire. Dans ce cas, et quel que soit $q \in \mathbb{Z}^*$,

$$H = \lim_{a \to 0} \frac{\ln S_q(a)}{q \ln a}, \tag{1.20}$$

ce qui suggère que $H$ peut être estimé par régression linéaire de $\ln S_q(a)$ vs. $q \ln a$. En particulier, une méthode réputée efficace repose sur l'utilisation des coefficients de la transformée en ondelettes discrète comme quantité multi-résolution [82, 5, 115].

**Définition 1.9** (Transformée en ondelettes discrète [139])**.** *Soit $\psi_0$ une ondelette mère, à savoir, $\psi_0 \in L^2(\mathbb{R})$ et*

$$(\forall l \in \{0, \ldots, N_\psi - 1\}) \quad \int_{\mathbb{R}} t^l \psi_0(t) dt \equiv 0 \tag{1.21}$$

*où $N_\psi$ désigne le nombre de moments nuls de $\psi_0$. Notons également*

$$\{\psi_{j,k}(t) = 2^{-j/2} \psi_0(2^{-j}t - k)\}_{j \in \mathbb{Z}, k \in \mathbb{N}} \tag{1.22}$$

*la collection des versions dilatées et translatées de $\psi_0$ qui forment une base orthogonale de $L^2(\mathbb{R})$. Le coefficient d'ondelette de $\{X(t)\}_{t \in \mathbb{R}}$ à l'échelle $a = 2^j$ et à la position $t = 2^j k$ est alors représenté de manière non redondante par la quantité $D(j,k)$, où*

$$D(j,k) = \langle \psi_{j,k}, X \rangle = \int_{\mathbb{R}} 2^{-j/2} \psi_0(2^{-j}t - k) X(t) \mathrm{d}t. \tag{1.23}$$

*Le lecteur est invité à se référer à [12] pour une introduction aux ondelettes bidimensionnelles.*

En guise d'exemple, nous proposons de détailler comment estimer $H$ via (1.20) à partir des coefficients d'ondelettes pour les fBm. Le même raisonnement sera généralisé au cadre multivarié dans le Chapitre 5.

**Exemple 1.3** (Mouvement brownien fractionnaire [82])**.** *Soit un fBm $\{X(t)\}_{t \in \mathbb{R}}$ de paramètres $\sigma^2 > 0$ et $0 < H < 1$. Etant donné que $X$ est entièrement caractérisé par ses propriétés statistiques d'ordre 2, nous considérons sa fonction de partition à l'ordre $q = 2$*

*exprimée à partir des coefficients d'ondelettes, également appelée* spectre en ondelettes, *qui s'écrit*

$$S_{q=2}(2^j) \equiv \mathbb{E}D(j,\cdot)D(j,\cdot)^* = \sigma^2 \eta_H 2^{2j(H+1/2)}, \tag{1.24}$$

*où*

$$\eta_H = -\frac{1}{2}\int_{\mathbb{R}} |u|^{2H} \mathrm{d}u \int_{\mathbb{R}} \psi_0(v)\psi_0(v-u)^* \mathrm{d}v. \tag{1.25}$$

*Dès lors, $H$ (ou plutôt $2H+1$) peut être estimé comme le coefficient de régression linéaire de $\ln S_{q=2}(2^j)$ vs. $\ln 2^j$, et $\sigma^2$ (ou plutôt $\sigma^2\eta_H$) peut ensuite être déduit via l'ordonnée à l'origine.*

*Démonstration.* A partir de (1.3),

$$\mathbb{E}D(j,\cdot)D(j,\cdot)^* = \int_{\mathbb{R}^2} \psi_{j,k}(t)\psi_{j,k}(s)\mathbb{E}X(t)X(s)^* \mathrm{d}t\mathrm{d}s \tag{1.26}$$

$$= -\frac{\sigma^2}{2}\int_{\mathbb{R}^2} \psi_{j,k}(t)\psi_{j,k}(s)^* |t-s|^{2H} \mathrm{d}t\mathrm{d}s \tag{1.27}$$

$$= -\frac{\sigma^2}{2}\int_{\mathbb{R}^2} \psi_{j,k}(v)\psi_{j,k}(v-u)^* |u|^{2H} \mathrm{d}v\mathrm{d}u \tag{1.28}$$

D'après la définition des ondelettes $\psi_{j,k}(v) = 2^{-j/2}\psi_0(v2^{-j}-k)$, il s'en suit donc que

$$\mathbb{E}D(j,\cdot)D(j,\cdot)^* = -\frac{\sigma^2}{2}2^{-j}\int_{\mathbb{R}^2} \psi_0(v2^{-j}-k)\psi_0((v-u)2^{-j}-k)^* |u|^{2H} \mathrm{d}v\mathrm{d}u \tag{1.29}$$

$$= -\frac{\sigma^2}{2}2^{-j}\int_{\mathbb{R}^2} \psi_0(v)\psi_0(v-u)^* |2^j u|^{2H} \left(2^j \mathrm{d}v\right)\left(2^j \mathrm{d}u\right) \tag{1.30}$$

$$= \sigma^2 2^{2j(H+1/2)} \left(-\frac{1}{2}\int_{\mathbb{R}} |u|^{2H}\mathrm{d}u \int_{\mathbb{R}} \psi_0(v)\psi_0(v-u)^*\mathrm{d}v\right) \tag{1.31}$$

□

**Remarque 1.3.** *En pratique, pour l'analyse d'invariance d'échelle il est parfois d'usage d'utiliser des coefficients d'ondelettes re-normalisés par la norme $\ell_1$, i.e.,*

$$d(j,k) = 2^{-j/2}D(j,k) \tag{1.32}$$

*afin de faire disparaître l'exposant* $1/2$ *dans* (1.24) *(cf. [16]).*

### 1.3.3 Processus multifractal et coefficients dominants

Pour les processus multifractals, le support de $\mathcal{D}$ n'est plus réduit à un singleton et son estimation est alors plus délicate. Plusieurs conséquences découlent de la transformée de Legendre reliant $\zeta$ à $\mathcal{D}$ (voir Proposition 1.1). Premièrement, la dérivée de $\zeta$ en $q \sim 0$ donne la position du maxima de $\mathcal{D}$, tel qu'illustré par la Figure 1.3 en pointillés. Deuxièmement, les dérivées de $\zeta$ pour $q > 0$ (resp. $q < 0$) permettent de mesurer la partie croissante (resp. décroissante) de $\mathcal{D}$. Afin de caractériser entièrement $\mathcal{D}$, il est alors nécessaire d'examiner $\zeta$, et par conséquent la fonction de structure $S_q$ (voir (1.13)), pour des moments $q$ négatifs et positifs.

Cependant, puisque la décomposition en ondelettes de $X$ possède nécessairement un grand nombre de coefficients $D(j,k)$ proches de 0, la fonction de partition $S_q(2^j) = \mathbb{E}|D(j,\cdot)|^q$ est susceptible de diverger pour $q < 0$. C'est, entre autres, afin de pallier cette limitation, que les coefficients d'ondelettes dominants ont été introduits [114, 128, 213, 115]. Dans la suite, afin d'étudier les processus multifractals, nous utiliserons systématiquement comme quantité multi-résolution, les coefficients d'ondelettes dominants définis comme suit :

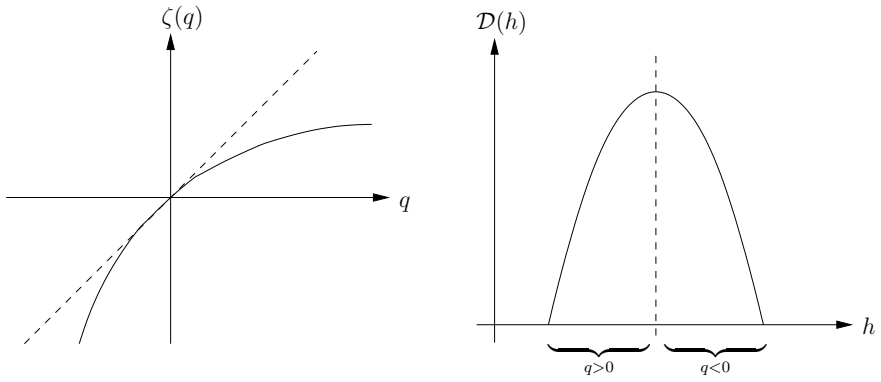



FIGURE 1.3 – **Illustration de la correspondance entre $\zeta$ et $\mathcal{D}$.** La dérivée de $\zeta$ en $q = 0$ donne la position du maxima de $\mathcal{D}$ (pointillés). La partie croissante de $\mathcal{D}$ est obtenue à partir de la connaissance de $\zeta$ pour $q > 0$, alors que sa partie décroissante est liée aux valeurs pour $q < 0$.

**Définition 1.10** (Coefficients d'ondelettes dominants [114])**.** *Le coefficient d'ondelettes dominant autour de la position $k$ et à l'échelle $j$, noté $L(j,k)$, est défini comme le supremum local autour d'un voisinage spatial, de tous les coefficients d'ondelettes à travers les échelles plus fines $2^{j'} \leq 2^j$. Formellement,*

$$L(j,k) = \sup_{\omega_{j',k'} \subset \Omega_{j,k}} |D(j',k')|, \tag{1.33}$$

*où $\omega_{j,k} = [k2^j, (k+1)2^j)$ et $\Omega_{j,k} = \cup_{p\in\{-1,0,1\}} \omega_{j,k+p}$. Une illustration est disponible Figure 1.4 où $L(j,k)$ est indiqué par une croix noire et où le voisinage $\Omega_{j,k}$ est affiché en vert.*

A présent, si l'on considère le développement polynomial de $\zeta$ en 0, i.e.,

**Définition 1.11** (Coefficients $(c_p)_p$ [13])**.**

$$\zeta(q) \underset{q\sim 0}{=} \sum_{p=1}^{\infty} c_p \frac{q^p}{p!}, \tag{1.34}$$

*où $c_1$ correspond à la partie linéaire de $\zeta$, où $c_2$ quantifie la première déviation de la linéarité, et où les coefficients $(c_p)_{p\geq 3}$ correspondent à des ordres de nonlinéarité plus élevés.*

alors le spectre multifractal peut être approché à partir des coefficients $(c_p)_{p\in\mathbb{N}^*}$. En particulier, dans la suite nous considérons souvent l'approximation parabolique du spectre multifractal suivante :

**Proposition 1.2** (Développements du spectre multifractal en log-cumulants)**.** *Soient $(c_p)_{p\in\mathbb{N}^*}$ les coefficients définis en* (1.34)*. Alors en première approximation, nous pouvons considérer que*

$$\mathcal{D}(h) \approx d + \frac{c_2}{2!}\left(\frac{h - c_1}{c_2}\right)^2 \tag{1.35}$$

*où $c_1$ peut être interprété comme la position du maximum de $\mathcal{D}$ et $c_2$ sa largeur.*

Afin d'estimer ces coefficients, nous considèrons la quantité $\ln S_q(2^j)$ similaire à l'énergie libre en physique statistique. Tout comme les cumulants d'ordre 1 et 2 de l'énergie libre donnent l'énergie moyenne et la chaleur, nous considérons le développement en cumulants suivant

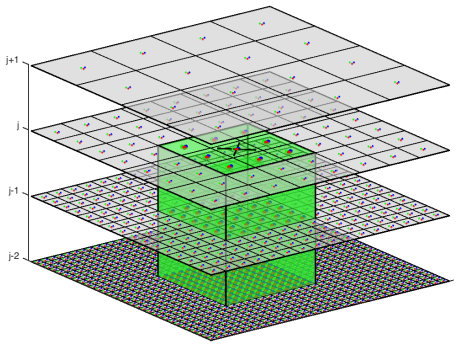



FIGURE 1.4 – **Coefficients d'ondelettes dominants.** Le coefficient d'ondelette dominant (croix noire) est défini comme le suprémum local de tous les coefficients d'ondelettes présents dans un voisinage spatial et à travers toutes les échelles les plus fines (vert). Figure empruntée à [212].

**Définition 1.12** (Coefficients $(C_p)_{p\in\mathbb{N}^*}$)**.**

$$\ln\left(\mathbb{E}e^{q\ln L(j,\cdot)}\right) = \sum_{p=1}^{+\infty} C_p(2^j)\frac{q^p}{p!}, \tag{1.36}$$

*où $C_p(2^j)$ est le cumulant d'ordre $p$ de $\ln L(j,\cdot)$.*

Les coefficients $(C_p)_{p\in\mathbb{N}^*}$ sont directement liés aux coefficients $(c_p)_{p\in\mathbb{N}^*}$ permettant de caractériser le spectre multifractal comme suit :

**Proposition 1.3** (Relation entre $C_p$ et $c_p$ [213].)**.** *Soient $C_p(2^j)$ le cumulant d'ordre $p$ de $\ln L(j,\cdot)$ et $c_p$ le coefficient d'expansion polynomial d'ordre $p$ de $q \mapsto \zeta(q)$ en 0 (cf. (1.34)). Alors*

$$C_p(2^j) = c_{0,p} + \ln(2^j)c_p. \tag{1.37}$$

*Démonstration.* A partir de la Définition 1.8, nous obtenons

$$\ln\left(\mathbb{E}|L(j,\cdot)|^q\right) = \ln F(q) + \zeta(q)\ln(2^j). \tag{1.38}$$

Dès lors, si l'on considère d'un côté le développement en cumulant $(C_p)$ de $\ln L(j,\cdot)$ dérivé en (1.36), et de l'autre les développements de Taylor de $q \to 0$ de $\zeta$ (cf. (1.34)) et $\ln F$ :

$$\ln F(q) = \sum_{p=1}^{\infty} c_{0,p}\frac{q^p}{p!}, \tag{1.39}$$

alors nous pouvons alors montrer que

$$C_p(2^j) = c_{0,p} + \ln(2^j)c_p. \tag{1.40}$$

□

**Remarque 1.4.** *Les coefficients $(c_p)_{p\geq 1}$ sont appelés log-cumulants de $\ln L(j,\cdot)$.*

A partir de la relation (1.37), il apparaît que les log-cumulants peuvent être obtenus par régression linéaire de $C_p(2^j)$ vs. $\ln(2^j)$ [213].

## 1.4 Questions ouvertes

La formulation actuelle du formalisme multifractal repose sur l'hypothèse que les propriétés multifractales du processus d'intérêt $X$ sont homogènes, c'est-à-dire que $\mathcal{D}$ décrit les propriétés de $X(t)$ pour tout $t \in \mathbb{R}^d$. Mais dans de nombreuses situations, il se peut que celles-ci puissent varier dans le temps ou spatialement. On retrouve par exemple cette situation lors de l'analyse de photographies satellitaires comprenant à la fois des textures de nuage et de neige [181]. Bien qu'il soit difficile de les distinguer visuellement, ces deux textures sont connues pour être invariantes d'échelle et caractérisées par un spectre multifractal différent. Dans ce cas, la fonction de partition $S_q$ ne doit pas être calculée sur la totalité de l'enregistrement disponible mais sur chacune des régions où les propriétés multifractales puissent être considérées homogènes. Cependant, ces régions sont *a priori* inconnues et doivent être estimées conjointement avec $\mathcal{D}$, ce qui complique la procédure d'estimation. Cette question sera abordée au Chapitre 3 à l'aide des approches variationnelles basées sur la variation totale décrites au Chapitre 2.

Dans ce chapitre nous avons présenté le formalisme multifractal pouvant s'appliquer à tout processus univarié, qu'il prenne par exemple la forme d'un signal ou d'une image $X$. Cependant, avec le développement des méthodes d'acquisition et l'accès à un plus grand nombre de capteurs, les processus enregistrés sont de plus en plus multivariés par nature, i.e., $X = (X_1, \ldots, X_M)^\top$. Dès lors, analyser leurs propriétés d'invariance d'échelle conjointement permettrait de mieux comprendre les mécanismes sous-jacents à leur élaboration. Un premier pas dans cette direction consiste à étendre la définition de l'autosimilarité (voir Définition 1.1) aux processus multivariés. Dans ce cas, la fonction de partition est un mélange de plusieurs lois de puissance que les techniques d'estimation univariées ne permettent pas de démélanger. La conception d'un paradigme d'estimation multivarié complet reste un problème ouvert dans la littérature. C'est dans ce contexte que le Chapitre 5 apporte sa contribution en proposant, à notre connaissance, le premier procédé d'estimation complet de processus bivariés autosimilaires. La méthode sera appliquée sur des données réelles issues du trafic internet dans le Chapitre 6.

# Chapitre 2

# Variation totale

## Sommaire


**2.1 Introduction aux problèmes $\ell_0$-TV et $\ell_1$-TV . . . . . 29**
2.1.1 Problème non convexe $\ell_0$-TV . . . . . 29
2.1.2 Problème convexe $\ell_1$-TV . . . . . 30
2.1.3 Schémas itératifs d'optimisation convexe . . . . . 32
**2.2 Résolution du problème $\ell_1$-TV . . . . . 34**
2.2.1 Passage au dual . . . . . 35
2.2.2 Augmentation de la dimensionnalité . . . . . 37
2.2.3 Lagrangien augmenté . . . . . 38
2.2.4 Algorithme primal-dual . . . . . 39
2.2.5 Comparaison des algorithmes . . . . . 40
2.2.5.1 Choix des paramètres des algorithmes . . . . . 40
2.2.5.2 Coût de calcul . . . . . 42
**2.3 Evaluation des performances d'estimation . . . . . 42**
2.3.1 Rappels sur l'indice de Jaccard . . . . . 44
2.3.2 Application à la détection de discontinuité . . . . . 45
2.3.3 Choix du filtre passe-bas gaussien . . . . . 46
**2.4 Extension bidimensionnelle . . . . . 47**
2.4.1 Problème non convexe $\ell_0$-TV 2D . . . . . 48
2.4.2 Problème convexe $\ell_1$-TV 2D . . . . . 49
2.4.3 Problème convexe $\ell_0$-TV 2D à $Q$ niveaux . . . . . 49
**2.5 Questions ouvertes . . . . . 51**


**Quelques grandes classes de méthodes pour détecter les discontinuités.** Une première classe de méthodes linéaires permettant de détecter les discontinuités des propriétés statistiques consiste à filtrer les données de façon à éliminer le bruit présent dans les hautes fréquences en appliquant un filtre passe bas. Toutefois, la solution débruitée présente des oscillations au voisinage des discontinuités (phénomène de Gibbs) [116]. Des filtres non linéaires allant du filtre médian glissant [199, 101, 218] au seuillage des coefficients d'ondelettes [140, 46, 70] ont alors été développés. Cependant, la première classe de méthodes a tendance à lisser le signal, masquant ainsi la détection précise de la position des discontinuités. Quant à la seconde, reposant sur des opérations de seuillage dur (norme $\ell_0$) ou doux (norme $\ell_1$) des coefficients d'ondelettes, elle implique une perte des détails [46, 77].

En parallèle, une grande partie de la littérature dédiée aux problèmes de détection de discontinuités s'est développée en inférence bayésienne et tests d'hypothèses. Les méthodes fréquentistes, telles que l'approche Cusum, reposent sur un test d'hypothèse à partir du rapport de vraisemblance [210, 163, 106, 61, 135]. Les méthodes bayésiennes permettent d'incorporer des *a priori* supplémentaires comme par exemple la distribution des discontinuités. La première méthode bayésienne pour détecter des discontinuités a été développée en [96] pour le contrôle de qualité des machines. La maintenance des appareils est décidée quand la distribution postérieure d'être dans un état défectueux dépasse un certain seuil [186, 187, 170, 171]. Notons que ces approches permettent une détection des discontinuités en ligne. Lorsque la distribution postérieure ne possède pas d'expression explicite, des algorithmes de Monte Carlo par Chaînes de Markov (MCMC) doivent être envisagés, pouvant conduire, pour certains problèmes, à des temps de calcul élevés [68, 151].

**Approches variationnelles.** Le problème de détection de discontinuités peut également être considéré du point de vue des approches variationnelles où la solution débruitée minimise un critère composé d'un terme d'attache aux données et d'un terme de régularisation. Dans ce second chapitre, nous nous intéressons à l'utilisation de la variation totale (TV) comme terme de régularisation. Pénaliser la variation totale par une norme favorisant la parcimonie, e.g. $\ell_1$ (dit $\ell_1$-TV ou seulement TV) ou $\ell_0$ (noté $\ell_0$-TV par abus de langage), permet à la solution restaurée de préserver un nombre restreint de discontinuités. Le terme variation totale a été introduit en débruitage d'image dans [182] où la variation totale $\ell_1$-TV de l'image analysée est minimisée sous contrainte que le bruit soit de moyenne nulle et de variance connue *a priori*. Les contraintes sont introduites à l'aide de multiplicateurs de Lagrange et l'algorithme proposé permet de résoudre les équations de stationnarité d'Euler Lagrange à l'aide d'une méthode de descente de gradient. D'un côté, les récents travaux autour de l'opérateur proximal permettent de résoudre le problème $\ell_1$-TV plus rapidemment et avec davantage de flexibilité [38, 209]. De l'autre côté, le problème $\ell_0$-TV, pourtant non convexe, peut être résolu de manière exacte et efficacement par des algorithmes de programmation dynamique [217, 216, 192, 93, 193], mais son extension à différents espaces ou à des critères génériques impliquants la présence d'un terme $\ell_0$-TV n'est pas systématique.

Les premières contributions de la $\ell_1$-TV dans la communauté du traitement du signal ont été formulées dans le cadre de la théorie « taut-string » [58, 76]. Les méthodes « taut-string », ainsi que les méthodes « run », reposent sur l'encadrement du signal à analyser par des bornes locales. Le signal débruité est ensuite défini à partir des positions et des valeurs pour lesquelles les bornes sont égales. Il a plus tard été montré en [141] que la solution « taut-string » est également la solution du problème de régularisation $\ell_1$-TV. Des travaux récents continuent d'être développés pour calculer directement la solution [52], étendue à différents espaces non nécessairement vectoriels [194] et pour divers problèmes de régression TV (régression de Poisson, TV pondérée) [76, 19, 121]. Si les méthodes en ligne Cusum, « run » et « taut-string » possèdent l'avantage de pouvoir calculer la solution à la volée, leur extension au cadre vectoriel ou à des problèmes impliquant des contraintes sur la dynamique, s'avère cependant beaucoup plus compliquée.

Dans la Section 2.1, nous formaliserons les problèmes $\ell_0$-TV et $\ell_1$-TV pour les signaux, et nous rappellerons diverses notions d'optimisation convexe non lisse (i.e., non différentiable) qui seront utiles tout au long de ce manuscrit. Nous en profiterons pour donner plusieurs algorithmes proximaux permettant de résoudre le problème $\ell_1$-TV et les comparererons dans la Section 2.2. Puis, nous définirons dans la Section 2.3 un critère permettant d'évaluer la qualité de l'estimation des discontinuités de signaux constants par morceaux. Pour finir, nous discuterons dans la Section 2.4 de l'extension des problèmes $\ell_0$-TV et $\ell_1$-TV aux images.

## 2.1 Introduction aux problèmes $\ell_0$-TV et $\ell_1$-TV

### 2.1.1 Problème non convexe $\ell_0$-TV

Considérons une observation $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{b} \in \mathbb{R}^N$ d'un signal constant par morceaux $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ corrompu par un bruit additif $\boldsymbol{b} \in \mathbb{R}^N$. Dans la suite du manuscrit, sauf mention contraire, nous supposerons que $\boldsymbol{b}$ est un bruit blanc gaussien centré de variance $\sigma^2$ inconnue, i.e., $\boldsymbol{b} \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_{\boldsymbol{N}})$.

Afin d'obtenir une solution débruitée constante par morceaux, nous pouvons maximiser la vraisemblance $p(\boldsymbol{y}|\boldsymbol{x}) = (2\pi\sigma^2)^{-N/2} \exp\left(-\|\boldsymbol{y} - \boldsymbol{x}\|^2/2\sigma^2\right)$ ou de manière équivalente minimiser l'anti-log-vraisemblance, sous contrainte que le nombre de discontinuités soit limité, i.e.,

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \quad \text{sujet à} \quad \sum_{n=1}^{N-1} \delta(x_{n+1} - x_n) \leq K, \tag{2.1}$$

où $\delta$ est la fonction de Dirac égale à 0 si son argument est nul et 1 sinon, et $K$ est le nombre maximal de discontinuités autorisé. Une méthode reposant sur le critère de Schwarz a été proposée pour estimer $K$ efficacement [131, 219]. Le problème (2.1) peut être réécrit de manière équivalente

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 \quad \text{sujet à} \quad \|L\boldsymbol{x}\|_0 \leq K, \tag{2.2}$$

où la pseudo-norme $\ell_0$ compte le nombre d'éléments non nuls et où $L$ est l'opérateur de premières différences, défini par

$$(\forall n \in \{1, \ldots, N-1\}), \quad (L\boldsymbol{x})_n = x_{n+1} - x_n. \tag{2.3}$$

Le problème contraint (2.2) peut également être reformulé sous forme régularisée, ce qui mène au problème suivant :

**Problème 2.1** (Problème $\ell_0$-TV)**.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\chi \geq 0$. On appelle problème $\ell_0$-TV le problème consistant à trouver*

$$\widehat{\boldsymbol{x}}_\chi = \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \chi\|L\boldsymbol{x}\|_0 \tag{2.4}$$

*où $L$ est l'opérateur linéaire de premières différences défini en* (2.3)*.*

**Remarque 2.1.** *Le Problème 2.1, également appelé problème de Pott univarié, est connu sous le nom de problème de Mumford-Shah constant par morceaux dans le cadre continu et bidimensionnel.*

L'hyperparamètre $\chi$ est appelé paramètre de régularisation et permet de relativiser les contributions du terme d'attache aux données et du terme de régularisation. En divisant le second membre de (2.4) par $\sigma^2$, nous remarquons qu'il varie naturellement comme $\chi = \sigma^2\tilde{\chi}$. Pour $\chi \to 0$, seul le terme d'attache aux données contribue significativement dans le critère et par conséquent $\widehat{\boldsymbol{x}}_\chi \approx \boldsymbol{y}$ est composé d'une multitude de segments. A l'inverse, pour $\chi \to +\infty$, quasiment seul le terme de régularisation contribue et $\widehat{\boldsymbol{x}}_\chi$ est composé d'un seul segment. Par conséquent, le choix de $\chi$ impacte fortement la qualité de l'estimateur $\widehat{\boldsymbol{x}}_\chi$. Dans ce chapitre, nous considérons que $\chi$ est connu. Son estimation sera discutée dans le Chapitre 8.

Bien que le Problème 2.1 ($\ell_0$-TV) soit non convexe, il peut être résolu de manière exacte à l'aide de méthodes de programmation dynamique [217, 216, 192, 93, 193]. Le principe de l'algorithme réside sur le fait que la solution $\widehat{\boldsymbol{x}}_\chi^{(n)} \in \mathbb{R}^n$ sur $(y_1, \ldots, y_n)$ peut être obtenue en temps polynomial $O(n^2)$ à condition que les solutions $(\widehat{\boldsymbol{x}}_\chi^{(1)}, \widehat{\boldsymbol{x}}_\chi^{(2)}, \ldots, \widehat{\boldsymbol{x}}_\chi^{(n-1)})$ sur chaque observation partielle $\{(y_1), (y_1, y_2), \ldots, (y_1, \ldots, y_{n-1})\}$ soient connues. Afin de calculer $\widehat{\boldsymbol{x}}_\chi^{(n)}$, $n$ solutions candidates $\{\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(n)}\}$ définies par

$$(\forall r \in \{1, \ldots, n\}) \quad \boldsymbol{x}^{(r)} = (\underbrace{\widehat{\boldsymbol{x}}_\chi^{(r-1)\top}}_{\text{taille } r-1}, \underbrace{\mu_{[r,n]} \mathbb{1}_{n-r+1}^\top}_{\text{taille } n-r+1})^\top, \tag{2.5}$$

où $\mu_{[r,n]}$ est la moyenne des observations $(y_r, \ldots, y_n)$, sont proposées. La solution retenue est alors celle possédant le critère (2.4) le plus faible. Une implémentation efficace reposant sur la technique d'accélération proposée dans [192, Algorithme 2] est disponible à l'adresse `pottslab.de`.

Le problème peut également est résolu de manière exacte lorsque le terme d'attache aux données $\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$ est remplacé par $\|\boldsymbol{y} - \boldsymbol{x}\|_1$ (i.e., bruit Laplacien $\boldsymbol{b}$) [194] ou par la divergence de Kullback-Leibler entre $\boldsymbol{x}$ et $\boldsymbol{y}$ (i.e., bruit Poissonien $\boldsymbol{b}$) [75]. Toutefois, l'extension du Problème 2.1 ($\ell_0$-TV) au cadre bidimensionnel, discutée plus en détails dans la Section 2.4, reste NP difficile. D'un côté, diverses stratégies ont été proposées pour approcher le Problème 2.1 ($\ell_0$-TV) par un problème plus simple où par exemple le nombre de valeurs $Q$ prises par $\overline{\boldsymbol{x}}$ est supposé connu *a priori*. D'un autre côté, une partie de la littérature s'est intéressée à la relaxation convexe du Problème 2.1 en remplaçant la pseudo-norme $\ell_0$ par la norme convexe $\ell_1$ [36, 18, 209, 52]. Dans la suite, nous nous intéresserons à cette seconde classe de méthodes.

### 2.1.2 Problème convexe $\ell_1$-TV

Une relaxation convexe du Problème 2.1 s'écrit

**Problème 2.2** (Problème $\ell_1$-TV)**.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\lambda \geq 0$. On appelle problème $\ell_1$-TV le problème consistant à trouver*

$$\widehat{\boldsymbol{x}}_\lambda = \arg\min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \lambda \|L\boldsymbol{x}\|_1 \tag{2.6}$$

*où $L$ est l'opérateur linéaire de premières différences défini en* (2.3) *et où $\|L\boldsymbol{x}\|_1$ est la variation totale de $\boldsymbol{x}$ [182].*

**Remarque 2.2.** *Si les amplitudes des discontinuités de $\boldsymbol{x}$ sont de même ordre de grandeur $\delta_x$, alors les Problèmes 2.1 et 2.2 sont équivalents pour $\chi \sim \delta_x \lambda$.*

L'emploi de la norme $\ell_1$ permet de sélectionner un nombre restreint de discontinuités, donnant par conséquent à $\widehat{\boldsymbol{x}}_\lambda$ un aspect constant par morceaux. En contrepartie, la norme $\ell_1$ introduit un biais sur l'estimation des valeurs de $\boldsymbol{x}$ sur chaque segment. En effet, considérons la réalisation $\boldsymbol{y} \in \mathbb{R}^N$ ($N = 10^3$) illustrée sur la Figure 2.1 (haut) en gris. La solution $\widehat{\boldsymbol{x}}_{\lambda=1}$ (rouge) du Problème 2.2 est composée de nombreux segments alors que $\widehat{\boldsymbol{x}}_{\lambda=10}$ (jaune) et $\widehat{\boldsymbol{x}}_{\lambda=100}$ (violet) comportent moins de segments mais leur estimation des valeurs de $\boldsymbol{x}$ (noir) est davantage biaisée.

La Figure 2.1 (bas) illustre ce compromis en fonction de $\lambda$ avec à gauche le nombre de discontinuités, i.e. $\|L\widehat{\boldsymbol{x}}_\lambda\|_0$, et à droite l'erreur quadratique relative $\mathrm{MSE}[\widehat{\boldsymbol{x}}_\lambda, \overline{\boldsymbol{x}}] = \|\widehat{\boldsymbol{x}}_\lambda - \overline{\boldsymbol{x}}\|^2 / \|\overline{\boldsymbol{x}}\|^2$. La valeur $\lambda_{\mathrm{MSE}}$ qui permet de minimiser la MSE conduit à une solution surestimant d'un ordre de grandeur le nombre de discontinuités de $\overline{\boldsymbol{x}}$. A l'inverse, la solution obtenue pour $\lambda_{\mathrm{Seg}}$ détient le bon nombre de segments mais possède un biais important.

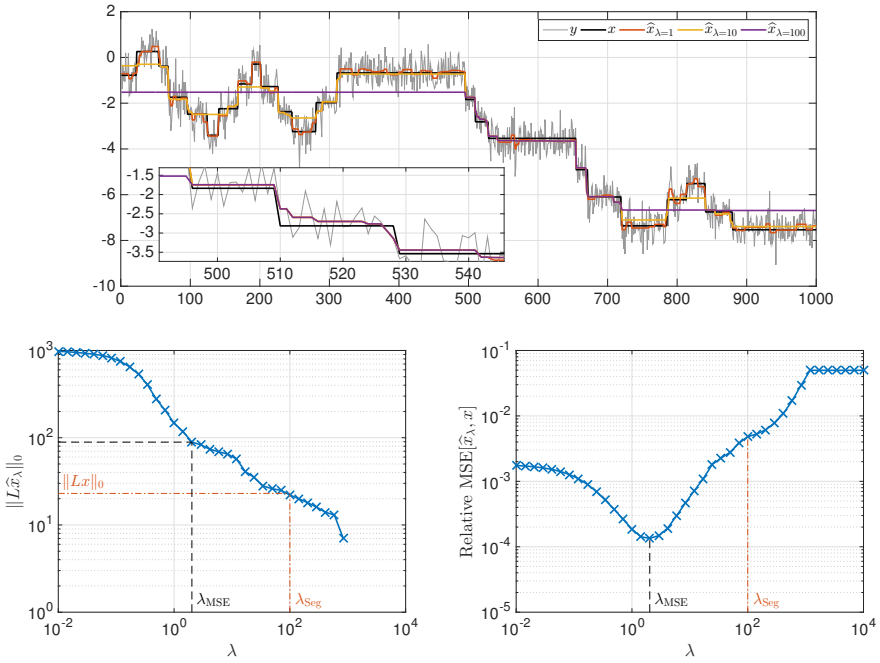



FIGURE 2.1 – **Illustration de $\widehat{\boldsymbol{x}}_\lambda$ solution du Problème 2.2 .** Haut : $\widehat{\boldsymbol{x}}_\lambda$ pour $\lambda = 1$ (rouge), 10 (jaune) et 100 (violet). Bas : compromis entre le biais et le nombre de segment. Bas gauche : le nombre de discontinuités de $\widehat{\boldsymbol{x}}_\lambda$ décroît avec $\lambda$. Bas droite : erreur quadratique relative entre $\widehat{\boldsymbol{x}}_\lambda$ et $\overline{\boldsymbol{x}}$ en fonction de $\lambda$. La valeur $\lambda_{\text{MSE}}$ qui permet de minimiser la MSE conduit à une solution surestimant d'un ordre de grandeur le nombre de discontinuités de $\overline{\boldsymbol{x}}$. A l'inverse, la solution obtenue pour $\lambda_{\text{Seg}}$ détient le bon nombre de segments mais possède un biais important.

**Remarque 2.3** (Effet d'escalier)**.** *L'emploi de la variation totale est susceptible de causer à la solution $\widehat{\boldsymbol{x}}_\lambda$ un effet d'escalier. Cet effet se traduit par $\widehat{\boldsymbol{x}}_\lambda$ présentant plusieurs « marches d'escalier » au niveau des discontinuités de $\overline{\boldsymbol{x}}$. Cet effet est illustré Figure 2.1 (haut, encadré). Par conséquent, cet effet implique qu'il est difficile de quantifier la qualité d'un estimateur $\widehat{\boldsymbol{x}}_\lambda$ de $\overline{\boldsymbol{x}}$ seulement à partir de leur nombre de discontinuités respectif. En effet, la solution $\widehat{\boldsymbol{x}}_{\lambda=100}$, illustrée Fig. 2.1 (haut, violet) possède le même nombre de discontinuités que $\overline{\boldsymbol{x}}$, mais possède une MSE relative élevée. Afin de mesurer les performances d'estimation, nous promouvons l'utilisation de l'indice de Jaccard décrit dans la Section 2.3.*

En pratique, il est souvent préférable d'avoir un estimateur non biaisé $\widehat{\boldsymbol{x}}_\lambda$. Une méthode naïve consiste à ré-estimer $\widehat{\boldsymbol{x}}_\lambda$ *a posteriori* sur chacun de ses segments. Récemment, une méthode a été introduite en [60] pour supprimer une partie du biais de $\widehat{\boldsymbol{x}}_\lambda$ inhérente à l'utilisation de la norme $\ell_1$. Notons également que sous certaines conditions sur $\overline{\boldsymbol{x}}$ et sur $\sigma^2$, la solution du Problème 2.2 garantit de retrouver exactement le support de $L\overline{\boldsymbol{x}}$ [202, 201].

Il est bien connu et largement documenté dans la littérature que l'unique solution du Problème 2.2 peut être obtenue par des algorithmes de points fixes [18, 209]. Cependant, résoudre ce problème nécessite de traiter la non-différentiabilité de la norme $\ell_1$. Une première approche consiste à ajouter un paramètre additionnel pour lisser la norme $\ell_1$ [208]. Une seconde passe par la mise en oeuvre d'algorithmes proximaux dont le principe est rappelé dans la section suivante.

### 2.1.3 Schémas itératifs d'optimisation convexe

Le Problème 2.2 est un problème de minimisation d'une fonction

$$f : \mathbb{R}^N \longrightarrow ]-\infty, +\infty] \tag{2.7}$$

$$\boldsymbol{x} \longmapsto \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 + \|L\boldsymbol{x}\|_1 \tag{2.8}$$

appartenant à la classe $\Gamma_0(\mathbb{R}^N)$ des fonctions convexes, semi-continues inférieurement et propres de $\mathbb{R}^N$ dans $]-\infty, +\infty]$ [1]. Il fait partie du problème générique d'optimisation convexe suivant :

**Problème 2.3.** *Soit une fonction $f \in \Gamma_0(\mathbb{R}^N)$. Le problème générique d'optimisation convexe consiste à trouver*

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{Argmin}}\ f(\boldsymbol{x}), \tag{2.9}$$

*où la notation* Arg *indique que $\widehat{\boldsymbol{x}}$ est un minima global de $f$. Lorsque $f$ est strictement convexe, i.e., unicité du minimiseur, nous adopterons la notation*

$$\widehat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} f(\boldsymbol{x}). \tag{2.10}$$

Lorsque $f$ est différentiable et de gradient $\beta$-Lipschitz, l'algorithme de descente de gradient garantit de trouver le minimum $\widehat{\boldsymbol{x}}$ à condition que le pas $\gamma_k$ à chaque itération $k$ soit choisi tel que $\gamma_k \in ]0, 2/\beta[$. Les itérations sont de la forme

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} = (\mathbb{1} - \gamma_k \nabla f)\,(\boldsymbol{x}^{[k]}) \tag{2.11}$$

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} = \boldsymbol{x}^{[k]} - \gamma_k \nabla f(\boldsymbol{x}^{[k]}). \tag{2.12}$$

Lorsque $f$ n'est pas différentiable en $\boldsymbol{x}^{[k]}$, i.e., lorsque $f$ est non lisse, le gradient $\nabla f(\boldsymbol{x}^{[k]})$ n'est plus défini et $f$ est caractérisé en $\boldsymbol{x}^{[k]}$ par un ensemble de sous-gradients appelé sous-différentielle et défini comme suit

**Définition 2.1** (Sous-différentielle de Moreau [155, 156])**.** *Soit une fonction propre $f : \mathbb{R}^N \to ]-\infty, +\infty]$. La sous-différentielle de $f$, notée $\partial f$, est l'ensemble défini par*

$$(\forall \boldsymbol{u} \in \mathbb{R}^N),\ \partial f(\boldsymbol{u}) = \{\boldsymbol{w} \in \mathbb{R}^N | (\forall \boldsymbol{v} \in \mathbb{R}^N)\ \langle \boldsymbol{v} - \boldsymbol{u}, \boldsymbol{w}\rangle + f(\boldsymbol{u}) \leq f(\boldsymbol{v})\}. \tag{2.13}$$

**Remarque 2.4** ([156])**.** *Si $f$ est différentiable en $\boldsymbol{u}$, alors $\partial f(\boldsymbol{u}) = \{\nabla f(\boldsymbol{u})\}$.*

Dès lors, chaque étape de descente de gradient (2.12) peut être remplacée par une descente de sous-gradient où $\nabla f(\boldsymbol{x}^{[k]})$ est remplacé par un élément $\boldsymbol{w}^{[k]}$ de la sous-différentielle de $f$ en $\boldsymbol{x}^{[k]}$, i.e.,

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} = \boldsymbol{x}^{[k]} - \gamma_k\, \boldsymbol{w}^{[k]} \quad \text{où} \quad \boldsymbol{w}^{[k]} \in \partial f(\boldsymbol{x}^{[k]}) \tag{2.14}$$

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} \in (\mathbb{1} - \gamma_k \partial f)(\boldsymbol{x}^{[k]}). \tag{2.15}$$

Ce type d'algorithme est appelé algorithme de sous-gradient et a été introduit en [188]. Le lecteur pourra se référer à [169, 25, 157, 159] pour des travaux plus récents.

1. Les définitions des fonctions convexes, semi-continues inférieurement et propres sont rappelées en Annexe A.

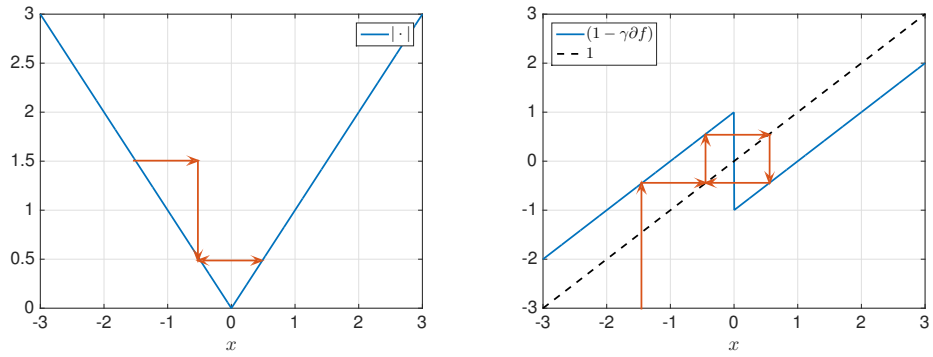



FIGURE 2.2 – **Algorithme de sous-gradient pour minimiser** $f = |\cdot|$**.** Pour $x^{[0]} = -1.5$ et $(\gamma_k)_{k\in\mathbb{N}} = \gamma = 1$.

**Exemple 2.1** (Algorithme de sous-gradient pour minimiser $f = |\cdot|$)**.** *Considérons le problème visant à minimiser la fonction* $f : x \in \mathbb{R} \mapsto |x|$ *illustrée Figure 2.2 (gauche) en bleu. Le problème de point fixe de l'algorithme de sous-gradient est représenté Figure 2.2 (droite) pour la suite constante* $(\gamma_k)_{k\in\mathbb{N}} = \gamma = 1$*. Les itérées* $x^{[k]}$ *pour l'initialisation* $x^{[0]} = -1.5$ *sont quant à elles représentées en orange. On remarque que si le pas est constant, les suites* $(x^{[2k+1]})_{k\in\mathbb{N}} = -0.5$ *et* $(x^{[2k]})_{k\in\mathbb{N}^*} = 0.5$ *sont constantes et l'algorithme ne converge pas vers la solution* $\widehat{x} = 0$*.*

Des garanties de convergence sont obtenues pour un pas décroissant conduisant à des problèmes numériques sur la solution estimée. Afin d'obtenir des résultats de convergence applicables en pratique, on considère des algorithmes de sous-gradient implicite, i.e., $\boldsymbol{w}^{[k]} \in \partial f(\boldsymbol{x}^{[k+1]})$, conduisant à l'algorithme du point proximal :

$$(\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} = \boldsymbol{x}^{[k]} - \gamma_k \boldsymbol{w}^{[k]} \quad \text{où} \quad \boldsymbol{w}^{[k]} \in \partial f(\boldsymbol{x}^{[k+1]}) \tag{2.16}$$

$$\Leftrightarrow \quad (\forall k \in \mathbb{N}) \quad \boldsymbol{x}^{[k+1]} = \operatorname{prox}_{\gamma_k f}(\boldsymbol{x}^{[k]}) \tag{2.17}$$

où $\operatorname{prox}_{\gamma_k f}$ est l'opérateur proximal de la fonction $\gamma_k f$. Le passage de (2.16) à (2.17) utilise la caractérisation sous-différentielle de l'opérateur proximal rappelée à la Proposition 2.1. La notion d'opérateur proximal introduit en [156] est rappelée ci-dessous.

**Définition 2.2** (Opérateur proximal [156, Notation 3.b])**.** *Soit une fonction* $f \in \Gamma_0(\mathbb{R}^N)$*. L'opérateur proximal de* $f$ *en* $\boldsymbol{u} \in \mathbb{R}^N$*, noté* $\operatorname{prox}_f(\boldsymbol{u})$ *est l'unique point qui minimise* $f + \frac{1}{2}\|\cdot - \boldsymbol{u}\|^2$*, i.e.,*

$$\operatorname{prox}_f : \mathbb{R}^N \to \mathbb{R}^N \tag{2.18}$$

$$\boldsymbol{u} \mapsto \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{x}\|_2^2 + f(\boldsymbol{x}), \tag{2.19}$$

Il est une généralisation de la projection $P_{\mathcal{C}}(\boldsymbol{u})$ de $\boldsymbol{u} \in \mathbb{R}^N$ sur un ensemble convexe $\mathcal{C} \subset \mathbb{R}^N$, i.e., $P_{\mathcal{C}}(\boldsymbol{u}) = \operatorname{prox}_{\imath_{\mathcal{C}}}(\boldsymbol{u})$ où $\imath_{\mathcal{C}}$ est la fonction indicatrice sur $\mathcal{C}$.

**Exemple 2.2** (Opérateur proximal de la norme $\ell_1$ [40])**.** *Soit* $\gamma \in ]0, +\infty[$ *et une fonction* $\mathbb{R}^N \to \mathbb{R} : \boldsymbol{u} \mapsto \gamma\|\boldsymbol{u}\|_1$*. On a alors, pour tout* $\boldsymbol{u} = (u_n)_{1\le n\le N} \in \mathbb{R}^N$*,* $\operatorname{prox}_{\gamma\|\cdot\|_1}(\boldsymbol{u}) = (\zeta_n)_{1\le n\le N}$ *avec*

$$(\forall n \in \{1, \dots, N\}), \quad \zeta_n = \begin{cases} u_n - \gamma & si \quad u_n > \gamma, \\ 0 & si \quad u_n \in [-\gamma, \gamma], \\ u_n + \gamma & si \quad u_n < -\gamma. \end{cases} \tag{2.20}$$

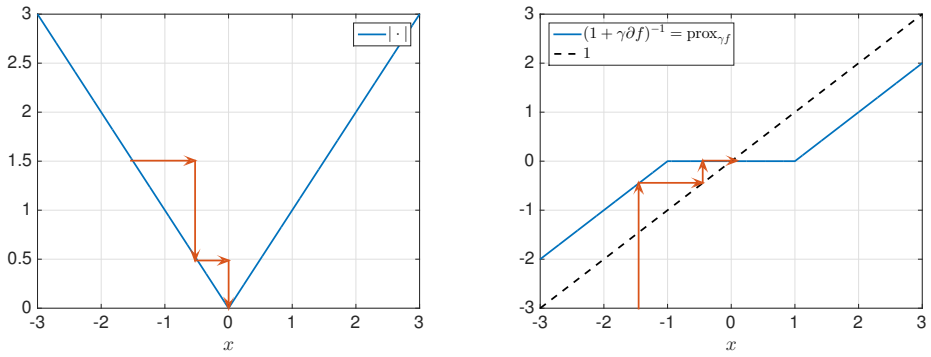



FIGURE 2.3 – **Algorithme proximal pour minimiser** $f = |\cdot|$**.** Pour $x^{[0]} = -1.5$ et $(\gamma_k)_{k\in\mathbb{N}} = \gamma = 1$.

*L'opérateur* $\mathrm{prox}_{\gamma\|\cdot\|_1}$ *correspond à l'opérateur de seuillage doux noté* $\mathrm{soft}_\gamma = \mathrm{sign}(\cdot)\max\{|\cdot| - \gamma, 0\}$*, illustré Figure 2.3 (droite, en bleu) pour* $\gamma = 1$*.*

**Exemple 2.3** (Algorithme proximal pour minimiser $f = |\cdot|$)**.** *Considérons le problème visant à minimiser la fonction* $f : x \in \mathbb{R} \mapsto |x|$ *illustrée Figure 2.3 (gauche) en bleu. Le problème de point fixe de l'algorithme proximal correspondant est représenté Figure 2.3 (droite) pour la suite constante* $(\gamma_k)_{k\in\mathbb{N}} = \gamma = 1$*. Les itérées* $x^{[k]}$ *pour l'initialisation* $x^{[0]} = -1.5$ *sont quant à elles représentées en orange. Contrairement à l'algorithme de sous-gradient, l'algorithme proximal converge en deux itérations vers* $\widehat{x} = 0$ *pour un pas constant.*

La caractérisation de l'opérateur proximal à partir de la sous-différentielle est due à Moreau [156, Proposition 6.a] et s'énonce ainsi :

**Proposition 2.1** (Caractérisation sous-différentielle de l'opérateur proximal [156, Proposition 6.a])**.** *Soient une fonction* $f \in \Gamma_0(\mathbb{R}^N)$ *et* $\boldsymbol{p} \in \mathbb{R}^N$*, alors*

$$\boldsymbol{p} = \mathrm{prox}_f \boldsymbol{x} \quad \Leftrightarrow \quad \boldsymbol{x} - \boldsymbol{p} \in \partial f(\boldsymbol{p}) \quad \Leftrightarrow \quad \boldsymbol{p} = (\mathbb{1} + \partial f)^{-1}(\boldsymbol{x}) \tag{2.21}$$

*où l'application* $(\mathbb{1} + \partial f)^{-1}$ *est appelée résolvante de l'opérateur* $\partial f$*.*

## 2.2 Résolution du problème $\ell_1$-TV

Dans cette section, nous détaillerons et motiverons les principaux algorithmes permettant de résoudre le Problème 2.2. Nous avons pris le parti de présenter les algorithmes proximaux pour leur flexibilité par rapport aux méthodes en ligne et car nous les utiliserons à plusieurs reprises dans ce manuscrit comme méthodes de référence.

D'après la Définition 2.2 de l'opérateur proximal d'une fonction, le minimiseur $\widehat{\boldsymbol{x}}_\lambda$ de (2.6) peut se formuler comme suit

$$\widehat{\boldsymbol{x}}_\lambda = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 + \underbrace{\lambda\|L\boldsymbol{x}\|_1}_{g\circ L(\boldsymbol{x})} = \mathrm{prox}_{g\circ L}(\boldsymbol{y}), \tag{2.22}$$

de gradient Lipschitz, où $g = \lambda\|\cdot\|_1$ est une fonction convexe non lisse et $L$ est un opérateur linéaire. De nombreux opérateurs proximaux possèdent une forme explicite (voir [40, 49]) dont celui de la norme $\ell_1$ rappelé à l'Exemple 2.2.

Cependant, la difficulté provient ici de l'opérateur linéaire $L$. En effet, $\mathrm{prox}_{g\circ L}$ possède une forme explicite si $L$ satisfait la condition faisant l'objet de la proposition suivante

**Proposition 2.2** (Voir [48, Proposition 11]). *Soient $g \in \Gamma_0(\mathbb{R}^M)$ et $L \in \mathbb{R}^{M\times N}$ un opérateur linéaire. Si*

$$LL^* = \nu\mathbb{1} \quad où \quad \nu \in ]0, +\infty[ \tag{2.23}$$

*alors $g \circ L \in \Gamma_0(\mathbb{R}^N)$ et*

$$\mathrm{prox}_{g\circ L} = \mathbb{1} + \nu^{-1} L^* \circ (\mathrm{prox}_{\nu g} - \mathbb{1}) \circ L. \tag{2.24}$$

Or, cette condition n'est pas vérifiée pour $L$ défini en (2.3). Afin de pallier cette difficulté, nous allons voir différentes méthodes d'éclatement proximales (« splitting ») permettant de converger vers $\widehat{\boldsymbol{x}}_\lambda$. Trois approches sont possibles.

La première repose sur la formulation du problème dual associé à (2.6) afin de déplacer $L$ dans le terme différentiable. Le problème peut ensuite ensuite être résolu à l'aide d'un algorithme forward-backward (Section 2.2.1).

La deuxième consiste à augmenter la dimensionnalité du problème en incluant des variables additionnelles. Nous présenterons à cet effet l'algorithme de Douglas-Rachford (Section 2.2.2) et l'algorithme ADMM (Section 2.2.3).

La troisième cherche à résoudre conjointement le problème primal et le problème dual à l'aide d'un algorithme primal-dual (Section 2.2.4).

### 2.2.1 Passage au dual

Une première stratégie pour trouver le minimiseur de (2.6) consiste à résoudre son problème dual au sens de la dualité de Fenchel-Moreau-Rockafellar [81, 156, 180] de façon à déplacer l'opérateur $L$ dans le terme différentiable.

En optimisation convexe, à chaque problème peut être associé un problème dual impliquant les fonctions conjuguées du problème initial, appelé par opposition problème primal. Nous suivrons ici les définitions du problème primal et du problème dual suivantes :

**Définition 2.3** (Conjugué de Fenchel [81]). *Le conjugué de Fenchel d'une fonction $f : \mathbb{R}^N \to ]-\infty, +\infty]$ est la fonction $f^* : \mathbb{R}^N \to ]-\infty, +\infty]$ définie par :*

$$(\forall \boldsymbol{u} \in \mathbb{R}^N) \quad f^*(\boldsymbol{u}) = \sup_{\boldsymbol{v}\in \mathrm{dom}\, f} \left(\langle \boldsymbol{u}, \boldsymbol{v}\rangle - f(\boldsymbol{v})\right) \tag{2.25}$$

*où* $\mathrm{dom}\, f$ *est le domaine de la fonction $f$ (voir Définition A.3).*

**Problème 2.4** (Problème de minimisation composite $f + g \circ L$ [221]). *Soient deux fonctions $f : \mathbb{R}^N \to ]-\infty, +\infty]$ et $g : \mathbb{R}^M \to ]-\infty, +\infty]$ et une matrice $L \in \mathbb{R}^{M\times N}$. Nous appellerons problème primal*

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} f(\boldsymbol{x}) + g(L\boldsymbol{x}), \tag{2.26}$$

*et problème dual*

$$-\min_{\boldsymbol{u}\in\mathbb{R}^M} f^*(L^*\boldsymbol{u}) + g^*(-\boldsymbol{u}), \tag{2.27}$$

*où $f^*$ et $g^*$ sont respectivement les conjugués de Fenchel de $f$ et $g$.*

**Remarque 2.5.** *Notons que la formulation duale usuelle implique $-\boldsymbol{u}$ à la place de $\boldsymbol{u}$ dans* (2.27)*. Nous choisissons cette définition par soucis de cohérence avec les résultats obtenus en [52] que nous étendrons au cadre vectoriel au Chapitre 4.*

**Remarque 2.6** (Voir [81]). *Le conjugué de Fenchel d'une fonction différentiable est également appelé transformée de Legendre.*

**Proposition 2.3** (Voir [81]). *Le conjugué de Fenchel d'une fonction est nécessairement convexe.*

En effet, $f^*$ est convexe puisqu'il est défini en chaque point $\boldsymbol{u}$ comme le supremum d'une famille de fonction affines (donc convexes) en $\boldsymbol{u}$.

Afin d'obtenir le problème dual associé au Problème 2.2, nous avons besoin de connaitre les fonctions conjuguées de $f = \frac{1}{2}\|\boldsymbol{y} - \cdot\|_2^2$ et $g = \lambda\|\cdot\|_1$. Pour se faire, nous présentons les conjugués de Fenchel des deux fonctions suivantes [30, 22] :

**Exemple 2.4** (Conjugué de la norme $\ell_p^p$ [22, Exemple 13.2]). *Soit $\|\cdot\|_p$ une norme $\ell_p$ avec $p \in ]1, +\infty[$. Le conjugué de Fenchel de $f = \frac{1}{p}\|\cdot\|_p^p$ est la fonction $f^* = \frac{1}{p^*}\|\cdot\|_{p^*}^{p^*}$ telle que $\frac{1}{p} + \frac{1}{p^*} = 1$.*

**Exemple 2.5** (Conjugué de la norme $\ell_p$ [22, Exemple 13.3]). *Soit $\|\cdot\|_p$ une norme $\ell_p$ avec $p \geq 1$ et un paramètre $\lambda \geq 0$. Le conjugué de Fenchel de $f = \lambda\|\cdot\|_p$ est la fonction indicatrice sur la boule $\ell_{p^*}$ de rayon $\lambda$ telle que $\frac{1}{p} + \frac{1}{p^*} = 1$, i.e.,*

$$(\forall \boldsymbol{u} \in \mathbb{R}^N) \quad f^*(\boldsymbol{u}) = \imath_{\|\cdot\|_{p^*} \leq \lambda}(\boldsymbol{u}) \equiv \begin{cases} 0 & \text{si } \|\boldsymbol{u}\|_{p^*} \leq \lambda, \\ +\infty & \text{sinon.} \end{cases} \tag{2.28}$$

Par conséquent, nous obtenons directement que $g^* = \imath_{\|\cdot\|_\infty \leq \lambda}$ et $f^* = \frac{1}{2}\|\boldsymbol{y} + \cdot\|_2^2$. Nous pouvons alors définir le problème dual associé au Problème 2.2.

**Problème 2.5** (Problème $\ell_1$-TV dual). *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\lambda \geq 0$. On appelle problème $\ell_1$-TV dual le problème consistant à trouver*

$$\widehat{\boldsymbol{u}}_\lambda \in \underset{u \in \mathbb{R}^{N-1}}{\operatorname{Argmin}} \frac{1}{2}\|\boldsymbol{y} + L^*\boldsymbol{u}\|_2^2 + \imath_{\|\cdot\|_\infty \leq \lambda}(\boldsymbol{u}). \tag{2.29}$$

Finalement, le Problème 2.5 consiste à minimiser la somme de deux fonctions convexes

$$\widehat{\boldsymbol{u}} \in \underset{\boldsymbol{u} \in \mathbb{R}^{N-1}}{\operatorname{Argmin}} \phi(\boldsymbol{u}) + \psi(\boldsymbol{u}) \tag{2.30}$$

où $\phi = \frac{1}{2}\|\boldsymbol{y} + L^* \cdot\|_2^2$ est dérivable et de gradient $\beta$-Lipschitz avec $\beta = \|L\|^2$, où $\|\cdot\|$ est la norme d'opérateur défini par

$$\|L\| = \sup_{\boldsymbol{u} \in \mathbb{R}^N, \boldsymbol{u} \neq \mathbf{0_N}} \frac{\|L\boldsymbol{u}\|_2}{\|\boldsymbol{u}\|_2}, \tag{2.31}$$

et $\psi = \imath_{\|\cdot\|_\infty \leq \lambda}$ est non lisse. Puisque $\phi + \psi \in \Gamma_0(\mathbb{R}^{N-1})$ est coercive [2], alors (2.30) admet au moins une solution $\widehat{\boldsymbol{u}}$ [51, Proposition 3.1]. D'après la règle de Fermat, $\widehat{\boldsymbol{u}}$ est entièrement caractérisé par

$$0 \in \partial(\phi + \psi)(\widehat{\boldsymbol{u}}) = \{\nabla\phi(\widehat{\boldsymbol{u}})\} + \partial\psi(\widehat{\boldsymbol{u}}) \tag{2.32}$$

$$\Leftrightarrow \quad -\nabla\phi(\widehat{\boldsymbol{u}}) \in \partial\psi(\widehat{\boldsymbol{u}}) \tag{2.33}$$

$$\Leftrightarrow \quad (\widehat{\boldsymbol{u}} - \gamma\nabla\phi(\widehat{\boldsymbol{u}})) - \widehat{\boldsymbol{u}} \in \gamma\partial\psi(\widehat{\boldsymbol{u}}) \tag{2.34}$$

$$\Leftrightarrow \quad \widehat{\boldsymbol{u}} = \operatorname{prox}_{\gamma\psi}(\widehat{\boldsymbol{u}} - \gamma\nabla\phi(\widehat{\boldsymbol{u}})) \tag{2.35}$$

La dernière équation suggère que $\widehat{\boldsymbol{u}}$ peut-être obtenu comme le point fixe d'une certaine application impliquant $\operatorname{prox}_{\gamma\psi}$ et $\nabla\phi = L(L^*\cdot + \boldsymbol{y})$. Lorsque $\operatorname{prox}_{\gamma\psi}$ n'est pas connu, l'identité de Moreau est utile pour calculer l'opérateur proximal d'une fonction à partir de celui de son conjugué, et inversement.

2. La définition de la coercivité est rappelée en Annexe A.

**Proposition 2.4** (Identité de Moreau [156, Proposition 6.c]). *Soient $\tau \geq 0$, $f : \mathbb{R}^N \to ]-\infty, +\infty]$ et $f^*$ son conjugué de Fenchel, alors*

$$(\forall \boldsymbol{u} \in \mathbb{R}^N) \quad \boldsymbol{u} = \mathrm{prox}_{\tau f^*}(\boldsymbol{u}) + \tau \mathrm{prox}_{f/\tau}(\boldsymbol{u}/\tau) \tag{2.36}$$

Dans notre cas où $\psi = g^* = \imath_{\|\cdot\|_\infty \leq \lambda}$ avec $g = \lambda\|\cdot\|_1$, nous obtenons ainsi que $\mathrm{prox}_{\gamma\psi} = \cdot - \gamma \mathrm{soft}_{\lambda/\gamma}(\cdot/\gamma)$.

Le Problème 2.2 ($\ell_1$-TV) peut être résolu par l'algorithme forward-backward qui combine une étape explicite (descente de gradient) et une étape implicite (itération proximale) [47, 51]. Les itérations sont présentées par l'Algorithme 1.

**Algorithme 1** Résolution du Problème 2.2 ($\ell_1$-TV) basée sur l'algorithme forward-backward

**Entrée :** Paramètres $\beta = \|L\|^2$ et $\epsilon \in ]0, \min(1, 1/\beta)]$.

1: **Pour** k=0,1,... **faire**
2: $\quad \gamma_k \in [\epsilon, 2/\beta - \epsilon]$
3: $\quad \alpha_k \in [\epsilon, 1]$
4: $\quad \boldsymbol{u}^{[k+\frac{1}{2}]} = \boldsymbol{u}^{[k]} - \gamma_k L(L^*\boldsymbol{u}^{[k]} + \boldsymbol{y})$
5: $\quad \boldsymbol{u}^{[k+1]} = \boldsymbol{u}^{[k]} + \alpha_k(\boldsymbol{u}^{[k+\frac{1}{2}]} - \gamma_k \mathrm{soft}_{\lambda/\gamma_k}(\boldsymbol{u}^{[k+\frac{1}{2}]}/\gamma_k) - \boldsymbol{u}^{[k]})$

**Sortie :** $\widehat{\boldsymbol{x}}_\lambda = \boldsymbol{y} + L^*\widehat{\boldsymbol{u}}_\lambda$ avec $\widehat{\boldsymbol{u}}_\lambda = \lim_{k\to+\infty} \boldsymbol{u}^{[k]}$.

Sous certaines conditions énoncées dans le théorème de Fenchel-Rockafellar, et reprises dans le lemme suivant, la borne inférieure du problème primal est égale au minimum du problème dual. On dit alors que le *saut de dualité* est nul.

**Lemme 2.1** (Formule de dualité de Fenchel-Rockafellar [221, Corollaire 2.8.5]). *Soient deux fonctions $f \in \Gamma_0(\mathbb{R}^N)$, $g \in \Gamma_0(\mathbb{R}^M)$ et $L \in \mathbb{R}^{M\times N}$ tel que $0 \in \mathrm{sri}\,(L(\mathrm{dom}\, f) - \mathrm{dom}\, g)$ [3]. On a alors*

$$\inf_{\boldsymbol{x}\in\mathbb{R}^N} f(\boldsymbol{x}) + g(L\boldsymbol{x}) = -\min_{u\in\mathbb{R}^M} f^*(L^*\boldsymbol{u}) + g^*(-\boldsymbol{u}). \tag{2.37}$$

*De plus, $\widehat{\boldsymbol{x}} \in \mathbb{R}^N$ est un minimiseur de $f + g \circ L$ si et seulement si il existe $\widehat{\boldsymbol{u}} \in \mathbb{R}^M$ tel que $L^*\widehat{\boldsymbol{u}} \in \partial f(\widehat{\boldsymbol{x}})$ et $-\widehat{\boldsymbol{u}} \in \partial g(L\widehat{\boldsymbol{x}})$.*

Cette condition a l'avantage de pouvoir nous permettre de calculer la solution $\widehat{\boldsymbol{x}}_\lambda$ du Problème 2.2 (primal) à partir de la solution $\widehat{\boldsymbol{u}}_\lambda$ du Problème 2.5 (dual) :

$$\widehat{\boldsymbol{x}}_\lambda = \boldsymbol{y} + L^*\widehat{\boldsymbol{u}}_\lambda. \tag{2.38}$$

### 2.2.2 Augmentation de la dimensionnalité

Une autre stratégie consiste à introduire une variable auxiliaire $\boldsymbol{z}$ assujettie à la contrainte $\boldsymbol{z} = L\boldsymbol{x}$. Par conséquent, si nous définissons un nouvel opérateur linéaire $A = [L, -\mathbb{1}_{N-1}] \in \mathbb{R}^{(N-1)\times(2N-1)}$, la contrainte $\boldsymbol{z} = L\boldsymbol{x}$ peut se réécrire $(\boldsymbol{x}, \boldsymbol{z}) \in \mathrm{Ker}A$ et être introduite dans le problème variationnel comme suit :

$$\min_{(\boldsymbol{x},\boldsymbol{z})\in\mathbb{R}^N\times\mathbb{R}^{N-1}} \underbrace{\frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2}_{f(\boldsymbol{x})} + \underbrace{\lambda\|\boldsymbol{z}\|_1}_{g(\boldsymbol{z})} + \underbrace{\imath_{\mathrm{Ker}A}(\boldsymbol{x}, \boldsymbol{z})}_{\psi(\boldsymbol{x},\boldsymbol{z})} \tag{2.39}$$

3. La définition de l'intérieur relatif fort (sri) d'un ensemble est rappelé dans l'Annexe A

Cette réécriture du problème $\ell_1$-TV peut alors s'interpréter comme un problème de minimisation de la somme de deux fonctions $\phi = f + g$ et $\psi$ dépendant chacune de deux variables, i.e.,

$$\min_{(\boldsymbol{x},\boldsymbol{z})\in\mathbb{R}^N\times\mathbb{R}^{N-1}} \phi(\boldsymbol{x},\boldsymbol{z}) + \psi(\boldsymbol{x},\boldsymbol{z}) \tag{2.40}$$

où $\phi(\boldsymbol{x},\boldsymbol{z}) = f(\boldsymbol{x}) + g(\boldsymbol{z})$ est une fonction séparable en $\boldsymbol{x}$ et $\boldsymbol{z}$. Les opérateurs proximaux associés à $\phi$ et $\psi$ possèdent une forme explicite connue [49]. En effet, pour $\gamma > 0$,

$$\operatorname{prox}_{\gamma\phi}(\boldsymbol{x},\boldsymbol{z}) = \big(\operatorname{prox}_{\gamma f}(\boldsymbol{x}), \operatorname{prox}_{\gamma g}(\boldsymbol{z})\big) = \left(\frac{\boldsymbol{x}+\gamma\boldsymbol{y}}{1+\gamma}, \operatorname{soft}_{\gamma\lambda}(\boldsymbol{z})\right), \tag{2.41}$$

$$\operatorname{prox}_{\gamma\psi}(\boldsymbol{x},\boldsymbol{z}) = (\boldsymbol{x},\boldsymbol{z}) - A^*(AA^*)^{-1}A(\boldsymbol{x},\boldsymbol{z}). \tag{2.42}$$

Dans cette situation, l'algorithme de Douglas-Rachford permet de converger vers la solution $(\widehat{\boldsymbol{x}},\widehat{\boldsymbol{z}})$ de (2.39) en utilisant alternativement $\operatorname{prox}_{\gamma\phi}$ et $\operatorname{prox}_{\gamma\psi}$. Les itérations correspondantes pour résoudre le Problème 2.2 ($\ell_1$-TV) sont présentées dans l'Algorithme 2.

**Algorithme 2** Résolution du Problème 2.2 ($\ell_1$-TV) basée sur l'algorithme de Douglas-Rachford

**Entrée :** Paramètres $\epsilon \in ]0,1[$ et $\gamma > 0$.

1: **Pour** k=0,1,... **faire**

2: $\quad \alpha_k \in [\epsilon, 2-\epsilon]$

3: $\quad (\boldsymbol{x}^{[k+\frac{1}{2}]}, \boldsymbol{z}^{[k+\frac{1}{2}]}) = (\boldsymbol{x}^{[k]}, \boldsymbol{z}^{[k]}) - A^*(AA^*)^{-1}A(\boldsymbol{x}^{[k]}, \boldsymbol{z}^{[k]})$

4: $\quad \begin{pmatrix}\boldsymbol{x}^{[k+1]}\\ \boldsymbol{z}^{[k+1]}\end{pmatrix} = \begin{pmatrix}\boldsymbol{x}^{[k]}\\ \boldsymbol{z}^{[k]}\end{pmatrix} + \alpha_k \begin{pmatrix}(1+\gamma)^{-1}(2\boldsymbol{x}^{[k]} - \boldsymbol{x}^{[k+\frac{1}{2}]} + \gamma\boldsymbol{y}) - \boldsymbol{x}^{[k]}\\ \operatorname{soft}_{\gamma\lambda}(2\boldsymbol{z}^{[k]} - \boldsymbol{z}^{[k+\frac{1}{2}]}) - \boldsymbol{z}^{[k]}\end{pmatrix}$

**Sortie :** $\widehat{\boldsymbol{x}}_\lambda = \lim_{k\to+\infty} \boldsymbol{x}^{[k]}$

Contrairement à l'algorithme forward-backward, l'algorithme de Douglas-Rachford est adapté au cas où les deux fonctions à minimiser sont non lisses. L'algorithme de Douglas-Rachford a originellement été introduit en [74] pour résoudre numériquement le problème de conduction de la chaleur à 2 et 3 dimensions. Plus récemment, des améliorations et des résultats généraux sur la convergence de l'algorithme de Douglas-Rachford ont été développés dans [78, 48, 47].

### 2.2.3 Lagrangien augmenté

Les méthodes de directions alternées des multiplicateurs, communément appelées ADMM, sont une classe d'algorithmes permettant de résoudre des problèmes de minimisation sous contraintes. Elles reposent sur la décomposition du problème de minimisation en plusieurs problèmes séparables et sur une adaptation de la méthode des multiplicateurs de Lagrange [209, 29]. Dans cette section, nous présenterons comment résoudre le problème $\ell_1$-TV via un algorithme ADMM [29, 209, 193].

Le problème formulé en (2.39) peut être réécrit de manière équivalente

$$\min_{\boldsymbol{x}\in\mathbb{R}^N,\boldsymbol{z}\in\mathbb{R}^{N-1},\boldsymbol{r}\in\mathbb{R}^N,\boldsymbol{s}\in\mathbb{R}^{N-1}} f(\boldsymbol{x}) + g(\boldsymbol{z}) + \imath_{\operatorname{Ker}A}(\boldsymbol{r},\boldsymbol{s}) \quad \text{sujet à} \quad \begin{cases}\boldsymbol{r} = \boldsymbol{x},\\ \boldsymbol{s} = \boldsymbol{z}.\end{cases} \tag{2.43}$$

où cette fois-ci le critère à minimiser est la somme de 3 fonctions qui dépendent chacune de variables différentes.

Le Lagrangien augmenté associé à (2.43) est la somme du Lagrangien du problème non contraint et d'un terme de pénalisation quadratique supplémentaire. Si nous introduisons des variables duales $\tilde{\boldsymbol{u}}$ et $\tilde{\boldsymbol{t}}$ respectivement associées aux contraintes $\boldsymbol{r} = \boldsymbol{x}$ et $\boldsymbol{s} = \boldsymbol{z}$, celui-ci s'écrit

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{r}, \boldsymbol{s}, \tilde{\boldsymbol{u}}, \tilde{\boldsymbol{t}}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \imath_{\mathrm{Ker}A}(\boldsymbol{r}, \boldsymbol{s}) + \tilde{\boldsymbol{u}}^\top(\boldsymbol{x} - \boldsymbol{r}) + \tilde{\boldsymbol{t}}^\top(\boldsymbol{z} - \boldsymbol{s}) \\ + \underbrace{\frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{r}\|^2 + \frac{\rho}{2}\|\boldsymbol{z} - \boldsymbol{s}\|^2}_{\text{pénalisation}}\end{aligned} \tag{2.44}$$

où $\rho > 0$ est un paramètre de pénalisation fixé. Si l'on fait intervenir les variables duales mises à l'échelle $\boldsymbol{u} = \tilde{\boldsymbol{u}}/\rho$ et $\boldsymbol{t} = \tilde{\boldsymbol{t}}/\rho$ alors (2.44) peut-être réécrit sous la forme

$$\begin{aligned}\mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{u}, \boldsymbol{t}) = \quad & f(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{r} + \boldsymbol{u}\|^2 - \frac{\rho}{2}\|\boldsymbol{u}\|^2 \\ & + g(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{z} - \boldsymbol{s} + \boldsymbol{t}\|^2 - \frac{\rho}{2}\|\boldsymbol{t}\|^2 \\ & + \imath_{\mathrm{Ker}A}(\boldsymbol{r}, \boldsymbol{s}).\end{aligned} \tag{2.45}$$

Finalement, le problème consiste à résoudre

$$\max_{(\boldsymbol{u}\in\mathbb{R}^N, \boldsymbol{t}\in\mathbb{R}^N)} \inf_{(\boldsymbol{x}\in\mathbb{R}^N, \boldsymbol{z}\in\mathbb{R}^{N-1}, \boldsymbol{r}\in\mathbb{R}^N, \boldsymbol{s}\in\mathbb{R}^{N-1})} \mathcal{L}(\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{r}, \boldsymbol{s}, \boldsymbol{u}, \boldsymbol{t}). \tag{2.46}$$

A chaque itération, le Lagrangien augmenté est minimisé en alternant la minimisation par rapport à $\boldsymbol{x}$ (Algorithme 3, étape 3), $\boldsymbol{z}$ (Algorithme 3, étape 4), $\boldsymbol{r}$ et $\boldsymbol{s}$ (Algorithme 3, étape 6).

Les variables duales $\boldsymbol{u}$ et $\boldsymbol{t}$ sont quant à elle obtenues par ascension de gradient (Algorithme 3, étapes 8 et 9).

**Algorithme 3** Résolution du Problème 2.2 ($\ell_1$-TV) basée sur l'algorithme ADMM

**Entrée :** Paramètre $\rho > 0$.

1: **Pour** k=1,... **faire**
2: **Mise à jour des variables primales :**
3: $\boldsymbol{x}^{[k+1]} = \mathrm{argmin}_{\boldsymbol{x}\in\mathbb{R}^N}\ f(\boldsymbol{x}) + \frac{\rho}{2}\|\boldsymbol{x} - \boldsymbol{r}^{[k]} + \boldsymbol{u}^{[k]}\|_2^2 = \frac{\boldsymbol{y} + \rho(\boldsymbol{z}^{[k]} - \boldsymbol{u}^{[k]})}{1+\rho}$
4: $\boldsymbol{z}^{[k+1]} = \mathrm{argmin}_{\boldsymbol{z}\in\mathbb{R}^{N-1}}\ g(\boldsymbol{z}) + \frac{\rho}{2}\|\boldsymbol{z} - \boldsymbol{s}^{[k]} + \boldsymbol{t}^{[k]}\|_2^2 = \mathrm{soft}_{\lambda/\rho}\left(\boldsymbol{s}^{[k]} - \boldsymbol{t}^{[k]}\right)$
5: **Projection sur l'ensemble des contraintes :**
6: $\left(\boldsymbol{r}^{[k+1]}, \boldsymbol{s}^{[k+1]}\right) = P_{\mathrm{Ker}A}\left(\boldsymbol{x}^{[k+1]} + \boldsymbol{u}^{[k]}, \boldsymbol{z}^{[k+1]} + \boldsymbol{t}^{[k]}\right)$
7: **Mise à jour des variables duales :**
8: $\boldsymbol{u}^{[k+1]} = \boldsymbol{u}^{[k]} + (\boldsymbol{x}^{[k+1]} - \boldsymbol{r}^{[k+1]})$
9: $\boldsymbol{t}^{[k+1]} = \boldsymbol{t}^{[k]} + (\boldsymbol{z}^{[k+1]} - \boldsymbol{s}^{[k+1]})$

**Sortie :** $(\widehat{\boldsymbol{x}}_\lambda, \widehat{\boldsymbol{u}}_\lambda) = \lim_{k\to+\infty}(\boldsymbol{x}^{[k]}, \boldsymbol{u}^{[k]})$

### 2.2.4 Algorithme primal-dual

Une troisième méthode consiste à combiner une minimisation dans l'espace primal et l'espace dual. A partir de la Définition 2.3 du conjugué de Fenchel, nous pouvons remarquer qu'il est également possible d'exprimer

$$g(L\boldsymbol{x}) = \sup_{\boldsymbol{u}\in\mathbb{R}^{N-1}} \langle \boldsymbol{u} \,|\, L\boldsymbol{x}\rangle - g^*(\boldsymbol{u}). \tag{2.47}$$

Par conséquent, les problèmes primal et dual énoncés dans le Problème 2.4 peuvent être combinés en un seul problème

$$\min_{\boldsymbol{x}\in\mathbb{R}^N} \sup_{\boldsymbol{u}\in\mathbb{R}^{N-1}} \tilde{\mathcal{L}}(\boldsymbol{x},\boldsymbol{u}) \quad \text{où} \quad \tilde{\mathcal{L}}(\boldsymbol{x},\boldsymbol{u}) = \langle L\boldsymbol{x} \,|\, \boldsymbol{u}\rangle + f(\boldsymbol{x}) - g^*(\boldsymbol{u}), \tag{2.48}$$

appelé problème primal-dual. Ainsi, pour $f = \frac{1}{2}\|\boldsymbol{y} - \cdot\|_2^2$ et $g^* = \imath_{\|\cdot\|_\infty \leq \lambda}$ nous obtenons le problème $\ell_1$-TV primal dual suivant

**Problème 2.6** (Problème $\ell_1$-TV primal-dual)**.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\lambda \geq 0$. On appelle problème $\ell_1$-TV primal-dual le problème consistant à trouver*

$$(\widehat{\boldsymbol{x}}_\lambda, \widehat{\boldsymbol{u}}_\lambda) = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \sup_{\boldsymbol{u}\in\mathbb{R}^{N-1}} (L\boldsymbol{x})^\top \boldsymbol{u} + \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{x}\|_2^2 - \imath_{\|\cdot\|_\infty \leq \lambda}(\boldsymbol{u}). \tag{2.49}$$

La nouvelle formulation (2.48) montre que la solution $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}})$ est le point scelle du Lagrangien $\tilde{\mathcal{L}}$, i.e., $0 \in \partial\tilde{\mathcal{L}}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}})$. Afin de mettre au point un algorithme proximal, il est important de remarquer que pour $\tau > 0$ et $\sigma > 0$,

$$0 \in \partial\tilde{\mathcal{L}}(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}}) \quad \Leftrightarrow \quad \begin{cases} 0 \in \{L^*\widehat{\boldsymbol{u}}\} + \partial f(\widehat{\boldsymbol{x}}) \\ 0 \in \{L\widehat{\boldsymbol{x}}\} - \partial g^*(\widehat{\boldsymbol{u}}) \end{cases} \tag{2.50}$$

$$\Leftrightarrow \quad \begin{cases} -L^*\widehat{\boldsymbol{u}} \in \partial f(\widehat{\boldsymbol{x}}) \\ L\widehat{\boldsymbol{x}} \in \partial g^\star(\widehat{\boldsymbol{u}}) \end{cases} \tag{2.51}$$

$$\Leftrightarrow \quad \begin{cases} (\widehat{\boldsymbol{x}} - \tau L^*\widehat{\boldsymbol{u}}) - \widehat{\boldsymbol{x}} \in \tau\partial f(\widehat{\boldsymbol{x}}) \\ (\widehat{\boldsymbol{u}} + \sigma L\widehat{\boldsymbol{x}}) - \widehat{\boldsymbol{u}} \in \sigma\partial g^*(\widehat{\boldsymbol{u}}) \end{cases} \tag{2.52}$$

$$\Leftrightarrow \quad \begin{cases} \widehat{\boldsymbol{x}} = \operatorname{prox}_{\tau f}(\widehat{\boldsymbol{x}} - \tau L^*\widehat{\boldsymbol{u}}) \\ \widehat{\boldsymbol{u}} = \operatorname{prox}_{\sigma g^*}(\widehat{\boldsymbol{u}} + \sigma L\widehat{\boldsymbol{x}}) \end{cases} \tag{2.53}$$

A partir de la dernière équation, il apparaît que $\widehat{\boldsymbol{x}}$ (resp. $\widehat{\boldsymbol{u}}$) est le point fixe d'une application dépendant de $\widehat{\boldsymbol{u}}$ (resp. $\widehat{\boldsymbol{x}}$). Une stratégie pour obtenir $(\widehat{\boldsymbol{x}}, \widehat{\boldsymbol{u}})$ consiste à alterner des étapes de descente proximale comme énoncé dans l'Algorithme 4 suivant [38, 53].

**Algorithme 4** Résolution du Problème 2.2 ($\ell_1$-TV) basée sur l'algorithme primal-dual

**Entrée :** Paramètres $0 \leq \theta \leq 1$, $\tau$ et $\sigma$ tels que $\tau\sigma\|L\|_2^2 < 1$.

1: **Pour** k=0,1,... **faire**
2: $\quad \boldsymbol{x}^{[k+1]} = \frac{1}{1+\tau}\left(\boldsymbol{x}^{[k]} - \tau L^*(\boldsymbol{u}^{[k]}) + \tau\boldsymbol{y}\right)$
3: $\quad \boldsymbol{u}^{[k+1]} = (\boldsymbol{u}^{[k]} + \sigma L(\tilde{\boldsymbol{x}}^{[k]}) - \gamma\operatorname{soft}_{\lambda/\gamma}\left((\boldsymbol{u}^{[k]} + \sigma L(\tilde{\boldsymbol{x}}^{[k]})/\gamma\right)$
4: $\quad \tilde{\boldsymbol{x}}^{[k+1]} = \boldsymbol{x}^{[k+1]} + \theta(\boldsymbol{x}^{[k+1]} - \boldsymbol{x}^{[k]})$

**Sortie :** $(\widehat{\boldsymbol{x}}_\lambda, \widehat{\boldsymbol{u}}_\lambda) = \lim_{k\to+\infty}(\boldsymbol{x}^{[k]}, \boldsymbol{u}^{[k]})$

### 2.2.5 Comparaison des algorithmes

Dans cette section, nous comparons les performances des quatre algorithmes présentés précédemment pour résoudre le Problème 2.2.

#### 2.2.5.1 Choix des paramètres des algorithmes

Jusqu'à présent, nous n'avons pas précisé comment choisir les paramètres des algorithmes $\ell_1$-TV afin d'obtenir la meilleure vitesse de convergence. Par souci de simplicité, nous nous

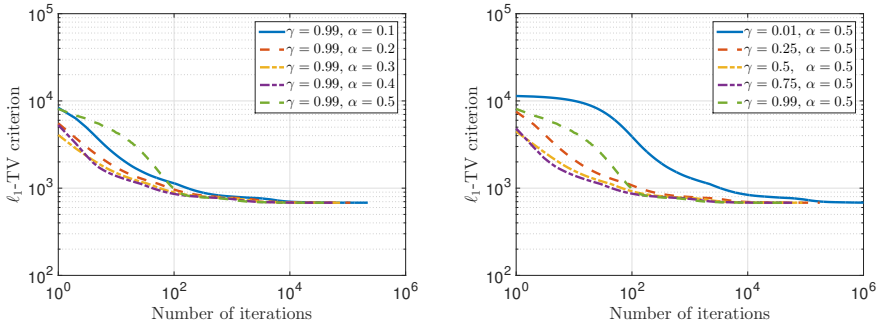



FIGURE 2.4 – **Algorithme forward-backward.** Influence des paramètres $\alpha$ et $\gamma$. Paramètres optimaux $\alpha = 0.5$ et $\gamma = 0.99$.

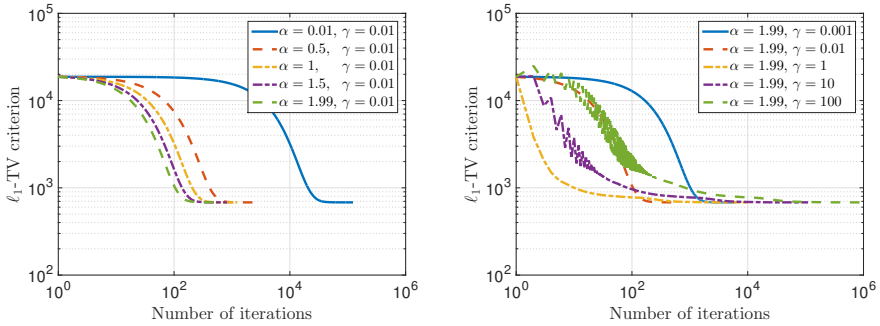



FIGURE 2.5 – **Algorithme Douglas-Rachford.** Influence des paramètres $\alpha$ et $\gamma$. Paramètres optimaux $\alpha = 1.99$ et $\gamma = 0.01$.

restreignons à des paramètres constants au fil des itérations et nous déciderons qu'un algorithme a suffisamment convergé lorsque $\|\boldsymbol{x}^{[k]} - \boldsymbol{x}^{[\infty]}\|^2 / \|\boldsymbol{x}^{[\infty]}\|^2 \leq 10^{-4}$, où $\boldsymbol{x}^{[\infty]}$ est obtenu en pratique pour $10^8$ itérations.

Les résultats qui suivent sont obtenus pour une même réalisation $\boldsymbol{y} \in \mathbb{R}^N$ ($N = 10^3$) et pour $\lambda = 20$. Des résultats similaires ont été observés pour différents $\lambda$ et différents rapports signal sur bruit.

Considérons l'Algorithme 1 forward-backward et les suites constantes $\alpha_k \equiv \alpha$ et $\gamma_k \equiv \gamma$ assujetties respectivement à $\alpha \in [\epsilon, 1]$ et $\gamma \in [\epsilon, 2/\beta - \epsilon]$ où $\epsilon \in ]0, \min(1, 1/\beta)]$ et $\beta = \|L\| = 2$. L'impact des paramètres $\alpha$ et $\gamma$ sur la vitesse de convergence de l'algorithme est illustré Figure 2.4. A $\gamma$ fixé, la Figure 2.4 (gauche) montre que le choix $\alpha = 0.5$ permet au critère de converger plus rapidement. De même, pour $\alpha = 0.5$ fixé, la Figure 2.4 (droite) montre que $\gamma = 0.99$ (i.e., $\gamma \to 1$) permet de stabiliser le critère en moins d'itérations. Dans la suite nous choisirons donc $\alpha = 0.5$ et $\gamma = 0.99$.

L'algorithme 2 Douglas-Rachford dépend des paramètres $\alpha \in [\epsilon, 2 - \epsilon]$ et $\gamma > 0$ où $\epsilon \in ]0, 1[$. L'influence du paramètre $\alpha$ est étudié Figure 2.5 (gauche) à $\gamma$ fixé, et montre que le choix $\alpha = 1.99$ (i.e., $\alpha \to 2$) permet de converger plus rapidement. Différentes gammes de $\gamma$ sont examinées Figure 2.5 (droite). Nous trouvons que les paramètres optimaux sont $\alpha = 1.99$ et $\gamma = 0.01$.

L'algorithme 3 ADMM dépend quant à lui seulement du paramètre $\rho$. Différentes gammes sont étudiées Figure 2.6 et les résultats montrent que le choix optimal est $\rho = 100$.

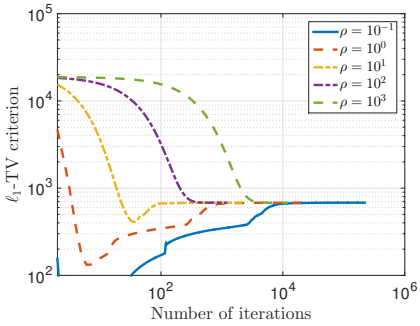



FIGURE 2.6 – **Algorithme ADMM.** Influence du paramètre $\rho$. Paramètre optimal : $\rho = 100$.

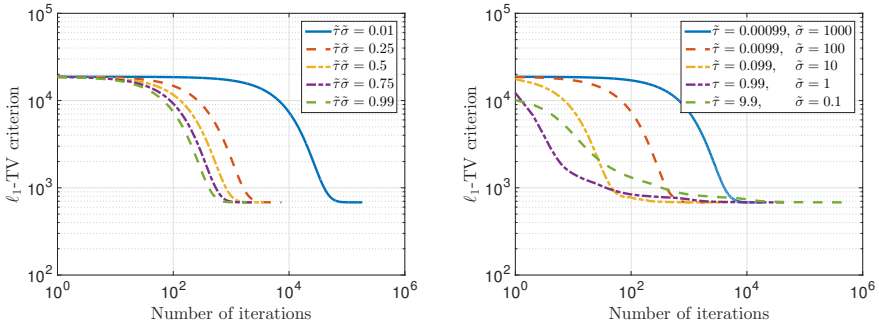



FIGURE 2.7 – **Algorithme Primal-Dual.** Influence des paramètres $\tilde{\tau} = \tau/\|L\|_2$ et $\tilde{\sigma} = \sigma/\|L\|_2$. Paramètres optimaux $\tilde{\tau} = 0.0099$ et $\tilde{\sigma} = 100$.

L'algorithme 4 Primal-Dual dépend de trois paramètres $\theta$, $\tau$ et $\sigma$ vérifiant $0 \leq \theta \leq 1$ et $\tau\sigma\|L\|_2^2 < 1$. Notons $\tilde{\tau} = \tau\|L\|_2$ et $\tilde{\sigma} = \sigma\|L\|_2$. Nous choisissons $\theta = 0$. L'impact du produit $\tilde{\tau}\tilde{\sigma} < 1$ est illustré Figure 2.7 (gauche) et montre que $\tilde{\tau}\tilde{\sigma} = 0.99$ (i.e., $\tilde{\tau}\tilde{\sigma} \to 1$) permet au critère de converger plus rapidement. Le rapport entre $\tilde{\tau}$ et $\tilde{\sigma}$ est quant à lui examiné Figure 2.7 (droite) et les valeurs optimales retenues sont $\tilde{\tau} = 0.099$ et $\tilde{\sigma} = 100$.

#### 2.2.5.2 Coût de calcul

Le temps d'exécution de chaque algorithme est représenté Figure 2.8 (gauche) en fonction de $\lambda$. On remarque que le temps de calcul de chaque algorithme dépend peu de $\lambda$ excepté l'algorithme forward-backward dont le temps d'exécution croît notablement avec $\lambda$. Pour $\lambda = 20$, les critères correspondants sont représentés Figure 2.8 (droite) en fonction du nombre d'itérations.

Les temps de calcul sont également comparés Figure 2.9 en fonction de la taille $N$ de l'observation pour $\lambda = 10$ (gauche) et $\lambda = 100$ (droite). On observe que le temps de calcul croît de la même manière en fonction de $N$, quel que soit l'algorithme. Globalement, l'algorithme Douglas-Rachford et l'algorithme ADMM sont les plus rapides. Des résultats identiques sont obtenus pour différents rapport signal sur bruit.

## 2.3 Evaluation des performances d'estimation

Soit une observation $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{b} \in \mathbb{R}^N$ d'un signal constant par morceaux $\overline{\boldsymbol{x}} \in \mathbb{R}^N$ corrompu par un bruit additif gaussien $\boldsymbol{b} \in \mathbb{R}^N$. Nous avons vu que la solution $\widehat{\boldsymbol{x}}_\chi$ du Problème 2.1 (resp.

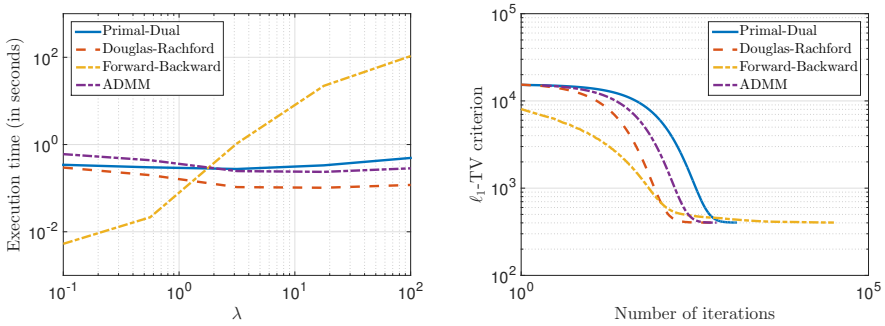



FIGURE 2.8 – **Comparaison des algorithmes $\ell_1$-TV (1).** Gauche : Temps de calcul en fonction de $\lambda$. Droite : Critères pour $\lambda = 20$.

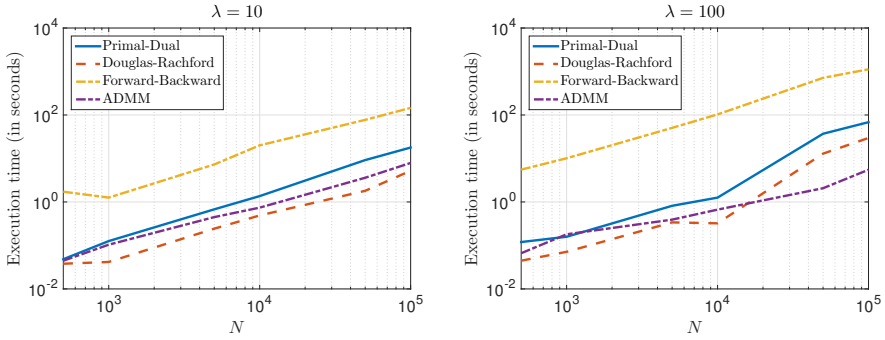



FIGURE 2.9 – **Comparaison des algorithmes $\ell_1$-TV (2).** Temps de calcul en fonction de la taille $N$ du signal. Gauche : $\lambda = 10$. Droite : $\lambda = 100$.

$\widehat{\boldsymbol{x}}_\lambda$ du Problème 2.2) est constante par morceaux et par conséquent fournit une estimation de $\overline{\boldsymbol{x}}$ pour un choix de $\chi$ (resp. $\lambda$) donné. Notons $\widehat{\boldsymbol{x}}$ l'une de ces solutions.

Afin de quantifier la qualité de l'estimateur $\widehat{\boldsymbol{x}}$ de $\overline{\boldsymbol{x}}$, nous pouvons par exemple comparer le nombre de segments de $\widehat{\boldsymbol{x}}$ et $\overline{\boldsymbol{x}}$. Cependant, nous avons vu que cela ne constitue pas un bon critère de performance à cause de l'effet d'escalier (voir Remarque 2.3).

Une seconde idée consiste à mesurer l'erreur quadratique moyenne relative entre $\widehat{\boldsymbol{x}}$ et $\overline{\boldsymbol{x}}$, i.e.,

$$\mathrm{MSEr}(\widehat{\boldsymbol{x}}, \overline{\boldsymbol{x}}) = \frac{\|\widehat{\boldsymbol{x}} - \overline{\boldsymbol{x}}\|^2}{\|\overline{\boldsymbol{x}}\|^2}. \tag{2.54}$$

Cependant, cette mesure ne permet pas de mesurer si les positions des discontinuités de $\overline{\boldsymbol{x}}$ ont été correctement estimées.

Une troisième idée consiste donc à considérer non pas $\widehat{\boldsymbol{x}}$ et $\overline{\boldsymbol{x}}$ mais leur vecteur d'indicatrices des discontinuités $\widehat{\boldsymbol{r}}$ et $\overline{\boldsymbol{r}}$ définis comme suit :

**Définition 2.4** (Indicatrice des discontinuités)**.** *Soit un $\boldsymbol{x} \in \mathbb{R}^N$ un signal constant par morceaux. Alors on définit le vecteur d'indicatrices des discontinuités $\boldsymbol{r} \in \mathbb{R}^{N-1}$ comme suit :*

$$(\forall n \in \{1, \dots, N-1\}) \quad r_n = \begin{cases} 1, \text{ si } x_n \neq x_{n+1}, \\ 0, \text{ sinon.} \end{cases} \tag{2.55}$$

Dès lors, la première mesure qui nous viendrait à l'esprit est la suivante

$$S(\widehat{\boldsymbol{r}}, \overline{\boldsymbol{r}}) = \frac{\|\widehat{\boldsymbol{r}} - \overline{\boldsymbol{r}}\|_0}{\|\overline{\boldsymbol{r}}\|_0}. \tag{2.56}$$

| $n=$ | 1 | 2 | 3 | 4 | 5 | $J(\boldsymbol{\alpha},\cdot)$ |
|---|---|---|---|---|---|---|
| $\alpha_n$ | 1 | 0 | 0 | 0 | 1 | 1 |
| $\beta_n^{(1)}$ | 1 | 0 | 0 | 0 | 0 | 1/2 |
| $\beta_n^{(2)}$ | 1 | 1 | 0 | 0 | 0 | 1/3 |
| $\beta_n^{(3)}$ | 1 | 1 | 0 | 0 | 1 | 2/3 |
| $\beta_n^{(4)}$ | 1 | 1 | 0 | 1 | 1 | 1/2 |

TABLE 2.1 – **Valeurs typiques de l'indice de Jaccard.**

Cependant $S$ est une mesure très sévère qui ne permet de prendre ne compte que les estimées parfaitement localisées. Mais en pratique, si les discontinuités de $\widehat{\boldsymbol{x}}$ sont localisées proches de celles de $\overline{\boldsymbol{r}}$, alors $\widehat{\boldsymbol{x}}$ est un estimateur satisfaisant dans une certaine mesure.

Dans la section suivante, nous montrons comment utiliser l'indice de Jaccard pondéré afin de mesurer la similarité entre $\overline{\boldsymbol{r}}$ et $\widehat{\boldsymbol{r}}$.

### 2.3.1 Rappels sur l'indice de Jaccard

L'indice de Jaccard a été introduit pour la première fois dans [110] pour mesurer la similarité entre deux ensembles. Il est défini comme suit :

**Définition 2.5** (Indice de Jaccard [110])**.** *Soient deux ensembles $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$. Alors l'indice de Jaccard entre $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ vaut*

$$J(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{\mathrm{Card}(\boldsymbol{\alpha}\cap\boldsymbol{\beta})}{\mathrm{Card}(\boldsymbol{\alpha}\cup\boldsymbol{\beta})}, \tag{2.57}$$

*où* Card *désigne l'opérateur de cardinalité. $J(\boldsymbol{\alpha},\boldsymbol{\beta})$ varie entre* 0*, lorsque $\boldsymbol{\alpha}\cap\boldsymbol{\beta}=\emptyset$, et* 1*, lorsque $\boldsymbol{\alpha}=\boldsymbol{\beta}$*

L'indice de Jaccard associé à 4 différents exemples est reporté dans la Table 2.1. Nous remarquons que l'indice de Jaccard est un score très sévère. En particulier, si $\boldsymbol{\beta}$ a correctement identifié la moitié des valeurs non nulles de $\boldsymbol{\alpha}$ et mal identifié l'autre moitié (voir cas $\boldsymbol{\beta}^{(2)}$), alors $J(\boldsymbol{\alpha},\boldsymbol{\beta})=1/3$.

Récemment, l'indice de Jaccard a connu un regain d'intérêt au sein de la communauté de traitement de graphes afin de mesurer la similarité entre deux noeuds $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ d'un graphe. En particulier, une extension de l'indice de Jaccard a été proposé dans [100] afin de mesurer la similarité entre deux noeuds d'un graphe pondéré :

**Définition 2.6** (Indice de Jaccard entre deux noeuds d'un graphe pondéré [100])**.** *Soient deux noeuds $\boldsymbol{\alpha}\in\mathcal{V}$ et $\boldsymbol{\beta}\in\mathcal{V}$ d'un graphe pondéré $G=(\mathcal{V},\mathcal{E})$ dont l'ensemble des noeuds et des arêtes sont respectivement notés $\mathcal{V}$ et $\mathcal{E}$. Alors l'indice de Jaccard entre $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$ vaut :*

$$J_w(\boldsymbol{\alpha},\boldsymbol{\beta}) = \frac{N(\boldsymbol{\alpha},\boldsymbol{\beta})}{D(\boldsymbol{\alpha},\boldsymbol{\beta})} \tag{2.58}$$

*où $N(\boldsymbol{\alpha},\boldsymbol{\beta})$ représente le poids total des arêtes $w\in\mathcal{E}$ reliant des noeuds $\boldsymbol{\gamma}\in\mathcal{V}$ communs à $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$, i.e.,*

$$N(\boldsymbol{\alpha},\boldsymbol{\beta}) = 2w_{\boldsymbol{\alpha},\boldsymbol{\beta}} + \sum_{\substack{\boldsymbol{\gamma}\in\mathcal{V}\\ \boldsymbol{\gamma}\sim\boldsymbol{\alpha}\\ \boldsymbol{\gamma}\sim\boldsymbol{\beta}}} \min(w_{\boldsymbol{\alpha},\boldsymbol{\gamma}}, w_{\boldsymbol{\beta},\boldsymbol{\gamma}}), \tag{2.59}$$

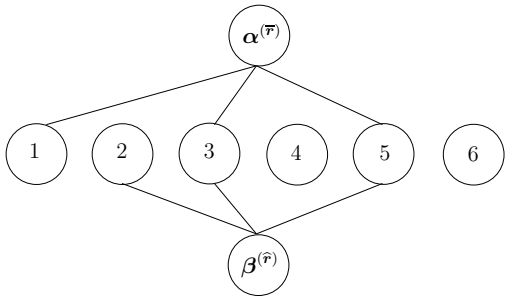



FIGURE 2.10 – **Exemple de graphe tripartite.** Une arête est créée entre chaque noeud $n$ et $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ (resp. $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$) si $\overline{r}_n > 0$ (resp. $\widehat{r}_n > 0$). Dans cet exemple, $\overline{\boldsymbol{r}} = (1,0,1,0,1,0)^\top$ et $\widehat{\boldsymbol{r}} = (0,1,1,0,1,0)^\top$.

*et où $D(\boldsymbol{\alpha},\boldsymbol{\beta})$ représente le poids total des arêtes reliant tous les noeuds voisins de $\boldsymbol{\alpha}$ et $\boldsymbol{\beta}$, i.e.,*

$$D(\boldsymbol{\alpha},\boldsymbol{\beta}) = 2w_{\boldsymbol{\alpha},\boldsymbol{\beta}} + \sum_{\substack{\boldsymbol{\gamma}\in\mathcal{V}\\ \boldsymbol{\gamma}\sim\boldsymbol{\alpha}\\ \boldsymbol{\gamma}\sim\boldsymbol{\beta}}} \frac{w_{\boldsymbol{\alpha},\boldsymbol{\gamma}} + w_{\boldsymbol{\beta},\boldsymbol{\gamma}}}{2} + \sum_{\substack{\boldsymbol{\gamma}\in\mathcal{V}\\ \boldsymbol{\gamma}\sim\boldsymbol{\alpha}\\ \boldsymbol{\gamma}\nsim\boldsymbol{\beta}}} w_{\boldsymbol{\alpha},\boldsymbol{\gamma}} + \sum_{\substack{\boldsymbol{\gamma}\in\mathcal{V}\\ \boldsymbol{\gamma}\nsim\boldsymbol{\alpha}\\ \boldsymbol{\gamma}\sim\boldsymbol{\beta}}} w_{\boldsymbol{\beta},\boldsymbol{\gamma}}. \tag{2.60}$$

### 2.3.2 Application à la détection de discontinuité

Mesurer la similarité entre les vecteurs d'indicatrices $\overline{\boldsymbol{r}} \in \mathbb{R}^{N-1}$ et $\widehat{\boldsymbol{r}} \in \mathbb{R}^{N-1}$ peut être interprété comme mesurer la similarité entre deux noeuds d'un graphe tripartite $G = (\mathcal{V},\mathcal{E})$. Nous présentons comment construire ce graphe illustré par la Figure 2.10.

**Ensemble des noeuds $\mathcal{V}$.** Le graphe est formé de $N+1$ noeuds composés des $(N-1)$ pas de temps et des deux noeuds $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ et $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$ dont nous cherchons à mesurer la similarité.

**Ensemble des arêtes $\mathcal{E}$.** Pour chaque indice $n \in \{1,\dots,N-1\}$ une arête de poids $\overline{r}_n$ (resp. $\widehat{r}_n$) est créée entre les noeuds $n$ et $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ (resp. $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$), i.e.,

$$w_{\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})},n} = \overline{r}_n \quad \left(\text{resp. } w_{\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})},n} = \widehat{r}_n\right). \tag{2.61}$$

Etant donné qu'il n'existe aucune arête reliant $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ et $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$, l'indice de Jaccard pondéré (défini en (2.58)) entre les noeuds $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ et $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$ peut se réécrire comme suit :

**Définition 2.7** (Indice de Jaccard entre $\overline{\boldsymbol{r}}$ et $\widehat{\boldsymbol{r}}$)**.** *Soient $\overline{\boldsymbol{r}} \in \mathbb{R}_+^{N-1}$ et $\widehat{\boldsymbol{r}} \in \mathbb{R}_+^{N-1}$ deux vecteurs à valeurs non négatives. Soient les deux noeuds $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ et $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$ du graphe pondéré tripartite $G = (\mathcal{V},\mathcal{E})$ défini précédemment. Alors l'indice de Jaccard entre $\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})}$ et $\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}$ vaut :*

$$J_w(\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})},\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}) = \frac{\sum_{1\le n\le N-1} \min(\overline{r}_n,\widehat{r}_n)}{\sum_{\substack{1\le n\le N-1\\ \overline{r}_n>0\\ \widehat{r}_n>0}} \frac{\overline{r}_n+\widehat{r}_n}{2} + \sum_{\substack{1\le n\le N-1\\ \widehat{r}_n=0}} \overline{r}_n + \sum_{\substack{1\le n\le N-1\\ \overline{r}_n=0}} \widehat{r}_n}. \tag{2.62}$$

*De même, nous définissons l'erreur de Jaccard : $J_{\text{error}}(\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})},\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}) = 1 - J_w(\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})},\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})})$.*

**Remarque 2.7.** *Dans la suite ce manuscrit, nous simplifierons la notation de l'indice de Jaccard comme suit :*

$$J_w(\overline{\boldsymbol{r}},\widehat{\boldsymbol{r}}) = J_w(\boldsymbol{\alpha}^{(\overline{\boldsymbol{r}})},\boldsymbol{\beta}^{(\widehat{\boldsymbol{r}})}). \tag{2.63}$$

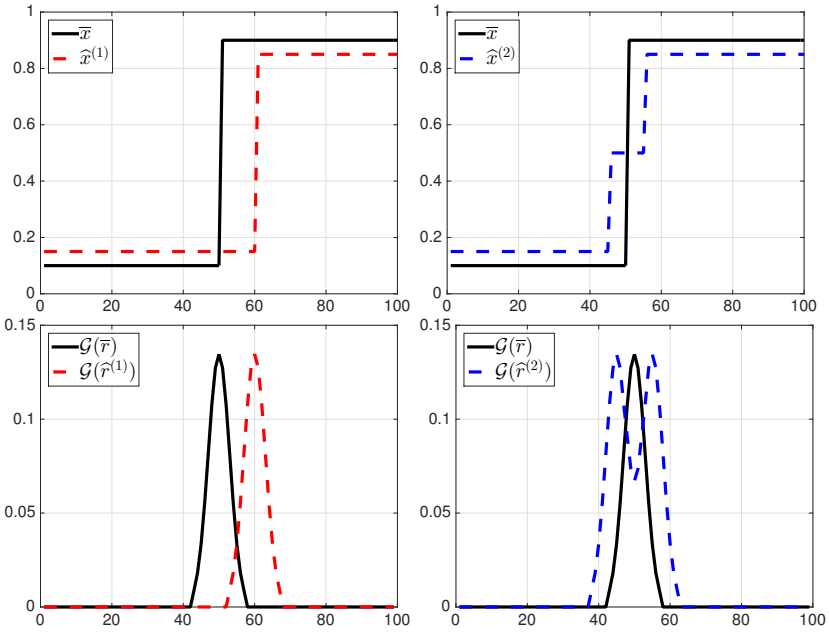



FIGURE 2.11 – **Exemples d'estimées constantes par morceaux**. Haut : Exemples d'estimées constantes par morceaux $\widehat{\boldsymbol{x}}^{(1)}$ et $\widehat{\boldsymbol{x}}^{(2)}$ de $\overline{\boldsymbol{x}}$. Bas : indicatrices des discontinuités correspondantes convoluées avec un filtre passe-bas gaussien à support compact $\mathcal{G}$ d'écart type 3 et de taille 15 points.

Afin de prendre en compte mais pénaliser les discontinuités de $\widehat{\boldsymbol{x}}$ proches de celles de $\overline{\boldsymbol{x}}$, nous proposons de mesurer la similarité non pas entre $\overline{\boldsymbol{r}} \in \{0,1\}^{N-1}$ et $\widehat{\boldsymbol{r}} \in \{0,1\}^{N-1}$, mais entre $\mathcal{G}(\overline{\boldsymbol{r}}) \in \mathbb{R}_+^{N-1}$ et $\mathcal{G}(\widehat{\boldsymbol{r}}) \in \mathbb{R}_+^{N-1}$ où $\mathcal{G}$ modélise un filtre passe-bas gaussien à support compact. Le choix du support et de l'écart-type de $\mathcal{G}$ est discuté dans la section suivante.

### 2.3.3 Choix du filtre passe-bas gaussien

Considérons l'exemple illustré par la Figure 2.11 (haut gauche) où $\widehat{\boldsymbol{x}}^{(1)}$ (rouge) est une estimée d'un signal constant par morceaux $\overline{\boldsymbol{x}}$ (noir) qui a mal localisé la seule discontinuité de $\overline{\boldsymbol{x}}$ d'une distance $d^{(1)} = 10$. Par conséquent, $J(\overline{\boldsymbol{r}}, \widehat{\boldsymbol{r}}^{(1)}) = 0$. Afin de quantifier cette imprécision autrement que par 0, nous considérons $\mathcal{G}(\widehat{\boldsymbol{r}}^{(1)})$ et $\mathcal{G}(\boldsymbol{r})$ rapportés Figure 2.11 (bas gauche) pour $\mathcal{G}$ de support $K = 15$ et d'écart type $\sigma = 4$. Pour de telles valeurs, $J_w(\mathcal{G}(\boldsymbol{r}), \mathcal{G}(\widehat{\boldsymbol{r}}^{(1)})) = 0.09$.

Une seconde estimée $\widehat{\boldsymbol{x}}^{(2)}$ (bleu) est également considérée et illustrée par la Figure 2.11 (haut droite). Elle contient deux discontinuités respectivement localisés à distance $d_1^{(2)} = 5$ et $d_2^{(2)} = 5$ de la véritable discontinuité. Nous noterons $d^{(2)} = d_1^{(2)} + d_2^{(2)}$. Cet exemple est motivé par l'effet d'escalier des solutions du Problème 2.2 (voir Remarque 2.3). Pour un filtre $\mathcal{G}$ identique et pour $d^{(1)} = d^{(2)} = d$, on obtient $J_w(\mathcal{G}(\boldsymbol{r}), \mathcal{G}(\widehat{\boldsymbol{r}}^{(2)})) = 0.56$.

Cependant, en pratique, on peut préférer la solution $\widehat{\boldsymbol{x}}^{(1)}$ à $\widehat{\boldsymbol{x}}^{(2)}$ puisqu'elle est composée d'une seule discontinuité. Dans la suite, nous examinons comment choisir $K$ et $\sigma$ afin de favoriser $\widehat{\boldsymbol{x}}^{(1)}$ ou $\widehat{\boldsymbol{x}}^{(2)}$.

Les quantités $J_1 = J_w(\mathcal{G}(\overline{\boldsymbol{r}}), \mathcal{G}(\widehat{\boldsymbol{r}}^{(1)}))$ et $J_2 = J_w(\mathcal{G}(\overline{\boldsymbol{r}}), \mathcal{G}(\widehat{\boldsymbol{r}}^{(2)}))$ sont rapportées dans la Figure 2.12 en fonction de $K/d$ et $\sigma/K$. Comme attendu, $J_1$ n'augmente pas selon $K/d$ et $\sigma/K$. Toutefois, la discussion est plus subtile pour $J_2$. En effet, en présence de discontinuités proches l'une de l'autre, tel que modélisé par $\widehat{\boldsymbol{r}}^{(2)}$, et pour $K$ et $\sigma$ assez grands, $|(\mathcal{G}(\widehat{\boldsymbol{r}}^{(2)}))_n -$

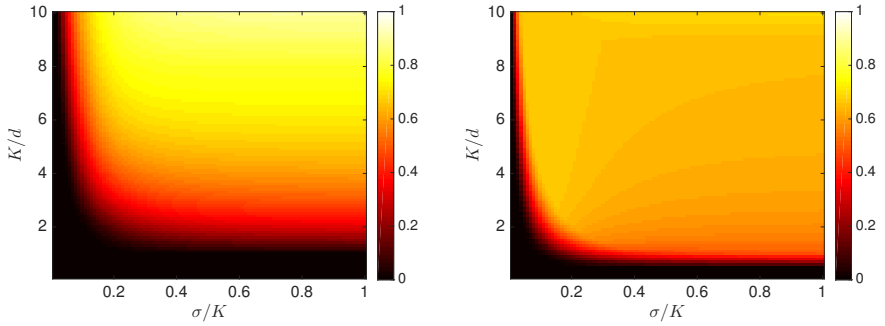



FIGURE 2.12 – **Indice de Jaccard pondéré.** $J_1$ (gauche) et $J_2$ (droite).

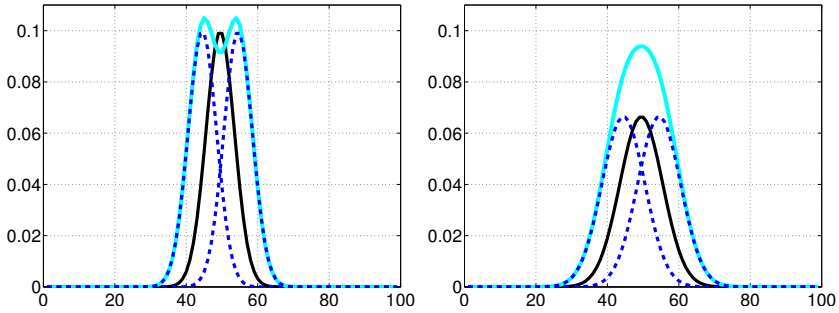



FIGURE 2.13 – **Impact du filtre passe-bas $\mathcal{G}$ sur deux discontinuités proches l'une de l'autre.** Le filtre passe-bas est gaussien à support compact de taille $K = 50$ et d'écart type $\sigma = 4$ (gauche) et $\sigma = 6$ (droite).

$(\mathcal{G}(\boldsymbol{r}))_n| \gg 0$ pour certains indices $n$ à cause de la convolution. La même conclusion est obtenue $K$ et $\sigma$ suffisamment faibles puisque $d_1^{(2)} > 0$ et $d_2^{(2)} > 0$. De plus, nous rappelons que d'après la Définition 2.7, plus $\alpha_n \neq \beta_n$ pour tout $n$, plus $J(\boldsymbol{\alpha}, \boldsymbol{\beta})$ est proche de 0. Il s'en suit que $J_2$ atteint sa plus grande valeur pour des régions intermédiaires $(K/d,\sigma/K)$. Par souci d'illustration, deux configurations sont rapportées par la Figure 2.13 pour $K$ fixé et $\sigma$ faible (gauche) et $\sigma$ grand (droite).

Le diagramme de phase indiquant les régions où $J_1 > J_2$ est illustré par la Figure 2.14 (gauche). Il est intéressant de remarquer qu'une condition nécessaire pour avoir $J_1 > J_2$ est de choisir $K$ au moins 4 fois supérieur à la distance $d$.

Pour finir, nous relâchons $d^{(1)} = d^{(2)}$ et examinons comment varie la frontière du diagramme de phase en fonction de $d^{(1)} - d^{(2)}$. En pratique, $d^{(2)} = 10$ et $d = d^{(1)}$ varie de 5 à 15. Les résultats correspondants sont rapportés par la Figure 2.14 (droite), et fournissent à l'utilisateur une façon de choisir $\sigma$ et $K$ selon la configuration qu'il souhaite privilégier.

## 2.4 Extension bidimensionnelle

Dans cette section, nous nous intéressons à la détection des discontinuités dans les images monovaluées. En d'autre termes, nous chercherons à partitionner une image en régions où l'intensité moyenne peut être considérée constante.

Dans la suite, nous considérons l'observation bidimensionnelle notée $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$, où $\Omega = \{1, \ldots, N_\mathrm{h}\} \times \{1, \ldots, N_\mathrm{v}\}$.

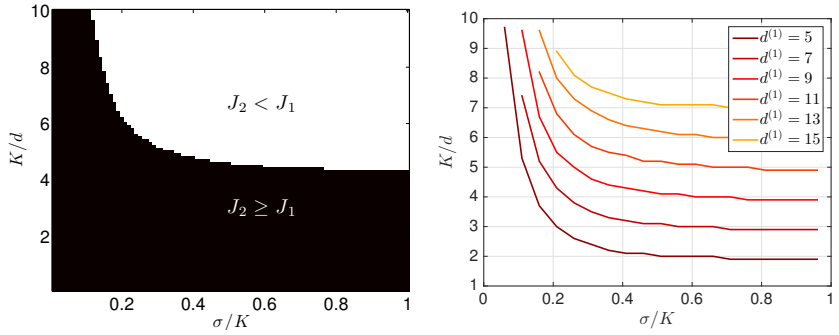



FIGURE 2.14 – **Diagramme de phase.** Comparaison entre $J_1$ et $J_2$ pour différentes valeurs de $K/d$ et $\sigma/K$..

### 2.4.1 Problème non convexe $\ell_0$-TV 2D

Le problème consistant à étiqueter une image $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$ en partition distinctes $\Omega = \bigcup_{q=1}^{+\infty} \Omega_q$ ayant chacune pour valeur moyenne $v_q$, peut être formulé via le problème suivant :

**Problème 2.7** (Problème $\ell_0$-TV 2D)**.** *Soit une image $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$, alors le problème $\ell_0$-TV 2D consiste à trouver $\boldsymbol{x}$ constant par morceaux, i.e., $\boldsymbol{x} = \sum_{q=1}^{+\infty} v_q \boldsymbol{i}_{\Omega_q}$ avec comme convention $v_q \leq v_{q+1}$, tel que*

$$\widehat{\boldsymbol{x}} \in \underset{\boldsymbol{x}=\sum_{q=1}^{+\infty} v_q \boldsymbol{i}_{\Omega_q} \in \mathbb{R}^{|\Omega|}}{\operatorname{Argmin}} \frac{1}{2} \sum_{q=1}^{+\infty} \boldsymbol{i}_{\Omega_q}^{\top} (\boldsymbol{y} - v_q)^2 + \lambda |\mathcal{C}| \qquad \textit{sujet à} \qquad \begin{cases} \bigcup_{q=1}^{+\infty} \Omega_q = \Omega, \\ (\forall q \neq p),\ \Omega_q \cap \Omega_p = \emptyset, \end{cases} \tag{2.64}$$

*où $\mathcal{C} = \Omega \left( \bigcap \bigcup_{q=1}^{+\infty} \partial \Omega_q \right)$ est l'ensemble des contours du partitionnement $\Omega = \bigcup_{q=1}^{\infty} \Omega_q$ défini par les vecteurs d'indicatrices*

$$(\forall q \geq 1) \quad (\forall \underline{\ell} \in \Omega) \quad \boldsymbol{i}_{\Omega_q, \underline{\ell}} = \begin{cases} 1 \ \textit{si} \ \underline{\ell} \in \Omega_q, \\ 0 \ \textit{sinon.} \end{cases} \tag{2.65}$$

**Remarque 2.8.** *Le Problème 2.7 est couramment appelé problème de Potts et problème de Mumford-Shah constant par morceaux dans le cadre continu.*

**Remarque 2.9.** *Si $\boldsymbol{y}$ peut être décrit par $Q$ lignes de niveau, alors pour tout $q \geq Q + 1$, $v_q = +\infty$ et $\boldsymbol{i}_{\Omega_q} = 0_\Omega$.*

Le critère présent dans (2.64) est composé de la somme d'un terme d'attache aux données quantifiant l'appartenance de chaque pixel $\underline{\ell}$ à une ligne de niveau $\Omega_q$ et d'un terme de pénalisation

$$|\mathcal{C}| = \frac{1}{2} \sum_{q=1}^{+\infty} |\partial \Omega_q| = \frac{1}{2} \sum_{q=1}^{+\infty} \operatorname{Per}(\Omega_q) \tag{2.66}$$

imposant aux partitions $\Omega_q$ d'avoir un périmètre $\operatorname{Per}(\Omega_q)$ minimal. Les contraintes assurent que les différentes partitions ne se chevauchent pas. L'hyperparamètre $\lambda$ permet de relativiser la contribution de ces deux termes. Choisir $\lambda$ élevé permet de favoriser la formation d'un petit nombre de large régions $\Omega_q$.

**Remarque 2.10.** *Le Problème 2.7 est une généralisation mutlidimensionnelle du Problème 2.1. En effet, nous pouvons considérer en première approximation que [133, 168]*

$$Per(\Omega_q) \approx TV(\boldsymbol{i}_{\Omega_q}), \tag{2.67}$$

*où TV désigne la variation totale bidimensionnelle définie par*

$$\mathrm{TV}(\boldsymbol{i}_{\Omega_q}) = \sum_{\underline{\ell}=1}^{|\Omega|} \sqrt{(D_{\mathrm{h}}\boldsymbol{i}_{\Omega_q})_{\underline{\ell}}^2 + (D_{\mathrm{v}}\boldsymbol{i}_{\Omega_q})_{\underline{\ell}}^2} = \sum_{\underline{\ell}=1}^{|\Omega|} \|(\mathrm{D}\boldsymbol{i}_{\Omega_q})_{\underline{\ell}}\|_2, \tag{2.68}$$

*où $D_{\mathrm{h}}$ (resp. $D_{\mathrm{v}}$) est l'opérateur discret de premières différences dans la direction horizontale (resp. verticale) et $\mathrm{D} = [D_{\mathrm{h}}, D_{\mathrm{v}}]$.*

**Remarque 2.11.** *La TV définie en* (2.68) *est largement utilisée et donne dans la pratique des résultats satisfaisants, mais elle souffre de plusieurs défauts : elle n'est pas rigoureusement isotropique et elle a tendance à lisser les contours. Le lecteur est invité à se référer à [54] pour une nouvelle définition de la TV isotropique pour les images discrètes.*

Cependant, le Problème 2.7 est NP difficile et ne peut pas être résolu de manière exacte. Afin de contourner cette limitation, des solutions approchées reposant sur l'utilisation d'algorithmes de « graph-cut » ont été proposées [79, 17]. Une partie de la littérature s'est également intéressée à la résolution exacte de deux problèmes approchés que nous détaillerons dans la suite : le premier étant une relaxation convexe du Problème 2.7 (voir Section 2.4.2) et le second étant un problème plus simple où le nombre de régions $Q$ est supposé connu *a priori* (voir Section 2.4.3).

### 2.4.2 Problème convexe $\ell_1$-TV 2D

Une relaxation convexe du Problème 2.7 consiste à estimer $\widehat{\boldsymbol{x}}$ comme suit :

**Problème 2.8** (Problème $\ell_1$-TV bidimensionnel)**.** *Soient $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$ et $\lambda \geq 0$. On appelle problème $\ell_1$-TV bidimensionnel le problème consistant à trouver*

$$\widehat{\boldsymbol{x}}_\lambda = \arg\min_{\boldsymbol{x}\in\mathbb{R}^{|\Omega|}} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|^2 + \lambda \mathrm{TV}(\boldsymbol{x}), \tag{2.69}$$

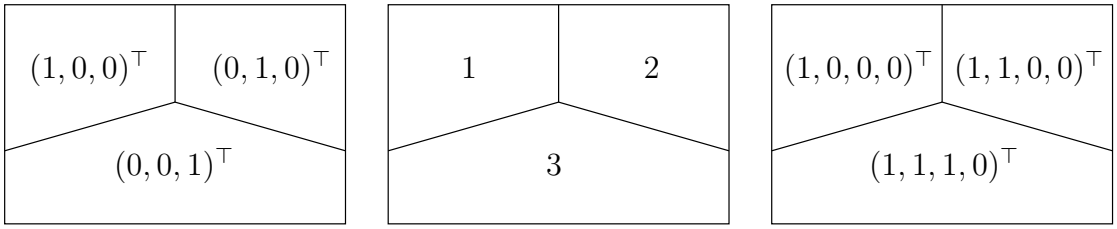
*où la TV bidimensionnelle est définie en* (2.68)*.*

Le Problème 2.8 apparaît donc comme une extension bidimensionnelle du Problème 2.2 où la seule différence notable réside dans le fait que la variation totale a été remplacée par son homologue bidimensionnel. Par conséquent, les algorithmes proximaux décrits à la Section 2.2 peuvent également être employés pour résoudre le Problème 2.8 (cf., e.g., [35, 23, 150]).

### 2.4.3 Problème convexe $\ell_0$-TV 2D à $Q$ niveaux

Afin de simplifier le Problème 2.7, nous pouvons également supposer que le nombre de partitions $Q$ et les valeurs des niveaux $(v_1, \ldots, v_Q)$ sont connus *a priori*. Ainsi, le Problème 2.7 se réduit au Problème de $\ell_0$-TV à $Q$ niveaux suivant

**Problème 2.9** (Problème $\ell_0$-TV 2D à $Q$ niveaux)**.** *Soit une image $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$ et un ensemble de valeurs moyennes $(v_q)_{1\leq q\leq Q}$ avec comme convention $v_q \leq v_{q+1}$. Alors le Problème $\ell_0$-TV*




(a) $\boldsymbol{i}_\Omega = (\boldsymbol{i}_{\Omega_1}, \ldots, \boldsymbol{i}_{\Omega_Q})$ (b) $\boldsymbol{u}$ (c) $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}})$

FIGURE 2.15 – **Illustration des trois variables d'étiquetage.** Dans cet exemple, les $Q = 3$ régions peuvent être étiquetées soit à partir de $\boldsymbol{i}_\Omega$ (a), $\boldsymbol{u}$ (b) ou $\boldsymbol{\theta}$ (c).

*à $Q$ niveaux consiste à trouver*

$$(\widehat{\boldsymbol{i}}_{\Omega_1}, \ldots, \widehat{\boldsymbol{i}}_{\Omega_Q}) = \underset{(\boldsymbol{i}_{\Omega_1}, \ldots, \boldsymbol{i}_{\Omega_Q})^\top \in \{0,1\}^{Q \times |\Omega|}}{\arg\min} \sum_{q=1}^{Q} \boldsymbol{i}_{\Omega_q}^\top (\boldsymbol{y} - v_q)^2 + \lambda \sum_{q=1}^{Q} TV(\boldsymbol{i}_{\Omega_q})$$
$$\text{sujet à} \quad \sum_{q=1}^{Q} \boldsymbol{i}_{\Omega_q} = \mathbb{1}_\Omega \quad (2.70)$$

Un exemple de vecteur $\boldsymbol{i}_\Omega$ est illustré Figure 2.15(a) pour $Q = 3$ régions, i.e., $\boldsymbol{i}_\Omega$ vaut $(1, 0, 0)$ dans la première région, $(0, 1, 0)$ dans la deuxième et $(0, 0, 1)$ dans la troisième.

Bien que le critère dans (2.70) soit convexe en $\boldsymbol{i}_\Omega$, le Problème reste non-convexe du à la contrainte et au caractère binaire de $\boldsymbol{i}_\Omega$. Afin de convexifier la contrainte, il a été proposé (cf. [39, 206] pour $Q = 2$ et [168] pour $Q \geq 2$) d'augmenter la dimension du problème en remplaçant la minimisation sur $\boldsymbol{i}_\Omega \in \{0, 1\}^{Q \times |\Omega|}$ par $\theta \in \{0, 1\}^{(Q+1) \times |\Omega|}$. La bijection entre ces deux problèmes, illustrée par la Figure 2.15, peut être interprétée via par l'intermédaire de l'étiquette $\boldsymbol{u}$ :

**Définition 2.8** (Etiquette $\boldsymbol{u}$)**.** *Soit le vecteur d'indicatrices $(\boldsymbol{i}_{\Omega_1}, \ldots, \boldsymbol{i}_{\Omega_Q})^\top \in \{0, 1\}^{Q \times |\Omega|}$ caractérisant le partionnement en $Q$ régions $(\Omega_1, \ldots, \Omega_Q)$ tels que $\bigcup_{q=1}^{Q} \Omega_q = \Omega$ et $(\forall q \neq p)$, $\Omega_q \cap \Omega_p = \emptyset$. On appelle étiquette $u \in \{1, \ldots, Q\}^{|\Omega|}$, la quantité*

$$\boldsymbol{u} = \sum_{q=1}^{Q} q \boldsymbol{i}_{\Omega_q}. \quad (2.71)$$

*Par conséquent, $(\forall \underline{\ell} \in \Omega)$, $u_{\underline{\ell}} = q$ si $\underline{\ell} \in \Omega_q$ et $0$ sinon.*

Dès lors, une bijection entre $\boldsymbol{u}$ défini sur un simplexe et des variables binaires $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}})$ peut être établie comme suit :

**Proposition 2.5** (Bijection entre $\boldsymbol{u}$ et $\theta$)**.** *Soit une étiquette $\boldsymbol{u} \in \{1, \ldots, Q\}^{|\Omega|}$ défini sur un simplexe de taille $Q$ et soit $Q + 1$ variables binaires $(\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}})^\top \in \{0, 1\}^{(Q+1) \times |\Omega|}$ telles que*

$$(\forall \underline{\ell} \in \Omega), \quad 1 = \theta_{1,\underline{\ell}} \geq \theta_{2,\underline{\ell}} \geq \ldots \geq \theta_{Q,\underline{\ell}} = 0. \quad (2.72)$$

*Alors il existe une bijection entre $\boldsymbol{u}$ et $\boldsymbol{\theta}$ :*

$$(\forall \underline{\ell} \in \Omega) \quad \theta_{q,\underline{\ell}} = \begin{cases} 1, & \text{si } u_{\underline{\ell}} \geq q, \\ 0, & \text{sinon,} \end{cases} \quad (2.73)$$

*qui peut être inversée en (cf. « layer cake formula »)*

$$(\forall \underline{\ell} \in \Omega) \quad u_{\underline{\ell}} = \sum_{q=1}^{Q} \theta_{q,\underline{\ell}}. \tag{2.74}$$

On remarque que le vecteur d'indicatrices $\boldsymbol{i}_\Omega$ peut être exprimé entièrement à partir de $\boldsymbol{\theta}$, i.e.,

**Proposition 2.6** (Correspondance entre $\boldsymbol{i}_\Omega$ et $\boldsymbol{\theta}$)**.** *Soit $\boldsymbol{i}_\Omega$ défini en* (2.65) *et $\boldsymbol{\theta}$ défini à la Proposition 2.5. Alors,*

$$(\forall q \in \{1, \ldots, Q\}) \quad \boldsymbol{i}_{\Omega_q} = (\boldsymbol{\theta_q} - \boldsymbol{\theta_{q+1}}). \tag{2.75}$$

Finalement, afin d'obtenir un problème convexe, on relâche la nature binaire de $\boldsymbol{\theta}$ et l'on suppose que $\boldsymbol{\theta}^\top \in [0,1]^{(Q+1)\times|\Omega|}$, i.e.,

**Problème 2.10** (Relaxation convexe du problème $\ell_0$-TV à $Q$ niveaux [37, 192])**.** *Soit une image $\boldsymbol{y} \in \mathbb{R}^{|\Omega|}$ et un ensemble de valeurs moyennes $(v_q)_{1\leq q\leq Q}$ avec comme convention $v_q \leq v_{q+1}$. Alors le problème consiste à trouver $\boldsymbol{\theta}^\top = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}})^\top \in [0,1]^{(Q+1)\times|\Omega|}$ tel que*

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}^\top \in [0,1]^{(Q+1)\times|\Omega|}}{\arg\min} \sum_{q=1}^{Q} (\boldsymbol{\theta_q} - \boldsymbol{\theta_{q+1}})^\top (\boldsymbol{y} - v_q)^2 + \lambda \sum_{q=1}^{Q} TV(\boldsymbol{\theta_q} - \boldsymbol{\theta_{q+1}})$$
$$\text{sujet à} \quad \begin{cases} \boldsymbol{\theta_1} \equiv 1, \\ \boldsymbol{\theta_{Q+1}} \equiv 0, \\ 1 \geq \boldsymbol{\theta_2} \geq \ldots \geq \boldsymbol{\theta_Q} \geq 0. \end{cases} \tag{2.76}$$

Le Problème 2.10 étant convexe, un algorithme proximal a été proposé en [168, 55] pour trouver efficacement le minimiseur global. Il a également été montré dans [168, Théorème 2] que quelque soit le seuil $\mu$, la solution seuillée, $\widehat{\boldsymbol{\theta}}^{\text{thr}} = 1$ si $\widehat{\boldsymbol{\theta}} \geq \mu$ et 0 sinon, définit une solution $\widehat{\boldsymbol{i}}_\Omega$ qui est un minimiseur du Problème 2.9.

## 2.5 Questions ouvertes

Dans ce second chapitre, nous avons présenté deux problèmes variationnels permettant d'estimer les discontinuités dans un signal constant par morceaux corrompu par un bruit additif gaussien. Résoudre le Problème 2.1 (non-convexe) permet d'obtenir une solution non-biaisée alors que la solution du Problème 2.2 (convexe) est biaisée due à l'utilisation de la norme $\ell_1$. La résolution de chacun des problèmes nécessite de minimiser la somme d'un terme quadratique d'attache aux données et d'un terme de régularisation ($\ell_0 \circ L$ ou $\ell_1 \circ L$). Le poids accordé à chacun des termes est contrôlé par un hyperparamètre ($\chi$ ou $\lambda$) qui impacte fortement la qualité de la solution. Une attention plus spécifique a été portée sur le Problème 2.2 pour lequel nous avons comparé différentes solutions algorithmiques en termes de temps d'éxécution. Nous avons également introduit une mesure basée sur l'indice de Jaccard pondéré afin de quantifier la qualité d'estimation d'un signal constant par morceaux. Pour terminer, nous avons présenté le Problème 2.10, en guise d'extension 2D du Problème 2.1, dont la résolution permet de segmenter une image en $Q$ classes.

Pour l'instant, nous avons seulement considéré le cas où $\boldsymbol{y}$ est une quantité scalaire, mais nous pouvons également étendre la définition et la résolution du Problème 2.2 et du Problème 2.10 lorsque $\boldsymbol{y}$ est remplacé par une quantité vectorielle $\mathbf{y} = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_M})^\top$.

Le Chapitre 3 sera dédié à l'application des méthodes d'estimation et de segmentation vectorielle pour les problèmes posés par l'analyse non homogène des processus multifractals.

Motivé par [52], nous proposerons dans le Chapitre 4 un algorithme à la volée permettant d'estimer les discontinuités présentes dans des signaux vectoriels au fur et à mesure de leur acquisition.

Dans le Chapitre 7, nous examinerons un cas pratique motivé par l'expérience de scintillement de boîtes quantiques où la question consiste cette fois-ci à détecter les changements d'intensité d'un processus poissonien. Nous y détaillerons comment prendre en compte cette information dans le problème variationnel et dans quelles mesures la solution du Problème 2.2 est acceptable.

Enfin, nous nous intéresserons dans le Chapitre 8 au choix du paramètre de régularisation dans le cadre du Problème 2.1 et proposerons un modèle hybride bayésien-variationnel pour l'estimation de $\chi$ de manière automatique.

# Deuxième partie

# Estimation constante par morceaux vectorielle

# Chapitre 3

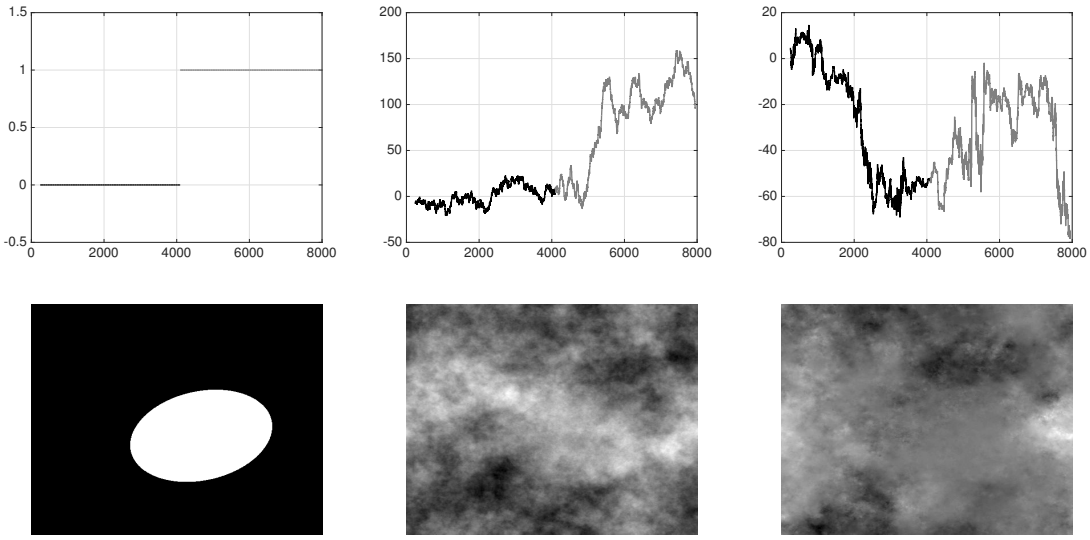
# Analyse multifractale inhomogène

**Sommaire**


**3.1 Modèle de processus multifractals par morceaux . . . . . 56**
3.1.1 Processus monofractal par morceaux . . . . . 56
3.1.2 Processus multifractal par morceaux . . . . . 57
**3.2 Formalisme multifractal local . . . . . 57**
**3.3 Estimation vectorielle par approche TV (1D) . . . . . 58**
3.3.1 Estimation vectorielle . . . . . 58
3.3.1.1 Présentation et résolution du problème $\ell_{1,2}$-TV . . . . . 58
3.3.1.2 Apport de l'estimation vectorielle . . . . . 61
3.3.2 Quantification des performances . . . . . 62
**3.4 Segmentation vectorielle par approche TV (2D) . . . . . 66**
3.4.1 Segmentation vectorielle . . . . . 66
3.4.2 Quantification des performances . . . . . 68
**3.5 Conclusion et perspectives . . . . . 71**


**Travaux préliminaires : processus monofractal par morceaux.** Comme rappelé au Chapitre 1, les propriétés d'invariance d'échelle de tout objet $\boldsymbol{z}$ sont quantifiées par la régularité locale, mesurée en pratique par le concept d'exposant de Hölder [114]. Alors que les premières études [181] supposaient que la régularité locale était homogène à travers le signal ou l'image (i.e., monofractal), de récentes contributions ont considéré un modèle plus réaliste de processus monofractal par morceaux où la régularité locale peut être hétérogène. Ce modèle complique le procédé d'estimation puisqu'il nécessite en plus de segmenter $\boldsymbol{z}$ en régions *a priori* inconnues où la régularité locale peut être considérée homogène. A cet effet, des approches variationnelles efficaces reposant sur l'utilisation de la variation totale ont récemment été mises en oeuvre dans [176, 158]. Les bonnes performances d'estimation et de segmentation scalaire, à la fois sur des données simulées et des données réelles, ouvrent la voie vers l'analyse de modèles d'invariance d'échelle constants par morceaux plus riches. C'est le sujet de ce chapitre.

**Contribution : processus multifractal par morceaux.** Dans [176, 158], les résultats étaient obtenus pour l'analyse de processus monofractals par morceaux. Ici, nous suggérons d'étendre cette étude au cas des processus multifractals par morceaux $\boldsymbol{z} \in \mathbb{R}^{|\Omega|}$, pouvant être unidimensionnels (signaux) ou bidimensionels (images), dont les propriétés multifractales ne sont pas homogènes sur $\Omega$ mais homogènes à l'intérieur de plusieurs régions distinctes $\Omega_1 \cap \ldots \cap \Omega_Q = \emptyset$ formant un partitionnement de $\Omega = \Omega_1 \cup \ldots \cup \Omega_Q$. L'objectif est d'estimer



(a) Masque (b) $\boldsymbol{z}$ monofractal par morceaux (c) $\boldsymbol{z}$ multifractal par morceaux

FIGURE 3.1 – **Exemples de processus invariants d'échelle non homogènes.** Deux processus, l'un monofractal par morceaux (b) et l'autre multifractal par morceaux (c) sont caractérisés par un spectre multifractal dans la région $\Omega_A$ (blanc) différent de celui dans la région $\Omega_B$ (noir).

à la fois les régions $(\Omega_1, \ldots, \Omega_Q)$ et le spectre multifractal (voir Section 1.2.2) à l'intérieur de chacune d'entre d'elles.

Ce problème est significativement plus difficile étant donné que pour un processus multifractal dont les propriétés multifractales sont constantes par morceaux, la régularité locale peut varier significativement d'une position à l'autre au sein d'une même région. Au lieu de considérer la régularité, nous nous intéresserons aux quantités vectorielles $(C_{1,j})_{j\in\mathbb{Z}}$, $(C_{2,j})_{j\in\mathbb{Z}}$, $c_1$ et $c_2$ permettant de définir le spectre multifractal (voir Chapitre 1). Le modèle de processus multifractal par morceaux est décrit dans la Section 3.1 et la définition des estimateurs locaux des quantités $(C_{1,j})_{j\in\mathbb{Z}}$, $(C_{2,j})_{j\in\mathbb{Z}}$, $c_1$ et $c_2$ fait l'objet de la Section 3.2.

Nous proposons ensuite deux applications, l'une unidimensionnelle appliquée à l'estimation vectorielle des discontinuités de $(c_1, c_2)$ et décrite à la Section 3.3 et l'autre bidimensionnelle reposant sur la segmentation de $(C_{1,j}, C_{2,j})_{j\in\mathbb{Z}}$ détaillée à la Section 3.4.

**Références.** Ce chapitre s'appuie sur les contributions [90, 91, 92] et est le résultat d'une collaboration avec H. Wendt.

## 3.1 Modèle de processus multifractals par morceaux

### 3.1.1 Processus monofractal par morceaux

Nous considérons le cas où $\boldsymbol{z}$ est un processus autosimilaire par morceaux caractérisé par l'exposant de Hurst $H_q$ sur chaque région $\Omega_q$ ($\forall q \in \{1, \ldots, Q\}$). Le modèle que nous adopterons dans la suite est illustré par un exemple de signal et de texture sur la Figure 3.1 (b)

pour $Q = 2$. Dans ce cas, l'exposant de Hölder $h$ vaut

$$(\forall \underline{\ell} \in \Omega_A) \;\; h_{\underline{\ell}} = H_A, \tag{3.1}$$
$$(\forall \underline{\ell} \in \Omega_B) \;\; h_{\underline{\ell}} = H_B, \tag{3.2}$$

avec $H_A \neq H_B$ et $\Omega = \Omega_A \cup \Omega_B$ séparable en deux régions distinctes, i.e., $\Omega_A \cap \Omega_B = \emptyset$, selon le masque présenté dans la Figure 3.1 (a). Dans l'exemple proposé, $H_A = 0.6$ et $H_B = 0.4$. Un comportement plus lisse est alors observé pour $\underline{\ell} \in \Omega_A$ que pour $\underline{\ell} \in \Omega_B$.

### 3.1.2 Processus multifractal par morceaux

Considérons à présent le cas illustré par la Figure 3.1 (c) où $\boldsymbol{z}$ modélise un processus multifractal par morceaux où

$$(\forall \underline{\ell} \in \Omega_A) \;\; h_{\underline{\ell}} \text{ est décrit par } \mathcal{D}_A, \tag{3.3}$$
$$(\forall \underline{\ell} \in \Omega_B) \;\; h_{\underline{\ell}} \text{ est décrit par } \mathcal{D}_B. \tag{3.4}$$

Dans la suite, pour des raisons pratiques, nous supposerons que le spectre multifractal peut être approché par une parabole (voir Proposition 1.2)

$$\mathcal{D}(h) = 2 + (h - c_1)^2/(2c_2). \tag{3.5}$$

Dans l'exemple présenté, $(c_1, c_2)_A = (0.6, -0.05)$ et $(c_1, c_2)_B = (0.4, -0.05)$. Nous remarquons qu'il est bien plus difficile de distinguer visuellement les deux régions $\Omega_A$ et $\Omega_B$ que dans le cas monofractal par morceaux. Cela est dû au fait que la régularité locale peut fluctuer au sein d'une même région.

## 3.2 Formalisme multifractal local

Dans cette section, nous rappelons comment établir des estimateurs locaux de $h$ et $\mathcal{D}$ à partir du formalisme multifractal basé sur les coefficients dominants décrit à la Section 1.3.

Dans la suite, nous supposons que $\boldsymbol{z}$ est invariant d'échelle sur la gamme $\{j_1, \ldots, j_2\}$ de taille $J = j_2 - j_1 + 1$ et nous notons $L_{j,\underline{k}}$ le coefficient d'ondelette dominant de $\boldsymbol{z}$ à l'échelle $2^j$ et à la position $\underline{\ell} = 2^j \underline{k}$ (voir Définition 1.10).

**Définition 3.1** (Estimateur local $\widehat{\boldsymbol{h}}$ de l'exposant de Hölder)**.** *L'exposant de Hölder peut être estimé par régression linéaire des coefficients d'ondelettes dominants, i.e.,*

$$(\forall \underline{\ell} \in \Omega) \quad \widehat{h}_{\underline{\ell}} = \sum_{j \in \{j_1, \ldots, j_2\}} w_{j,\underline{k}} \log_2 L_{j,\underline{k}}, \quad \textit{tel que} \quad \underline{\ell} = 2^j \underline{k} \tag{3.6}$$

*et où $w \in [0,1]^{J \times |\Omega|}$ désigne des poids de régression vérifiant*

$$\begin{cases} \sum_{j=j_1}^{j_2} j w_{j,\cdot} &= 1, \\ \sum_{j=j_1}^{j_2} w_{j,\cdot} &= 0, \end{cases} \tag{3.7}$$

*afin d'obtenir un estimateur non biaisé [213].*

Afin d'estimer le spectre multifractal $\mathcal{D}$, l'estimation strictement ponctuelle de $h$ doit être remplacée par une estimation locale de $C_1$ et $C_2$ dans un voisinage $\mathcal{W}$.

**Définition 3.2** (Estimateurs locaux $\widehat{\boldsymbol{C}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{C}}_{\mathbf{2}}$ des cumulants.)**.** *Les quantités $C_1$ et $C_2$ sont estimées localement par des moyennes en temps dans un voisinage glissant :*

$$(\forall j \in \{j_1, \dots, j_2\})\ (\forall \underline{\ell} \in \Omega) \quad \widehat{C}_{1,j,\underline{\ell}} = \frac{1}{|\mathcal{W}_{j,\underline{\ell}}|} \sum_{\underline{k} \in \mathcal{W}_{j,\underline{\ell}}} \ln L_{j,\underline{k}}, \tag{3.8}$$

$$(\forall j \in \{j_1, \dots, j_2\})\ (\forall \underline{\ell} \in \Omega) \quad \widehat{C}_{2,j,\underline{\ell}} = \frac{1}{|\mathcal{W}_{j,\underline{\ell}}| - 1} \sum_{\underline{k} \in \mathcal{W}_{j,\underline{\ell}}} \left( \ln L_{j,\underline{k}} - \widehat{C}_{1,j,\underline{\ell}} \right)^2, \tag{3.9}$$

*où $|\mathcal{W}_{j,\underline{\ell}}|$ est le nombre de coefficients dans le voisinage $\mathcal{W}_{j,\underline{\ell}}$ centré en $\underline{\ell}$ à l'échelle $2^j$.*

A partir de la Proposition 1.3, des estimateurs locaux de $c_1$ et $c_2$ peuvent ensuite être obtenus par régression linéaire à travers les échelles.

**Définition 3.3** (Estimateurs locaux $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$ des log-cumulants.)**.** *Soient les estimateurs locaux empiriques des cumulants $\widehat{\boldsymbol{C}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{C}}_{\mathbf{2}}$. Alors $c_1$ et $c_2$ peuvent être estimés localement comme suit*

$$(\forall p \in \{1, 2\})\ (\forall \underline{\ell} \in \Omega) \quad \widehat{c}_{p,\underline{\ell}} = (\log_2 \mathrm{e}) \sum_{j=j_1}^{j_2} w_j \widehat{C}_{p,j,\underline{\ell}}, \tag{3.10}$$

*où les poids de régression $w_j$ vérifient* (3.7)*.*

L'estimation de $C_1$, $C_2$, $c_1$ et $c_2$ est d'autant meilleure que la taille du voisinage $|\mathcal{W}_{j,\ell}|$ augmente, mais une mesure glissante en temps reposant sur un grand voisinage produit des estimées fortement dépendantes, compliquant ainsi la localisation fine des discontinuités.

## 3.3 Estimation vectorielle par approche TV (1D)

Dans cette section, nous allons nous intéresser au cas unidimensionel où $\boldsymbol{z} \in \mathbb{R}^N$ est un signal invariant d'échelle dont les propriétés multifractales sont caractérisées localement par un ensemble de $M$ signaux notés $\mathbf{y} = (\boldsymbol{y_m})_{1 \leq m \leq M}$ pouvant par exemple modéliser les quantités $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})^\top$, i.e., $M = 2$, décrites précédemment.

La question posée est alors de pouvoir détecter les portions où les propriétés multifractales de $\boldsymbol{z}$ sont homogènes en estimant les discontinuités présentes dans l'observation vectorielle $\mathbf{y}$. Pour y parvenir, nous présentons et quantifions dans la Section 3.3.1 l'apport d'une extension vectorielle du Problème 2.2 sur des données synthétiques. Puis, les performances d'estimation sont évaluées dans la Section 3.3.2 dans le cas d'une marche aléatoire multifractale à partir de $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})^\top$.

### 3.3.1 Estimation vectorielle

#### 3.3.1.1 Présentation et résolution du problème $\ell_{1,2}$-TV

Considérons une observation $\mathbf{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_M})^\top \in \mathbb{R}^{M \times N}$ d'un signal constant par morceaux $\overline{\mathbf{x}} = (\overline{\boldsymbol{x}}_{\mathbf{1}}, \dots, \overline{\boldsymbol{x}}_{\boldsymbol{M}})^\top \in \mathbb{R}^{M \times N}$ corrompu par un bruit additif $\mathbf{b} \in \mathbb{R}^{M \times N}$ supposé blanc gaussien centré de variance $\mathrm{diag}(\sigma_1^2, \dots, \sigma_M^2)$.

Par analogie avec le Problème 2.2, nous pouvons obtenir une solution constante par morceaux en résolvant par exemple le problème variationnel suivant

$$\widehat{\mathbf{x}}_\lambda = \underset{\mathbf{x} = (\boldsymbol{x_1}, \dots, \boldsymbol{x_M})^\top}{\arg\min} \ \frac{1}{2} \sum_{m=1}^{M} \|\boldsymbol{x_m} - \boldsymbol{y_m}\|^2 + \lambda \sum_{m=1}^{M} \sum_{n=1}^{N-1} |(L\boldsymbol{x_m})_n|. \tag{3.11}$$

Puisque (3.11) est séparable en $m \in \{1, \dots, M\}$, résoudre (3.11) revient à résoudre le Problème 2.2 indépendamment pour chacune des observation $\boldsymbol{y_m}$ ($m \in \{1, \dots, M\}$) pour le même paramètre de régularisation $\lambda$.

En pratique, la notation vectorielle $\overline{\boldsymbol{x}} = (\overline{\boldsymbol{x}}_{\mathbf{1}}, \dots, \overline{\boldsymbol{x}}_{\boldsymbol{M}})^\top$ est utilisée pour suggérer une dépendance entre les composantes. Nous supposerons ici que $\overline{\boldsymbol{x}}_{\mathbf{1}}, \dots, \overline{\boldsymbol{x}}_{\boldsymbol{M}}$ sont des signaux constants par morceaux partageant les mêmes positions de discontinuité. Afin de prendre en compte cette information supplémentaire, nous considérons le problème suivant :

**Problème 3.1** (Problème $\ell_{1,2}$-TV)**.** *Soient* $\mathbf{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_M})^\top \in \mathbb{R}^{M \times N}$ *et* $\lambda \geq 0$*. On appelle problème* $\ell_{1,2}$*-TV le problème consistant à trouver*

$$\widehat{\mathbf{x}}_\lambda = \underset{\mathbf{x}=(\boldsymbol{x_1}, \dots, \boldsymbol{x_M})^\top}{\arg\min} \frac{1}{2} \sum_{m=1}^{M} \|\boldsymbol{x_m} - \boldsymbol{y_m}\|^2 + \lambda \sum_{n=1}^{N-1} \sqrt{\sum_{m=1}^{M} |(L\boldsymbol{x_m})_n|^2}, \tag{3.12}$$

*où* $L$ *est l'opérateur linéaire de premières différences défini en* (2.3)*.*

**Remarque 3.1.** *Lorsque* $M = 1$*, le Problème 3.1 se réduit au Problème 2.2.*

Le second terme dans le membre de droite de (3.12) est la norme mixte $\ell_{1,2}$ de $L\mathbf{x}$, dont nous rappelons ci-dessous la définition générale.

**Définition 3.4** (Norme mixte $\ell_{p,q}$ [24])**.** *Soient* $\mathbf{u} = (u_{m,n})_{1 \leq m \leq M, 1 \leq n \leq N} \in \mathbb{R}^{M \times N}$*,* $p \geq 1$ *et* $q \geq 1$*. On appelle norme mixte* $\ell_{p,q}$ *de* $\mathbf{u}$*, la quantité définie par*

$$\|\mathbf{u}\|_{p,q} = \left( \sum_{n=1}^{N} \left( \sum_{m=1}^{M} |u_{m,n}|^q \right)^{p/q} \right)^{1/p}. \tag{3.13}$$

L'utilisation de la norme mixte permet de considérer une structure par blocs, où la norme $\ell_2$ est appliquée à chaque bloc, i.e., $\|L\boldsymbol{x_1}\|_2, \dots, \|L\boldsymbol{x_M}\|_2$, et où la norme $\ell_1$ est ensuite appliquée selon la direction des blocs. A partir des conditions de Karush-Kuhn-Tucker, nous rappelerons dans la Section 4.2 que l'emploi de la norme $\ell_{1,2}$ permet d'assurer une parcimonie de groupe sur $(L\widehat{\boldsymbol{x}}_{\mathbf{1}}, \dots, L\widehat{\boldsymbol{x}}_{\boldsymbol{M}})^\top$. Ce résultat est d'ailleurs illustré Figure 3.2 où pour une réalisation de $\overline{\mathbf{x}}$ (noir) et $\mathbf{y}$ (gris) donnée ($N = 500$, $M = 4$), la solution $\widehat{\mathbf{x}}_\lambda$ du Problème 3.1 est calculée et représentée pour $\lambda = 10$ (rouge), 50 (jaune) et 100 (violet). Quelque soit $\lambda$, $\widehat{\mathbf{x}}_\lambda$ partage les mêmes positions de discontinuités sur chacune de ses quatre composantes tracées sur la Figure 3.2 de haut en bas.

La forte analogie entre le Problème 2.2 et le Problème 3.1 nous permet de facilement étendre les algorithmes présentés à la Section 2.2 pour résoudre le Problème 3.1. En effet, il suffit de substistuer les quantités scalaires dans les Algorithmes 2-4 par leur homologue vectoriel, e.g. $\boldsymbol{y}$ et $\boldsymbol{x}$ remplacés par $\mathbf{y}$ et $\mathbf{x}$, et de remplacer toute occurence de $L$ par L tel que $\mathrm{L}\mathbf{x} = (L\boldsymbol{x_1}, \dots, L\boldsymbol{x_M})^\top$ et $\mathrm{prox}_{\gamma\|\cdot\|_1} = \mathrm{soft}_\gamma$ par $\mathrm{prox}_{\gamma\|\cdot\|_{1,2}}$ dont la définition est rappelée ci-dessous :

**Exemple 3.1** (Opérateur proximal de la norme $\ell_{1,2}$ [122, Théorème 3])**.** *Soit* $\gamma \in ]0, +\infty[$ *et une fonction* $\mathbb{R}^{M \times N} \to \mathbb{R} : \mathbf{u} \mapsto \gamma\|\mathbf{u}\|_{1,2}$*. On a alors, pour tout* $\mathbf{u} \in \mathbb{R}^{M \times N}$*,* $\mathrm{prox}_{\gamma\|\cdot\|_{1,2}}(\mathbf{u}) = (\zeta_n)_{1 \leq n \leq N}$ *avec*

$$(\forall n \in \{1, \dots, N\}), \quad \zeta_n = \begin{cases} (1 - \frac{\gamma}{\|\mathbf{u}_n\|_2})\mathbf{u}_n & \textit{si} \quad \|\mathbf{u}_n\|_2 > \gamma, \\ 0 & \textit{sinon.} \end{cases} \tag{3.14}$$

En particulier, nous détaillons dans l'Algorithme 5 les itérations de l'algorithme ADMM pour résoudre le Problème 3.1 proposé dans [209].

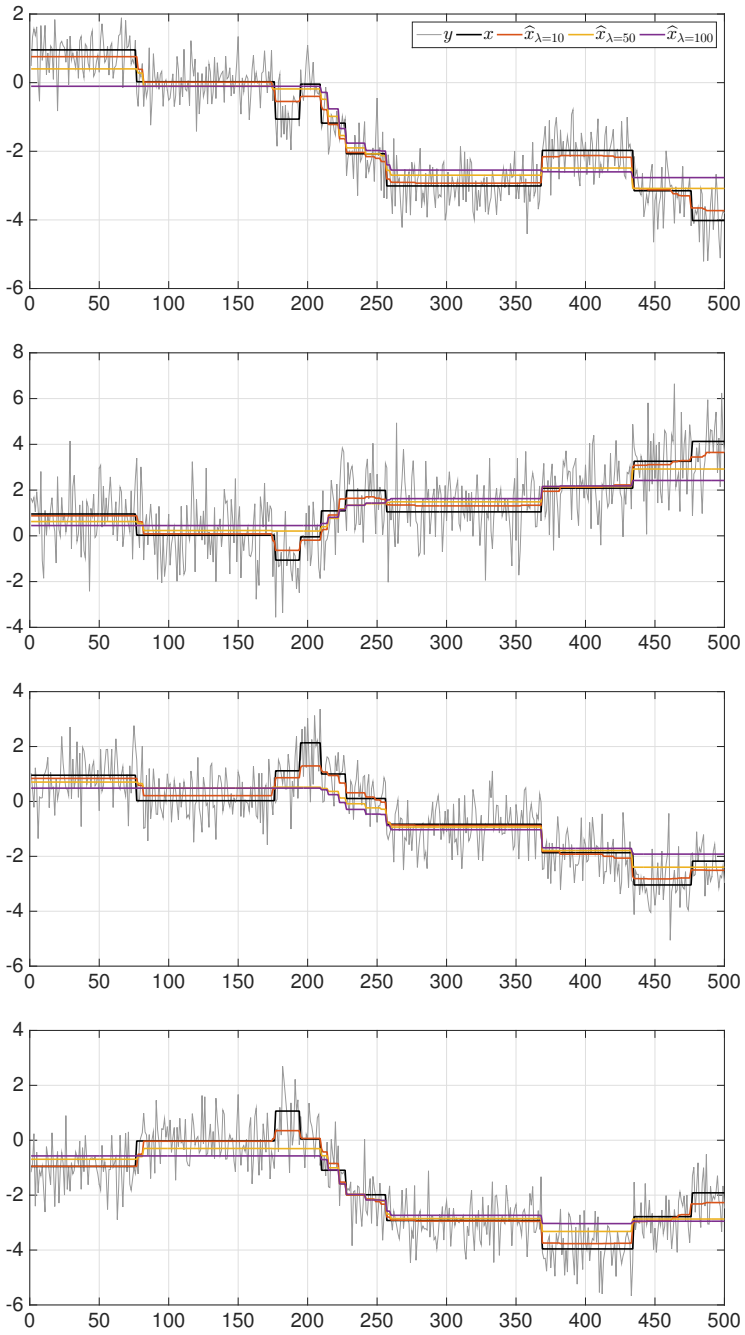



FIGURE 3.2 – **Illustration de $\widehat{\mathbf{x}}_\lambda$ solution du Problème 3.1.** $\overline{\mathbf{x}}$ (noir), $\mathbf{y}$ (gris) et $\widehat{\mathbf{x}}_\lambda$ pour $\lambda = 10$ (rouge), 50 (jaune) et 100 (violet). De haut en bas : 1ère jusqu'à la $M = 4$ème composante.

**Algorithme 5** Résolution du Problème 3.1 basée sur l'algorithme ADMM

**Entrée :** Paramètre $\rho > 0$.

**Pour** k=1,... **faire**

  **Mise à jour des variables primales :**

  $\mathbf{x}^{[k+1]} = \frac{\mathbf{y}+\rho(\mathbf{z}^{[k]}-\mathbf{u}^{[k]})}{1+\rho}$

  $\mathbf{z}^{[k+1]} = \text{prox}_{\frac{\lambda}{\rho}\|\cdot\|_{1,2}}(\mathbf{s}^{[k]} - \mathbf{t}^{[k]})$

  **Projection sur l'ensemble des contraintes :**

  $\left(\mathbf{r}^{[k+1]}, \mathbf{s}^{[k+1]}\right) = P_{\text{Ker}A}\left(\mathbf{x}^{[k+1]} + \mathbf{u}^{[k]}, \mathbf{z}^{[k+1]} + \mathbf{t}^{[k]}\right)$

  **Mise à jour des variables duales :**

  $\mathbf{u}^{[k+1]} = \mathbf{u}^{[k]} + (\mathbf{x}^{[k+1]} - \mathbf{r}^{[k+1]})$

  $\mathbf{t}^{[k+1]} = \mathbf{t}^{[k]} + (\mathbf{z}^{[k+1]} - \mathbf{s}^{[k+1]})$

**Sortie :** $(\widehat{\mathbf{x}}_\lambda, \widehat{\mathbf{u}}_\lambda) = \lim_{k\to+\infty}(\mathbf{x}^{[k]}, \mathbf{u}^{[k]})$

#### 3.3.1.2 Apport de l'estimation vectorielle

Dans cette section, nous examinons les configurations où la solution du Problème 3.1 appliquée à $\mathbf{y} = (\boldsymbol{y_1}, \ldots, \boldsymbol{y_M})^\top$ présente de meilleures performances d'estimation que les solutions obtenues en résolvant le Problème 2.2 indépendamment pour chacune des composantes $\boldsymbol{y_m}$ ($m \in \{1, \ldots, M\}$). Par souci de présentation des résultats, nous limitons l'étude au cas $M = 2$ choisi comme le plus simple représentatif de $M > 0$ et utilisé dans la section suivante pour l'analyse de $(\widehat{\boldsymbol{c}}_1, \widehat{\boldsymbol{c}}_2)^\top$.

**Configuration expérimentale.** Premièrement, on génère $\overline{\mathbf{x}} = (\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2)^\top \in \mathbb{R}^{2\times N}$ ($N = 10^3$) constant par morceaux tel que $\overline{\boldsymbol{x}}_1$ et $\overline{\boldsymbol{x}}_2$ partagent les mêmes discontinuités. Pour se faire, les longueurs de chaque segment sont tirées selon une loi de Pareto de paramètre d'échelle 50 et de paramètre de forme 1. Par conséquent, l'indicatrice des discontinuités $\overline{\mathbf{r}}$ de $\overline{\mathbf{x}}$ est égale à celle de $\overline{\boldsymbol{x}}_1$ et $\overline{\boldsymbol{x}}_2$ (voir Définition 2.4). Ensuite, les amplitudes absolues des discontinuités de $\overline{\boldsymbol{x}}_1$ et $\overline{\boldsymbol{x}}_2$ sont tirées indépendamment selon une gaussienne de moyenne 1 et de faible dispersion, e.g. 0.1. Le signe de chaque amplitude est ensuite tiré de manière équiprobable.

Deuxièmement, on génère $\mathbf{y} = (\boldsymbol{y}_1, \boldsymbol{y}_2)^\top$ tel que $\boldsymbol{y}_1 = \overline{\boldsymbol{x}}_1 + \boldsymbol{b}_1$ et $\boldsymbol{y}_2 = \overline{\boldsymbol{x}}_2 + \boldsymbol{b}_2$ avec $\boldsymbol{b}_1 \sim \mathcal{N}(0, \sigma_1 \mathbb{1}_N)$ et $\boldsymbol{b}_2 \sim \mathcal{N}(0, \sigma_2 \mathbb{1}_N)$. Puisque les amplitudes des discontinuités de $\overline{\boldsymbol{x}}_1$ et $\overline{\boldsymbol{x}}_2$ varient légèrement autour de 1, on définit les rapports signal sur bruit $\text{SNR}_1 = 1/\sigma_1$ et $\text{SNR}_2 = 1/\sigma_2$. Pour une même réalisation de $(\overline{\boldsymbol{x}}_1, \overline{\boldsymbol{x}}_2, \boldsymbol{b}_1, \boldsymbol{b}_2)$, différents $\text{SNR}_1$ et $\text{SNR}_2$ allant de $-10$ dB à $+10$ dB sont examinés en multipliant $\boldsymbol{b}_1$ et $\boldsymbol{b}_2$ par un facteur d'échelle.

**Critère de performance.** Nous cherchons à connaître le régime $(\text{SNR}_1, \text{SNR}_2)$, dans lequel la solution du Problème 3.1 appliqué à $\mathbf{y}$ fournie une meilleure estimation des positions des discontinuités de $\overline{\mathbf{x}}$, que les deux solutions obtenues en résolvant deux fois le Problème 2.2 pour $\boldsymbol{y} = \boldsymbol{y}_1$ et $\boldsymbol{y} = \boldsymbol{y}_2$. Pour se faire, nous allons quantifier la qualité des différentes estimées de $\overline{\mathbf{r}}$ à l'aide de l'erreur de Jaccard (voir Définition 2.7).

Etant donné que la qualité de l'estimation dépend fortement du choix de $\lambda$, nous balayons dans chacun des cas une gamme $\Lambda$ de 100 valeurs de $\lambda$ (équidistantes en échelle logarithmique)

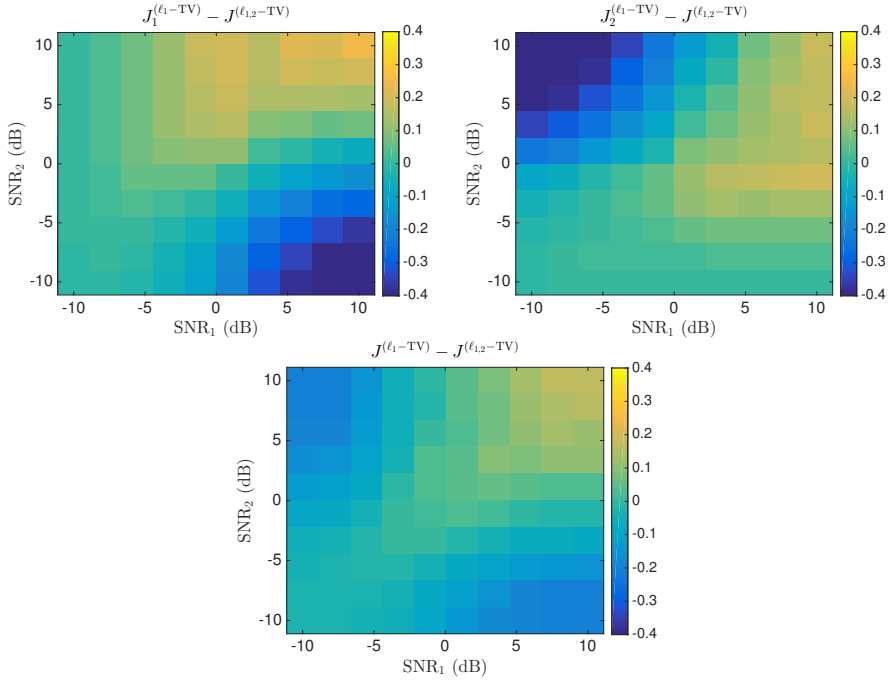



FIGURE 3.3 – **Apport de l'estimation vectorielle $\ell_{1,2}$-TV vs. deux estimations scalaires $\ell_1$-TV.**. Haut : $J_1^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$ et $J_2^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$ en fonction de $(\mathrm{SNR}_1, \mathrm{SNR}_2)$. Bas : $J^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$ en fonction de $(\mathrm{SNR}_1, \mathrm{SNR}_2)$.

allant de $10^{-1}$ à $10^2$, retenons l'erreur de Jaccard la plus faible, et notons :

$$J_1^{(\ell_1-\mathrm{TV})} = \min_{\lambda\in\Lambda} J_{\mathrm{error}}\left(\mathcal{G}\left(\widehat{\boldsymbol{r}}_{\mathbf{1},\lambda}^{(\ell_1-\mathrm{TV})}\right), \mathcal{G}\left(\overline{\boldsymbol{r}}_{\mathbf{1}}\right)\right), \tag{3.15}$$

$$J_2^{(\ell_1-\mathrm{TV})} = \min_{\lambda\in\Lambda} J_{\mathrm{error}}\left(\mathcal{G}\left(\widehat{\boldsymbol{r}}_{\mathbf{2},\lambda}^{(\ell_1-\mathrm{TV})}\right), \mathcal{G}\left(\overline{\boldsymbol{r}}_{\mathbf{2}}\right)\right), \tag{3.16}$$

$$J^{(\ell_1-\mathrm{TV})} = \min_{\lambda_1\in\Lambda}\min_{\lambda_2\in\Lambda} J_{\mathrm{error}}\left(\mathcal{G}\left([\widehat{\boldsymbol{r}}_{\mathbf{1},\lambda_1}^{(\ell_1-\mathrm{TV})}, \widehat{\boldsymbol{r}}_{\mathbf{2},\lambda_2}^{(\ell_1-\mathrm{TV})}]\right), \mathcal{G}\left([\overline{\boldsymbol{r}}_{\mathbf{1}}, \overline{\boldsymbol{r}}_{\mathbf{2}}]\right)\right), \tag{3.17}$$

$$J^{(\ell_{1,2}-\mathrm{TV})} = \min_{\lambda\in\Lambda} J_{\mathrm{error}}\left(\mathcal{G}\left(\widehat{\mathbf{r}}_{\lambda}^{(\ell_{1,2}-\mathrm{TV})}\right), \mathcal{G}\left(\overline{\mathbf{r}}\right)\right), \tag{3.18}$$

où $\mathcal{G}$ est un filtre passe-bas gaussien à support compact de taille 5 et d'écart-type 0.5. Les quantités $J_1^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$, $J_2^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$ et $J^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})}$ sont moyennées sur 50 réalisations et représentées Figure 3.3 en fonction de $(\mathrm{SNR}_1, \mathrm{SNR}_2)$.

**Apport de l'estimation vectorielle.** Les régions où $J_1^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})} > 0$ (resp. $J_2^{(\ell_1-\mathrm{TV})} - J^{(\ell_{1,2}-\mathrm{TV})} > 0$) correspondent aux configurations où l'estimation vectorielle donne de meilleures performances d'estimation de $\overline{\mathbf{r}}$ que la seule estimation de $\overline{\boldsymbol{r}}_{\mathbf{1}}$. (resp. $\overline{\boldsymbol{r}}_{\mathbf{2}}$). Nous remarquons Figure 3.3 (haut) que dès que la composante $\boldsymbol{y_1}$ (resp. $\boldsymbol{y_2}$) est significativement plus informative que la seconde, alors l'estimation scalaire donne globalement une erreur de Jaccard plus faible. Autrement, l'estimation vectorielle donne systématiquement de meilleurs résultats.

### 3.3.2 Quantification des performances

Maintenant que nous avons illustré l'intérêt d'une estimation vectorielle par rapport à plusieurs estimations scalaires dans un cadre simple de signaux dégradés par un bruit blanc

gaussien, nous allons à présent étudier les performances dans le cadre de l'estimation du spectre $\mathcal{D}$ de signaux multifractals par morceaux. Dans ce contexte, nous considérons l'observation $\mathbf{y} = (\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})^\top \in \mathbb{R}^{2\times N}$ (voir Définition 3.3).

**Configuration expérimentale.** Considérons une marche aléatoire multifractale $\boldsymbol{z}$ de taille $N = 2^9$, dont les propriétés multifractales sont homogènes par morceaux et changent à la position $\ell = 2^8$, selon une configuration motivée par exemple par l'observation de données d'activités cérébrales, où l'on passe du repos (longue mémoire $\boldsymbol{c}_{\mathbf{1}} \equiv 0.6 > 0.5$ et faiblement multifractal $\boldsymbol{c}_{\mathbf{2}} \equiv -0.0125 \simeq 0$) à une tache (faible corrélation $\boldsymbol{c}_{\mathbf{1}} \equiv 0.5$ mais forte multifractalité $\boldsymbol{c}_{\mathbf{2}} \equiv -0.0250 < 0$) (cf. e.g., [43]). Les quantités $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$, définies en (3.10), sont estimées à travers les octaves $j_1 = 2$ et $j_2 = 4$. Les performances d'estimation des deux méthodes seront évaluées et comparées, comme moyenne sur 100 réalisations, pour différentes tailles de fenêtres d'estimation $|\mathcal{W}_{j,\ell}|$ variant de $2^2$ à $2^8$ échantillons.

**Estimateurs proposés.** Nous proposons de raffiner les estimateurs $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})^\top \in \mathbb{R}^{2\times N}$ par une approche variationnelle faisant intervenir un terme de variation totale permettant de favoriser les solutions constantes par morceaux.

Une première approche consiste à résoudre indépendamment les deux Problèmes 2.2 suivants :

$$\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1},\lambda_1} = \underset{\boldsymbol{u}_{\mathbf{1}}\in\mathbb{R}^N}{\arg\min} \|\widehat{\boldsymbol{c}}_{\mathbf{1}} - \boldsymbol{u}_{\mathbf{1}}\|^2 + \lambda_1 \sum_{\ell=1}^{N-1} |(D\boldsymbol{u}_{\mathbf{1}})_\ell| \tag{3.19}$$

$$\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2},\lambda_2} = \underset{\boldsymbol{u}_{\mathbf{2}}\in\mathbb{R}^N}{\arg\min} \|\widehat{\boldsymbol{c}}_{\mathbf{2}} - \boldsymbol{u}_{\mathbf{2}}\|^2 + \lambda_2 \sum_{\ell=1}^{N-1} |(D\boldsymbol{u}_{\mathbf{2}})_\ell| \tag{3.20}$$

où $D \in \mathbb{R}^{(N-1)\times N}$ représente l'opérateur de premières différences défini en (2.3). Les solutions $\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1},\lambda_1}$ et $\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2},\lambda_2}$ étant biaisées (voir Section 2.1.2), nous définissons $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{disj}}_{\mathbf{1},\lambda_1}$ et $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{disj}}_{\mathbf{2},\lambda_2}$ les solutions de (3.19) et (3.20) ré-estimées *a posteriori* sur chaque segment.

Il est plus réaliste dans les applications de considérer que l'ensemble du spectre multifractal (donc *a priori* à la fois $\boldsymbol{c}_{\mathbf{1}}$ et $\boldsymbol{c}_{\mathbf{2}}$) change à un instant donné. Pour favoriser la détection de discontinuités conjointes dans les estimées $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$, nous considérons le Problème 3.1 appliqué à la quantité $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})$, i.e., :

$$(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})_\lambda = \underset{(\boldsymbol{u}_{\mathbf{1}},\boldsymbol{u}_{\mathbf{2}})\in\mathbb{R}^{2\times N}}{\arg\min} \sum_{p=1}^{2} \|\widehat{\boldsymbol{c}}_{\boldsymbol{p}} - \boldsymbol{u}_{\boldsymbol{p}}\|^2 + \lambda \sum_{\ell=1}^{N-1} \sqrt{\sum_{p=1}^{2} |(D\boldsymbol{u}_{\boldsymbol{p}})_\ell|^2}. \tag{3.21}$$

Afin d'assurer que les termes d'attaches aux données contribuent de manière égale dans le critère (3.21), il est nécessaire de normaliser au préalable la dynamique de $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$. De la même manière, nous noterons $(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})^{\mathrm{conj}}_{\lambda}$, la solution de (3.21) ré-estimées *a posteriori* sur chaque segment.

**Illustration.** La Figure 3.4 illustre la procédure : En haut, le signal multifractal par morceaux et le spectre multifractal associé sur chaque région, au centre (resp. en bas), $\boldsymbol{c}_{\mathbf{1}}$, $\widehat{\boldsymbol{c}}_{\mathbf{1}}$, $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{disj}}_{\mathbf{1},\lambda_{J_{\max}}}$, $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{conj}}_{\mathbf{1},\lambda_{J_{\max}}}$ (resp. $\boldsymbol{c}_{\mathbf{2}}$, $\widehat{\boldsymbol{c}}_{\mathbf{2}}$, $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{disj}}_{\mathbf{2},\lambda_{J_{\max}}}$, $\widehat{\widehat{\boldsymbol{c}}}^{\mathrm{conj}}_{\mathbf{2},\lambda_{J_{\max}}}$) pour $|\mathcal{W}_{j,\ell}| = 2^8$. Cet exemple montre que l'estimation conjointe (rouge) permet de gagner en précision sur la localisation de la discontinuité, comparée à une estimation disjointe (bleu).

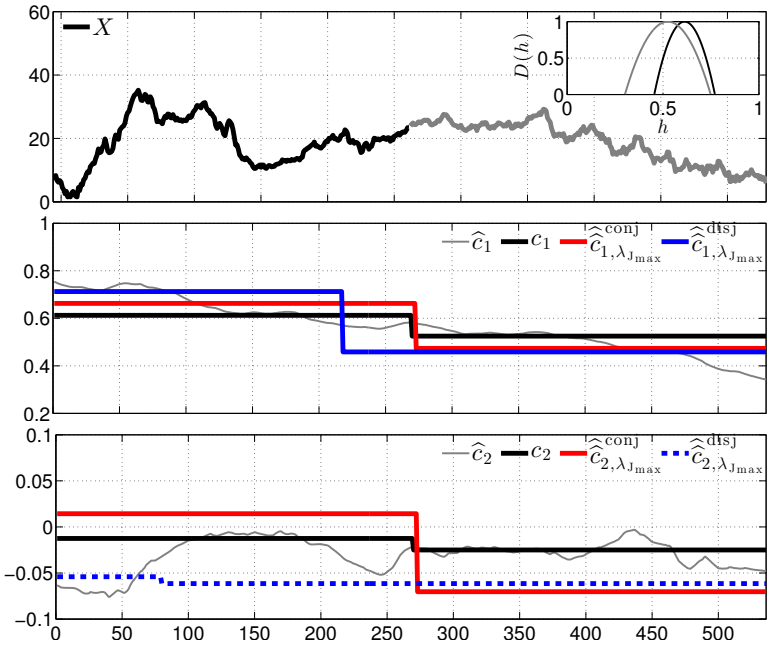



FIGURE 3.4 – **Illustration.** Comparaison entre les estimées $\widehat{\widehat{\boldsymbol{c}}}^{\text{disj}}_{\mathbf{1},\lambda_{J_{\max}}}$, $\widehat{\widehat{\boldsymbol{c}}}^{\text{disj}}_{\mathbf{2},\lambda_{J_{\max}}}$ et $(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})^{\text{conj}}_{\lambda_{J_{\max}}}$ de $(\boldsymbol{c_1}, \boldsymbol{c_2})$.

**Paramètre de régularisation.** Les performances d'estimation dépendent fortement du choix de $(\lambda_1, \lambda_2)$ et $\lambda$ qui n'ont pas le même poids dans (3.19)-(3.20) et (3.21). Cette question importante lorsque l'on fait face à des signaux réels a été discutée à la Section 3.3.1.2 pour un modèle de signal constant par morceaux corrompu par un bruit additif gaussien. Dans le cas présent, nous proposons de comparer les performances des deux approches pour deux valeurs différentes. Pour chaque problème (3.19)-(3.21), nous définissons :

$\lambda_{2\text{seg}}$ correspondant à la plus grande valeur de $\lambda$ telle que la solution possède au moins 2 segments.

$\lambda_{J_{\max}}$ est la plus grande valeur de $\lambda$ telle que la solution maximise l'indice de Jaccard (voir Section 2.3).

Pour notre étude, si nous notons $\widehat{\boldsymbol{r}}^{\text{disj}}_{\mathbf{1},\lambda}$, $\widehat{\boldsymbol{r}}^{\text{disj}}_{\mathbf{2},\lambda}$, $\widehat{\boldsymbol{r}}^{\text{conj}}_{\lambda}$ et $\boldsymbol{r}$ les vecteurs d'indicatrices des discontinuités respectivement associés à $\widehat{\boldsymbol{c}}^{\text{disj}}_{\mathbf{1},\lambda}$, $\widehat{\boldsymbol{c}}^{\text{disj}}_{\mathbf{2},\lambda}$, $(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})^{\text{conj}}_{\lambda}$ et $(\boldsymbol{c_1}, \boldsymbol{c_2})$, l'indice de Jaccard sera mesuré entre les quantités $\mathcal{G}(\widehat{\boldsymbol{r}}_\bullet)$ et $\mathcal{G}(\boldsymbol{r})$ où $\mathcal{G}$ modélise un filtre passe-bas gaussien à support compact. Pour favoriser les solutions avec peu de segments, nous choisissons $\mathcal{G}$ de taille $2^8$ points et de largeur à mi-hauteur $2^7$ points.

**Indice de Jaccard et nombre de segments.** La Figure 3.5 montre que l'estimation conjointe permet d'atteindre un indice de Jaccard plus élevé quelque soit la taille de la fenêtre d'estimation $|\mathcal{W}_{j,\ell}|$. Par conséquent, elle bénéficie d'une meilleure estimation de la position de la discontinuité. De plus, elle conduit également à un nombre de segments plus faible, ce qui dans notre cas est préférable.

**Distance à la discontinuité.** La Figure 3.6 compare les performances en terme de distance entre les discontinuités pour $\lambda = \lambda_{2\text{seg}}$ (Figure 3.6(a)) et $\lambda = \lambda_{J_{\max}}$ (Figure 3.6(b)). Pour l'estimation disjointe, $\widehat{\widehat{\boldsymbol{c}}}^{\text{disj}}_{\mathbf{1},\lambda}$ montre une meilleure localisation de la détection que $\widehat{\widehat{\boldsymbol{c}}}^{\text{disj}}_{\mathbf{2},\lambda}$. De plus, l'estimation conjointe, bénéficiant de l'information sur les deux composantes, présente

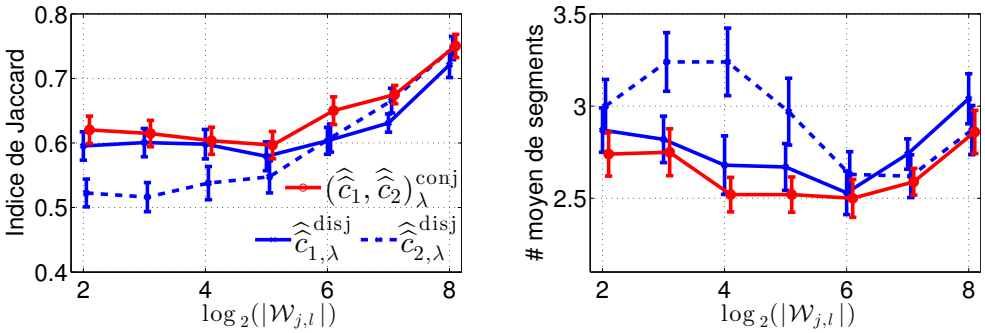
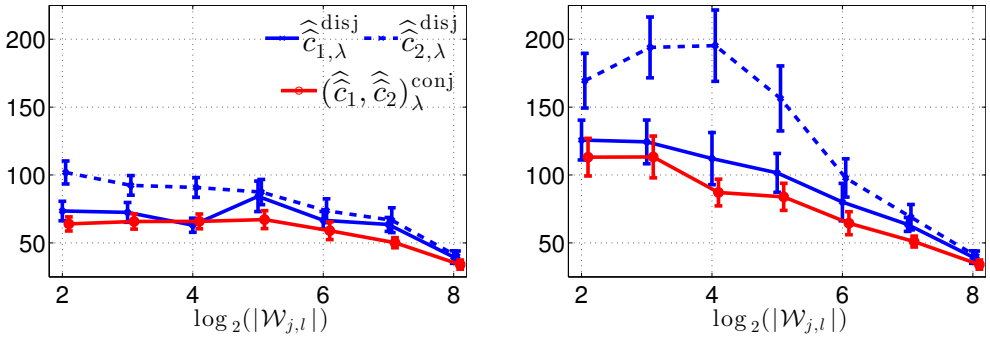
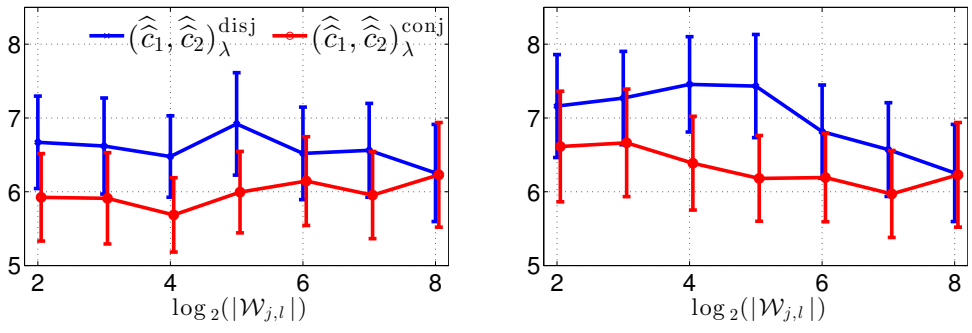



FIGURE 3.5 – **Performances pour $\lambda_{\mathrm{J}_{\max}}$.**




(a) $\lambda = \lambda_{2\text{seg}}$ (b) $\lambda = \lambda_{\mathrm{J}_{\max}}$

FIGURE 3.6 – **Distance à la discontinuité.**




(a) $\lambda = \lambda_{2\text{seg}}$ (b) $\lambda = \lambda_{\mathrm{J}_{\max}}$

FIGURE 3.7 – **Erreur quadratique moyenne relative.**

systématiquement une meilleure localisation de la détection des discontinuités.

**Erreur quadratique relative moyenne.** La Figure 3.7 compare, pour $\lambda = \lambda_{2\text{seg}}$ (a) et $\lambda = \lambda_{J_{\max}}$ (b), les performances d'estimation en terme d'erreur quadratique relative moyenne (MSEr)

$$\text{MSEr}\left[(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}}), (\boldsymbol{c}_{\mathbf{1}}, \boldsymbol{c}_{\mathbf{2}})\right] = \widehat{\mathbb{E}}\left[\frac{\|\widehat{\boldsymbol{c}}_{\mathbf{1}} - \boldsymbol{c}_{\mathbf{1}}\|^2}{\|\boldsymbol{c}_{\mathbf{1}}\|^2} + \frac{\|\widehat{\boldsymbol{c}}_{\mathbf{2}} - \boldsymbol{c}_{\mathbf{2}}\|^2}{\|\boldsymbol{c}_{\mathbf{2}}\|^2}\right] \tag{3.22}$$

où $\widehat{\mathbb{E}}$ est un estimateur empirique de la moyenne calculé sur 100 réalisations. Quelque soit la taille de la fenêtre d'estimation $|\mathcal{W}_{j,\ell}|$, on constate que

$$\text{MSEr}[(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})_{\lambda}^{\text{conj}}, (\boldsymbol{c}_{\mathbf{1}}, \boldsymbol{c}_{\mathbf{2}})] \leq \text{MSEr}[(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})_{\lambda}^{\text{disj}}, (\boldsymbol{c}_{\mathbf{1}}, \boldsymbol{c}_{\mathbf{2}})].$$

Pour $|\mathcal{W}_{j,\ell}| = 2^8$, la MSEr des deux méthodes est sensiblement la même. Cependant, ce résultat doit être nuancé avec les résultats présentés Figure 3.6 et Figure 3.5. En effet, à MSEr égale, $(\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1}}, \widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2}})_{\lambda}^{\text{conj}}$ détecte moins de segments et localise la discontinuité 15% plus près que $\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{1},\lambda}^{\text{disj}}$ et $\widehat{\widehat{\boldsymbol{c}}}_{\mathbf{2},\lambda}^{\text{disj}}$.

**Influence de $|\mathcal{W}_{j,\ell}|$.** Plus $|\mathcal{W}_{j,\ell}|$ est grand, plus les estimées $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$ sont lisses. Inversement, pour $|\mathcal{W}_{j,\ell}|$ faible, les estimées sont fortement variables. Dans les deux cas, cela rend difficile la détection et la localisation de la discontinuité. Les résultats expérimentaux que nous rapportons montrent que les détections de discontinuités de $\boldsymbol{c}_{\mathbf{1}}$ et $\boldsymbol{c}_{\mathbf{2}}$ sont mieux estimées par l'utilisation d'une grande fenêtre d'estimation. La segmentation vectorielle assure de localiser le plus précisément possible ces estimées pourtant fortement lissées au cours temps, un résultat peu intuitif *a priori*.

## 3.4 Segmentation vectorielle par approche TV (2D)

Dans cette section, nous considérons le cas bidimensionnel où $\boldsymbol{z} \in \mathbb{R}^{|\Omega|}$ désigne une image avec un total de $|\Omega|$ pixels.

Par analogie avec la Section 3.3, nous pouvons estimer $(C_{1,j})_{j\in\mathbb{Z}}$ et $(C_{2,j})_{j\in\mathbb{Z}}$ localement dans le voisinage de chaque pixel $\underline{\ell}$, concevoir les estimateurs locaux $\widehat{\boldsymbol{c}}_{\mathbf{1}}$ et $\widehat{\boldsymbol{c}}_{\mathbf{2}}$, puis appliquer un procédé de segmentation vectorielle de $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})$. Cependant, cette méthode repose sur l'hypothèse forte que les textures ou les signaux issus du monde réel suivent précisément le comportement d'échelle prescrit à la Proposition 1.3. Dans cette section, nous choisissons de relâcher cette contrainte et de segmenter directement les quantités vectorielles $\widehat{\boldsymbol{C}}_{\mathbf{1}}$ et/ou $\widehat{\boldsymbol{C}}_{\mathbf{2}}$ à travers les échelles.

Dans la suite, nous considérons la quantité $\mathbf{y} = (\boldsymbol{y}_{\mathbf{1}}, \ldots, \boldsymbol{y}_{\boldsymbol{M}})^\top \in \mathbb{R}^{M\times|\Omega|}$ pouvant soit modéliser $\widehat{\boldsymbol{h}}$ (i.e., $M = 1$), $\widehat{\boldsymbol{C}}_{\mathbf{1}}$ ou $\widehat{\boldsymbol{C}}_{\mathbf{2}}$ (i.e., $M = J$), ou leur concaténation $(\widehat{\boldsymbol{C}}_{\mathbf{1}}, \widehat{\boldsymbol{C}}_{\mathbf{2}})^\top$ (i.e., $M = 2J$).

### 3.4.1 Segmentation vectorielle

Une extension vectorielle du Problème 2.10 peut être formulée à l'aide de $\boldsymbol{\theta}$ qui dépend à la fois de $m \in \{1, \ldots, M\}$ et de la partition $q \in \{1, \ldots, Q+1\}$ de la manière suivante :

**Problème 3.2** (Problème $\ell_0$-TV 2D vectoriel à $Q$ niveaux)**.** *Soit une quantité* $\mathbf{y} = (\boldsymbol{y}_{\mathbf{1}}, \ldots, \boldsymbol{y}_{\boldsymbol{M}})^\top \in \mathbb{R}^{M\times|\Omega|}$ *et un ensemble de valeurs de niveaux* $(v_{q,m})_{1\leq q\leq Q, 1\leq m\leq M}$ *avec comme convention*

$v_{q,m} \leq v_{q+1,m}$. *Alors le problème consiste à trouver le vecteur* $\boldsymbol{\theta} = (\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_{Q+1}})^\top \in [0,1]^{(Q+1)\times M\times|\Omega|}$ *tel que*

$$\min_{\boldsymbol{\theta}\in[0,1]^{(Q+1)\times M\times|\Omega|}} \sum_{q=1}^{Q}\sum_{m=1}^{M}(\boldsymbol{\theta_{q,m}} - \boldsymbol{\theta_{q+1,m}})^\top(\boldsymbol{y_m} - v_{q,m})^2 + \lambda \sum_{m=1}^{M}\sum_{q=1}^{Q} TV(\boldsymbol{\theta_{q,m}} - \boldsymbol{\theta_{q+1,m}})$$
$$\text{sujet à} \quad (\forall m \in \{1,\ldots,M\}) \quad \begin{cases} \boldsymbol{\theta_{1,m}} \equiv 1, \\ \boldsymbol{\theta_{Q+1,m}} \equiv 0, \\ 1 \geq \boldsymbol{\theta_2} \geq \ldots \geq \boldsymbol{\theta_Q} \geq 0, \end{cases} \tag{3.23}$$

*avec* $\lambda \geq 0$ *et où TV désigne la variation totale définie en* (2.68).

**Remarque 3.2.** *Dans toute la suite, les niveaux* $v_{q,m} \in \mathbb{R}$ *seront choisis de manière équidistante entre* $\min_{\underline{\ell}\in\Omega} y_{m,\underline{\ell}}$ *et* $\max_{\underline{\ell}\in\Omega} y_{m,\underline{\ell}}$.

Il apparait clairement que le critère (3.23) est séparable en $m$, et aucun couplage entre les composantes $m \in \{1,\ldots,M\}$ n'est imposé. Par conséquent, résoudre le Problème 3.2 revient à trouver le meilleur étiquetage des régions $(\Omega_{1,m},\ldots,\Omega_{Q,m})$ indépendamment sur chacune des composantes $m \in \{1,\ldots,M\}$.

Cependant, nous devons garder à l'esprit que nous ne souhaitons pas seulement segmenter le vecteur des caractéristiques $\mathbf{y}$ mais nous cherchons avant tout à segmenter la texture $\boldsymbol{z}$. En d'autre termes, nous souhaiterions obtenir un partitionnement $(\Omega_1,\ldots,\Omega_Q)$ commun à toutes les composantes $(\boldsymbol{y_m})_{1\leq m\leq M}$.

Pour se faire, nous considérons l'extension du Problème 2.10 au cadre vectoriel proposé dans [37] pour l'étiquetage d'images RGB (i.e., $M = 3$). Il repose sur l'utilisation d'une seule variable $\boldsymbol{\theta}$ qui est cette fois-ci commune à toutes les composantes, i.e., $\boldsymbol{\theta_q} \equiv \boldsymbol{\theta_{q,m}}$ $(\forall m \in \{1,\ldots,M\})$.

**Problème 3.3** (Relaxation convexe du problème $\ell_0$-TV 2D vectoriel à $Q$ niveaux)**.** *Soit* $\mathbf{y} = (\boldsymbol{y_1},\ldots,\boldsymbol{y_M}) \in \mathbb{R}^{M\times|\Omega|}$ *et un ensemble de valeurs de niveaux* $(v_q)_{1\leq q\leq Q}$ *avec comme convention* $v_q \leq v_{q+1}$. *Alors le problème consiste à trouver* $\boldsymbol{\theta} = (\boldsymbol{\theta_1},\ldots,\boldsymbol{\theta_{Q+1}})^\top \in [0,1]^{(Q+1)\times|\Omega|}$ *tel que*

$$\min_{\boldsymbol{\theta}\in[0,1]^{(Q+1)\times|\Omega|}} \sum_{q=1}^{Q}(\boldsymbol{\theta_q} - \boldsymbol{\theta_{q+1}})^\top \sum_{m=1}^{M}(\boldsymbol{y_m} - v_{q,m})^2 + \lambda\sum_{q=1}^{Q} TV(\boldsymbol{\theta_q} - \boldsymbol{\theta_{q+1}})$$
$$\text{sujet à} \quad \begin{cases} \boldsymbol{\theta_1} \equiv 1, \\ \boldsymbol{\theta_{Q+1}} \equiv 0, \\ 1 \geq \boldsymbol{\theta_2} \geq \ldots \geq \boldsymbol{\theta_Q} \geq 0, \end{cases} \tag{3.24}$$

*avec* $\lambda \geq 0$ *et où TV désigne la variation totale définie en* (2.68).

**Remarque 3.3.** *Pour que tous les termes d'attache aux données soient du même ordre de grandeur dans le critère 3.24, il est important de normaliser* $\mathbf{y}$ *composante par composante.*

Afin d'estimer efficacement $\boldsymbol{\theta}$, nous utilisons la stratégie algorithmique proposée dans [55] qui consiste à utiliser une méthode d'éclatement proximal couplée à une stratégie efficace pour calculer les opérateurs proximaux. Le lecteur peut se référer à [55] pour des détails quant à la stratégie algorithmique.

### 3.4.2 Quantification des performances

**Configuration expérimentale.** Les performances de la solution du Problème 3.3 sont évaluées sur des données synthétiques, produites numériquement par l'inclusion d'un patch 2D $\Omega_A$ d'une réalisation de marche aléatoire multifractale (voir Exemple 1.2) de paramètres $(\boldsymbol{c_1}, \boldsymbol{c_2})_A$ sur un fond $\Omega_B$ de paramètres $(\boldsymbol{c_1}, \boldsymbol{c_2})_B$. Le fond et le patch sont normalisés afin d'assurer que la variance locale ne dépende pas de la position de l'image.

Les simulations sont conduites en utilisant la transformée standard des coefficients d'ondelettes 2D avec le tenseur orthogonal produit de Daubechies à 2 moments nuls sur $J = 3$ échelles. Le procédé de segmentation est appliqué pour $Q = 2$, $\lambda = 1$ et $M = J = 3$. Pour chaque composante $m \in \{1, \dots M\}$, les $(v_{q,m})_{1 \leq q \leq Q}$ sont initialement choisis de manière équidistante entre $\min_{\underline{\ell} \in \Omega} y_{m,\underline{\ell}}$ et $\max_{\underline{\ell} \in \Omega} y_{m,\underline{\ell}}$. Puis, nous alternons la minimisation de (3.24) et la ré-estimation $(v_{q,j})_{1 \leq q \leq Q}$ 5 fois.

**Segmentation scalaire vs. vectorielle.** Considérons le patch $\Omega_A$ et le fond $\Omega_B$ respectivement représentés par la Figure 3.8 (b) en blanc et en noir. Une réalisation de texture multifractale par morceaux $\boldsymbol{z}$ est fournie Figure 3.8 (a) où les paramètres du patch sont $(\boldsymbol{c_1}, \boldsymbol{c_2})_A \equiv (0.4, -0.001)$ et ceux du fond sont $(\boldsymbol{c_1}, \boldsymbol{c_2})_B \equiv (0.5, -0.1)$.

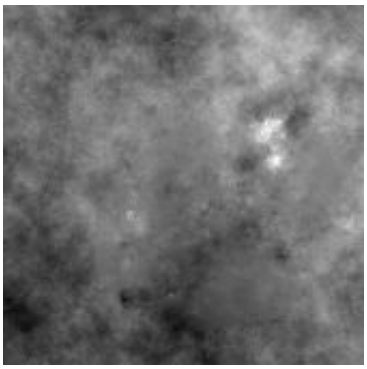
La segmentation scalaire de $\boldsymbol{y} = \widehat{\boldsymbol{h}}$, originellement envisagée pour l'analyse des processus monofractal par morceaux [176], est illustrée par la Figure 3.8-(c)). Nous observons que segmenter $\widehat{\boldsymbol{h}}$ fournit de mauvais résultats. En effet, puisque le processus $\boldsymbol{z}$ est multifractal, toutes les valeurs de régularité locale $h$ du support de $\mathcal{D}$ sont présentes dans tout interval ouvert de $\boldsymbol{z}$ pour des données de taille finie. Par conséquent, l'estimation locale de $h$ n'est pas pertinente. De plus, puisque le support de $\mathcal{D}_A$ et $\mathcal{D}_B$ se recouvrent, la seule quantité $\boldsymbol{h}$ ne peut pas permettre de discriminer entre $\Omega_A$ et $\Omega_B$.

Cette limitation démontre le besoin d'examiner des quantités multi-échelles liées au spectre multifractal, à savoir $\boldsymbol{C_1}$ et $\boldsymbol{C_2}$. En effet, les résultats de segmentation vectorielle de $\widehat{\boldsymbol{C}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$ et $(\widehat{\boldsymbol{C}}_{\boldsymbol{1}}, \widehat{\boldsymbol{C}}_{\boldsymbol{2}})^\top$, respectivement illustrés Figure 3.8-(d), Figure 3.8-(e) et Figure 3.8-(f), montrent que le masque est cette fois-ci estimé de façon satisfaisante.

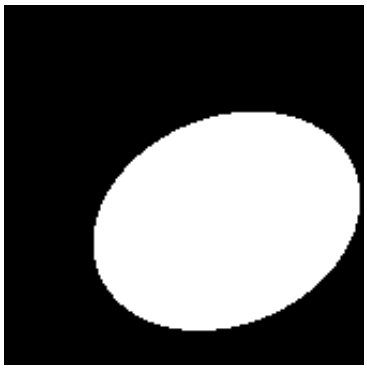
**Performances de segmentation.** Pour le même masque représenté par la Figure 3.8 (b), nous avons considéré 8 configurations reportées dans la Figure 3.9 permettant de modéliser différents écarts de $c_1$ et $c_2$ entre les régions $\Omega_A$ et $\Omega_B$. De plus, nous avons également examiné davantage l'impact de $\boldsymbol{c_2}$ : les résultats en bleu correspondent à un spectre $\mathcal{D}_B$ à support étendu alors que ceux en orange sont associés à un spectre $\mathcal{D}_B$ à support serré. Les performances d'estimation sont quantifiées en termes de pixels mal classés et sont moyennés sur 20 réalisations.

Sans surprise, segmenter $\widehat{\boldsymbol{h}}$ offre dans chaque configuration de mauvaises performances d'estimation. Par contre, les résultats de segmentation vectorielle de $\widehat{\boldsymbol{C}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$ et $(\widehat{\boldsymbol{C}}_{\boldsymbol{1}}, \widehat{\boldsymbol{C}}_{\boldsymbol{2}})^\top$ sont systématiquement satisfaisantes. De plus, toutes ces expériences reproduisent le comportement espéré suivant : les performances de segmentations associées à $\widehat{\boldsymbol{C}}_{\boldsymbol{1}}$ (resp. $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$) sont d'autant meilleures que la différence de $\boldsymbol{c_1}$ (resp. $\boldsymbol{c_2}$) entre $\Omega_A$ et $\Omega_B$ est grande. Cependant, une inspection plus minitieuse des résultats montre qu'une plus grande différence de $\boldsymbol{c_1}$ ne mène pas nécessairement à de meilleurs résultats de segmentation de $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$, comme le montre la ligne orange dans la Figure 3.9 (graphes de gauche). Globalement, on observe que plus $\mathbf{y}$ est informatif (e.g., $\mathbf{y} = (\widehat{\boldsymbol{C}}_{\boldsymbol{1}}, \widehat{\boldsymbol{C}}_{\boldsymbol{2}})^\top$), meilleure sera la segmentation.

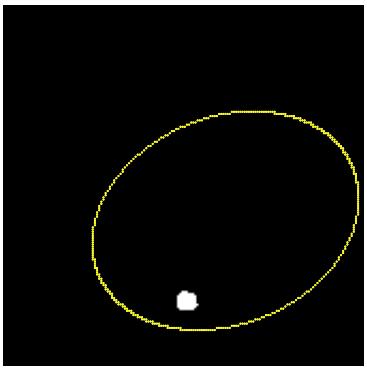
Finalement, il convient également de noter que la segmentation scalaire de $\widehat{\boldsymbol{h}}$ est seulement 2 fois plus rapide que la segmentation vectorielle de $\widehat{\boldsymbol{C}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$ ou $(\widehat{\boldsymbol{C}}_{\boldsymbol{1}}, \widehat{\boldsymbol{C}}_{\boldsymbol{2}})^\top$. En pratique, les simulations durent moins d'une minute par image de taille $|\Omega| = 2^9 \times 2^9$ pixels.



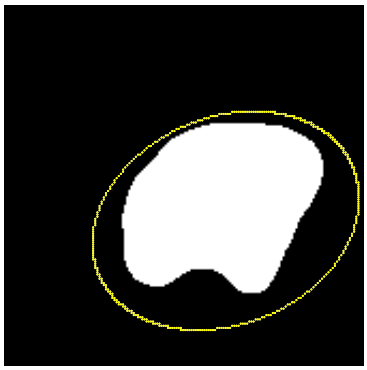
(a) $\boldsymbol{z}$ multifractal par morceaux



(b) Masque



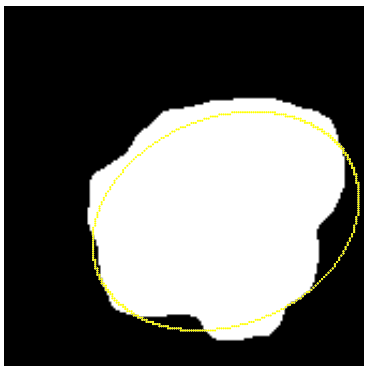
(c) Segmentation basée sur $\widehat{\boldsymbol{h}}$
Pixels mal classés : 29.0 %



(d) Segmentation basée sur $\widehat{\boldsymbol{C}}_{\mathbf{1}}$
Pixels mal classés : 5.7 %



(e) Segmentation basée sur $\widehat{\boldsymbol{C}}_{\mathbf{2}}$
Pixels mal classés : 3.0 %



(f) Segmentation basée sur $(\widehat{\boldsymbol{C}}_{\mathbf{1}}, \widehat{\boldsymbol{C}}_{\mathbf{2}})$
Pixels mal classés : 0.96 %

FIGURE 3.8 – **Illustration des résultats de segmentation conjointe.** L'image « multifractale par morceaux » $\boldsymbol{z}$ est présentée en (a) et a été générée à partir du masque en (b). La solution du Problème 3.2 pour $\boldsymbol{y} = \widehat{\boldsymbol{h}}$ est illustrée en (c). Les solutions du Problème 3.3 pour $\mathbf{y} = \widehat{\boldsymbol{C}}_{\mathbf{1}}$, $\widehat{\boldsymbol{C}}_{\mathbf{2}}$ et $(\widehat{\boldsymbol{C}}_{\mathbf{1}}, \widehat{\boldsymbol{C}}_{\mathbf{2}})^{\top}$ sont respectivement illustrées en (d), (e) et (f).

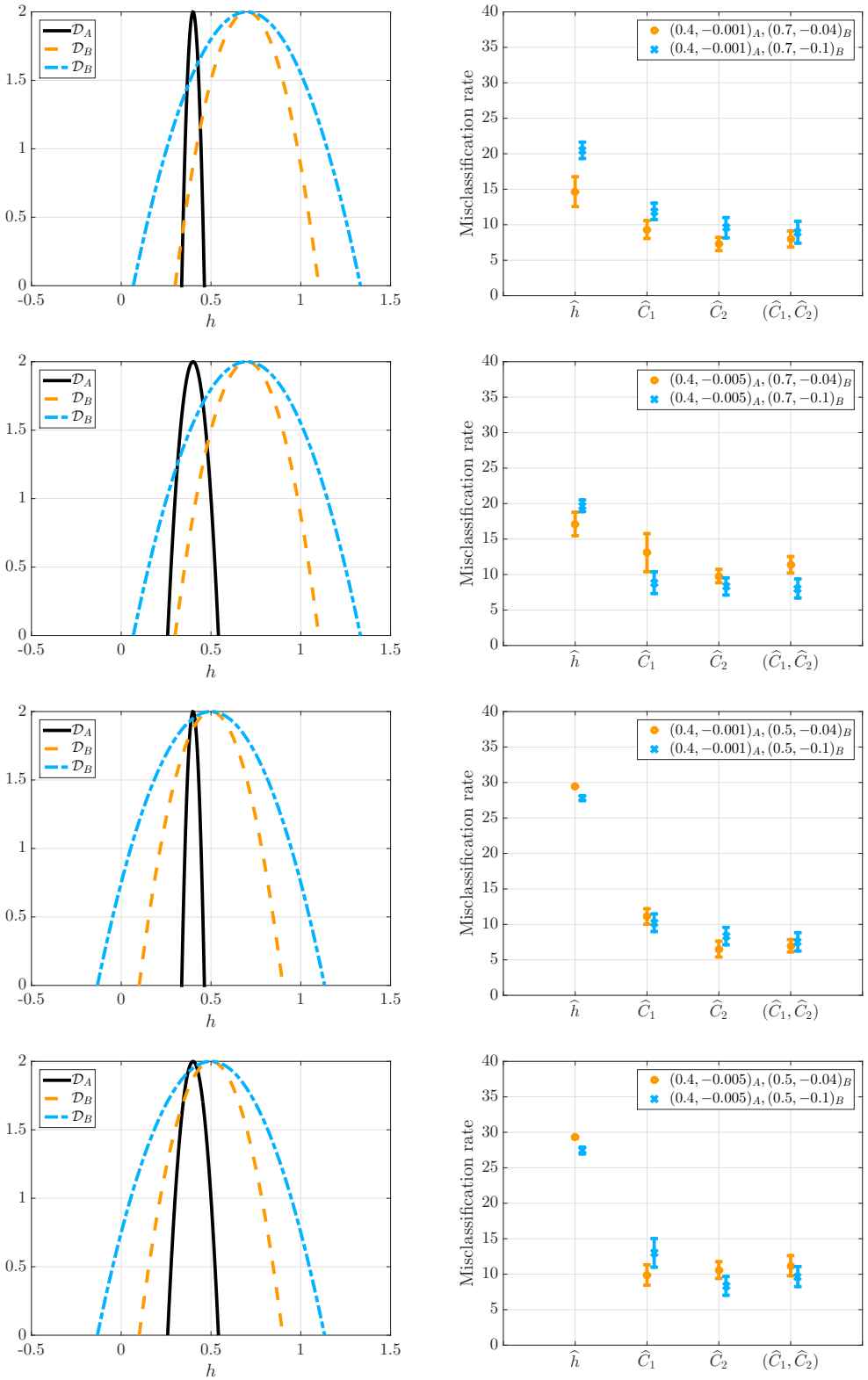



FIGURE 3.9 – **Performances de segmentation conjointe.** Comparaison des performances de segmentation selon le choix de la quantité **y** indiquée en abscisse. Chaque graphe représente deux configurations (bleu et orange) de $(\boldsymbol{c_1}, \boldsymbol{c_2})$ choisis sur $\Omega_A$ et $\Omega_B$.

## 3.5 Conclusion et perspectives

Ce chapitre propose une première tentative visant à analyser des processus multifractals par morceaux caractérisés par des propriétés multifractales homogènes différentes sur plusieurs régions distinctes. Elle repose sur l'estimation locale dans une fenêtre glissante des quantités multi-échelles $c_1$ et $c_2$, voire $C_1$ et $C_2$, impliquées dans la formulation du spectre multifractal, combinée à un procédé d'estimation ou de segmentation.

L'apport des approches TV vectorielles par rapport aux approches scalaires a tout d'abord été évalué sur des signaux constants par morceaux corrompus par un bruit Gaussien. Nous avons montré que lorsque le SNR est du même ordre de grandeur sur chacune des composantes, alors les approches vectorielles bénéficient d'une erreur de Jaccard plus faible, ce qui quantifie une meilleure localisation des positions des discontinuités. Des résultats similaires ont ensuite également été obtenus pour l'analyse d'une marche aléatoire multifractale par morceaux produit synthétiquement et reproduisant un cas d'intérêt pratique.

Nous avons montré dans le cas unidimensionel que l'estimation doit être conduite dans des voisinages $\mathcal{W}$ de grande taille. Cependant, la corrélation (temporelle ou spatiale) de $\widehat{\boldsymbol{c}}_{\boldsymbol{1}}$, $\widehat{\boldsymbol{c}}_{\boldsymbol{2}}$, $\widehat{\boldsymbol{C}}_{\boldsymbol{1}}$ et $\widehat{\boldsymbol{C}}_{\boldsymbol{2}}$, inhérente à la nécessité d'estimer ces quantités dans des voisinages de taille significative, est susceptible de compliquer *a priori* les méthodes d'estimation et de segmentation utilisées, reposant sur une formulation par variation totale, qui sont adéquates lorsque les signaux à régulariser ne sont pas corrélés. Nous envisageons de poursuivre ce travail en modifiant les termes d'attache aux données présents dans les équations (3.21), (3.24) et (B.1) afin de prendre en compte ces corrélations.

De récentes études [72, 73, 190] ont montrées l'intérêt d'analyser les propriétés multifractales du rythme cardiaque foetal (RCF) pour caractériser l'état de santé du foetus. Nous envisageons alors la possibilité d'estimer les discontinuités de $(\widehat{\boldsymbol{c}}_{\boldsymbol{1}}, \widehat{\boldsymbol{c}}_{\boldsymbol{2}})^\top$ comme un moyen de quantifier l'évolution de la santé du foetus au cours du temps. Afin d'offrir une estimation des discontinuités au fur et à mesure de l'acquisition du RCF, nous proposons dans le chapitre suivant une solution algorithmique permettant de fournir une solution approchée du Problème 3.1 à la volée.

Le procédé de segmentation proposé au Problème 3.3 démontre des performances très satisfaisantes mais repose sur la forte hypothèse que $\boldsymbol{\theta}_{\boldsymbol{q}} \equiv \boldsymbol{\theta}_{\boldsymbol{q},\boldsymbol{m}}$ $(\forall m \in \{1, \ldots, M\})$. D'un côté, cela a l'avantage de définir des partitions $(\Omega_1, \ldots, \Omega_Q)$ communes à toutes les $M$ composantes. De l'autre côté, cela implique que les performances sont très sensibles au choix arbitraire et à l'ordre des niveaux $v_{q,m}$. Afin de pallier cette limitation, différents termes de régularisation sont actuellement envisagés pour permettre une stratégie de segmentation plus flexible. Un premier pas dans cette direction peut par exemple être lié à l'utilisation du tenseur de structure de la variation totale, dont la méthode est décrite dans l'Annexe B, qui permet d'introduire des corrélations entre les composantes sans toutefois permettre une segmentation conjointe.

# Chapitre 4

# Estimation $\ell_{1,2}$-TV à la volée

**Sommaire**


**4.1 Nature non locale du problème $\ell_{1,2}$-TV . . . . . 74**
**4.2 Conditions de Karush-Kuhn-Tucker . . . . . 76**
4.2.1 Formulation duale . . . . . 76
4.2.2 Réécriture des conditions de KKT . . . . . 77
4.2.3 Solution approchée . . . . . 78
**4.3 Solution algorithmique à la volée . . . . . 79**
4.3.1 Cas idéal où $\widehat{\mathbf{z}}$ est connu . . . . . 79
4.3.1.1 Bornes inférieures et supérieures . . . . . 79
4.3.1.2 Règles de mise à jour & Règle 1 . . . . . 80
4.3.1.3 Prolongation du signal & Règle 2 . . . . . 81
4.3.1.4 Estimer la position de la discontinuité $n_{\text{rupt}}$ . . . . . 82
4.3.1.5 Commencer un nouveau segment . . . . . 83
4.3.2 Estimation du vecteur auxiliaire $\widehat{\mathbf{z}}$ . . . . . 83
4.3.2.1 Estimateur constant par morceaux de $\widehat{\mathbf{z}}$ . . . . . 84
4.3.2.2 Estimer les positions candidates de la discontinuité $n_{\text{rupt}}^{(q)}$ . 84
4.3.2.3 Estimer la position de la discontinuité $n_{\text{rupt}}$. . . . . 85
4.3.2.4 Commencer un nouveau segment . . . . . 85
**4.4 Résultats . . . . . 86**
4.4.1 Configuration expérimentale . . . . . 86
4.4.2 Conception de $\mathcal{Q}$ . . . . . 86
4.4.3 Performances hors ligne . . . . . 88
4.4.4 Performances en ligne . . . . . 89
**4.5 Conclusion et perspectives . . . . . 91**


**Extension vectorielle.** Dans le chapitre précédent, nous avons démontré l'intérêt de considérer des approches TV vectorielles pour détecter les régions où le spectre multifractal diffère via l'estimation des discontinuités conjointes à $(\widehat{\boldsymbol{c}}_{\mathbf{1}}, \widehat{\boldsymbol{c}}_{\mathbf{2}})$ voire $(\widehat{\boldsymbol{C}}_{\mathbf{1},\boldsymbol{j}}, \widehat{\boldsymbol{C}}_{\mathbf{2},\boldsymbol{j}})_j$. L'extension vectorielle émerge également dans de nombreux contextes comme par exemple dans certaines applications biomédicales où l'objectif est d'extraire les discontinuités conjointes présentes dans des données multivariées, e.g., données EEG [98]. Ce type de problème s'applique également aux données à valeurs complexes qui peuvent naturellement s'interpréter comme une quantité vectorielle, i.e., la partie réelle et la partie complexe des données. D'un point de vue bayésien, des solutions élégantes ont été proposées en [69, 167] et des stratégies

itératives efficaces ont récemment été introduites en [26, 107].

**Algorithmes à la volée.** Afin d'adresser le problème lié au temps de calcul dû à la nature itérative des algorithmes proximaux pour résoudre le Problème 2.2, des procédés alternatifs ont été étudiés, tel quel *l'algorithme « taut string »* utilisé communément dans la littérature statistique [141]. Très récemment, un algorithme rapide reposant sur la formulation duale et l'analyse des conditons de Karush-Kuhn-Tucker (KKT) a été proposé en [52] pour résoudre le Problème 2.2. Contrairement à l'algorithme « taut string » , ce dernier permet d'éviter des sommes cumulées, pouvant potentiellement mener à des valeurs excessives et par conséquent à des erreurs numériques. Une autre spécificité de cet algorithme concerne son comportement *à la volée*, c'est-à-dire qu'il ne nécessite pas l'observation de toute la série temporelle avant de fournir une solution sur une sous-partie. Les algorithmes à la volée pourraient être d'un intérêt crucial pour les problèmes de surveillance en temps réel rencontrés notamment dans le milieu médical [97, 33].

**Contribution** Dans ce contexte, le présent chapitre élabore à partir de [52] un algorithme à la volée permettant de résoudre le Problème 3.1. Nous illusterons en premier lieu le comportement *non local* du problème vectoriel par opposition au caractère *local* du problème scalaire, ce qui démontre que tout algorithme à la volée pour résoudre le Problème 3.1 ne fournira seulement une solution approchée. A partir de cette observation et de la réécriture des conditions de KKT résultant de la formulation duale du Problème 3.1 spécifiées à la Section 4.2, un algorithme rapide et à la volée est dérivé à la Section 4.3. Les performances en termes de qualité de la solution et de gains de temps de calcul sont présentées dans la Section 4.4.

**Référence.** Ce chapitre s'appuie sur la contribution [87] et est le résultat d'une collaboration avec L. Condat. Une vidéo démontrant le caractère à la volée de l'algorithme ainsi que les routines MATLAB correspondantes sont disponibles à l'adresse `http://perso.ens-lyon.fr/jordan.frecon`.

**Notations.** Soit $\mathbf{u} = (u_{m,n})_{1\leq m\leq M, 1\leq n\leq N} \in \mathbb{R}^{M\times N}$ un signal multivarié, où pour chaque $m \in \{1,\dots,M\}$, $\boldsymbol{u_m} = (u_{m,n})_{1\leq n\leq N} \in \mathbb{R}^N$ correspond à la $m$-ème composante alors que la $n$-ème valeur sera raccourcie en $\mathbf{u}_n = (u_{m,n})_{1\leq m\leq M} \in \mathbb{R}^M$. Pour tout $n \in \{1,\dots,N\}$, nous utiliserons également les fonctions suivantes : $\mathrm{abs}(\mathbf{u}_n) = (|u_{m,n}|)_{1\leq m\leq M} \in \mathbb{R}^M$ et $\mathrm{sgn}(\mathbf{u}_n) = \big(\mathrm{sgn}(u_{m,n})\big)_{1\leq m\leq M} \in \mathbb{R}^M$.

## 4.1 Nature non locale du problème $\ell_{1,2}$-TV

Il existe une différence de nature fondamentale entre l'estimation scalaire ($M = 1$) et vectorielle ($M > 1$). La première est intrinsèquement de nature locale [52, 136] alors que la seconde est non locale. Nous entendons par local le fait que la solution à une position donnée ne dépend pas du signal localisé avant (resp. après) la discontinuité précédente (resp. suivante). Afin d'expliquer cette notion, nous proposons l'expérience suivante dont les résultats sont illustrés Figure 4.1. Les résultats associés à l'estimation scalaire (resp. vectorielle) sont présentés sur le graphe de gauche (resp. droite).

Premièrement, considérons la quantité scalaire $\boldsymbol{y} \in \mathbb{R}^N$ ($N = 180$), composée de la somme d'un signal constant par morceaux et d'un bruit blanc Gaussien (gris, graphe en haut à droite). La solution $\widehat{\boldsymbol{x}}$ du Problème 2.2 est affichée en traits pleins rouges. Nous

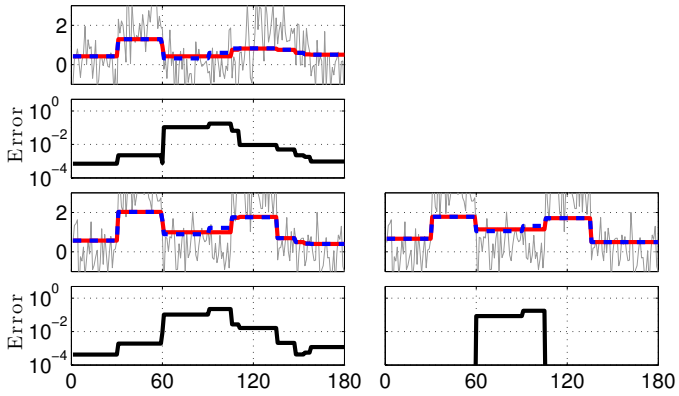



FIGURE 4.1 – **Nature non locale ($\ell_{1,2}$-TV) vs. locale ($\ell_1$-TV).** Gauche : solutions du Problème 3.1 pour $M = 2$ (haut : 1ère composante, bas : 2ème composante). Droite : solutions du Problème 2.2. Observations $\mathbf{y}$ (gris), solution $\widehat{\mathbf{x}}$ (rouge), concaténation des solutions $\widehat{\mathbf{x}}_-$ et $\widehat{\mathbf{x}}_+$ (bleu pointillés).

considérons également les solutions du Problème 2.2 appliqués à deux partitions de $\boldsymbol{y}$ obtenues en scindant $\boldsymbol{y}$ en deux, i.e., $\boldsymbol{y}_- = (y_n)_{1\leq n\leq N/2}$ et $\boldsymbol{y}_+ = (y_n)_{N/2+1\leq n\leq N}$. Les solutions $\widehat{\boldsymbol{x}}_-$ et $\widehat{\boldsymbol{x}}_+$ du Problème 2.2 respectivement associées à $\boldsymbol{y}_-$ et $\boldsymbol{y}_+$ sont concaténées en une solution représentée Figure 4.1 en traits pointillés bleus. Nous observons qu'il n'existe aucune différence entre $\widehat{\boldsymbol{x}}$ et la concaténation de $\widehat{\boldsymbol{x}}_-$ et $\widehat{\boldsymbol{x}}_+$, illustrée Figure 4.1, mis à part pour le segment qui contient le point de concaténation. La différence autour du point de concaténation est en effet attendue car $\widehat{\boldsymbol{x}}$ utilise une information (la continuité entre $\boldsymbol{y}_-$ et $\boldsymbol{y}_+$) qui n'est pas disponible lorsque les solutions $\widehat{\boldsymbol{x}}_-$ et $\widehat{\boldsymbol{x}}_+$ sont calculées. Le fait qu'il n'existe aucune différence ailleurs illustre la nature locale de la solution du Problème 2.2.

Cette expérience est maintenant reconduite pour $M = 2$ car c'est le plus simple représentatif du cas vectoriel $M > 1$. Nous considérons cette fois-ci une quantité vectorielle $\mathbf{y} = (\boldsymbol{y_1}, \boldsymbol{y_2})^\top \in \mathbb{R}^{2\times N}$ ($N = 180$) également composée de la somme de signaux constants par morceaux et de bruits blancs Gaussien (en gris, Figure 4.1, 1ère et 3ème lignes). Deux partitions $\mathbf{y}_- = (y_{1,n}, y_{2,n})_{1\leq n\leq N/2}$ et $\mathbf{y}_+ = (y_{1,n}, y_{2,n})_{N/2+1\leq n\leq N}$ obtenues en scindant $\mathbf{y}$ en deux sont également considérées. Les solutions du Problème 3.1, appliqué à $\mathbf{y}, \mathbf{y}_-, \mathbf{y}_+$, respectivement notées $\widehat{\mathbf{x}}$, $\widehat{\mathbf{x}}_-$ et $\widehat{\mathbf{x}}_+$ sont obtenues pour $\lambda = 20$ via l'Algorithme 5. Les solutions $\widehat{\mathbf{x}}_-$ et $\widehat{\mathbf{x}}_+$ respectivement associées à $\mathbf{y}_-$ et $\mathbf{y}_+$ sont concaténées et représentées Figure 4.1 en traits bleus pointillés, tandis que $\widehat{\mathbf{x}}$ est illustré en rouge. Contrairement au cas $M = 1$, la différence entre $\widehat{\mathbf{x}}$ et la concaténation de $\widehat{\mathbf{x}}_-$ et $\widehat{\mathbf{x}}_+$, affichée en noir sur le graphique du bas, est différente de 0 sur l'ensemble du support de $\mathbf{y}$, montrant clairement la nature non locale de $\widehat{\mathbf{x}}$ lorsque $M > 1$.

Dans le cas scalaire, la nature locale de la solution permet de mettre au point un algorithme à la volée, dit « taut-string », qui consiste à trouver la corde (« string ») de longueur minimale (appelée « taut string » ) contenue dans le tube de rayon $\lambda$ autour de la primitive de $\mathbf{y}$. La solution $\widehat{\boldsymbol{x}}$ du Problème 2.2 ($\ell_1$-TV) est ensuite obtenue en calculant la dérivée de la « taut string ». Une stratégie efficace a été proposée en [58] afin de déterminer directement (i.e., sans itérations) les points de contact entre la « taut string » et le tube. Bien que cette approche puisse être généralisée au cadre vectoriel, la détection des points de contact requiert en plus l'angle du contact entre la « taut string » et le tube. Or, cette information est justement non locale, ce qui implique que minimiser le Problème 3.1 à la volée est un

problème difficile qui ne peut pas être résolu localement. Cette interprétation sera discutée plus en détail dans la Section 4.2.2.

La nature non locale du Problème 3.1 implique que l'on ne puisse pas espérer trouver la solution exacte à l'aide d'un algorithme à la volée. Par conséquent, dans ce chapitre, nous présenterons un algorithme à la volée permettant de fournir une approximation de bonne qualité de la solution exacte du Problème 3.1. Un paramètre de contrôle $|\mathcal{Q}|$, qui sera défini dans la Section 4.2.3, permettra de contrôler le compromis entre la qualité de l'approximation et le temps d'exécution.

## 4.2 Conditions de Karush-Kuhn-Tucker

### 4.2.1 Formulation duale

Nous rappelons ci-dessous la formulation duale du Problème 3.1 au sens de la dualité de Fenchel-Moreau-Rockafellar (voir Section 2.2.1 et [81, 156, 180]) :

**Problème 4.1** (Problème $\ell_{1,2}$-TV dual)**.** *Soient* $\mathbf{y} = (\boldsymbol{y_1}, \dots, \boldsymbol{y_M})^\top \in \mathbb{R}^{M\times N}$ *et* $\lambda \geq 0$*. On appelle problème* $\ell_{1,2}$*-TV dual le problème consistant à trouver*

$$\widehat{\mathbf{u}} = \underset{\mathbf{u}=(\boldsymbol{u_1},\dots,\boldsymbol{u_M})^\top \in \mathbb{R}^{M\times(N-1)}}{\arg\min} \frac{1}{2}\sum_{m=1}^{M} \|\boldsymbol{y_m} + L^*\boldsymbol{u_m}\|^2 \quad \textit{assujetti à}$$
$$(\forall n \in \{1,\dots,N-1\}) \quad \|\mathbf{u}_n\| \leq \lambda, \quad (4.1)$$

*où, pour tout* $m \in \{1,\dots,M\}$ *et* $n \in \{2,\dots,N-2\}$,

$$(L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}})_n = \widehat{u}_{m,n-1} - \widehat{u}_{m,n} \quad (4.2)$$

*et*

$$\begin{cases} (L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}})_1 &= -\widehat{u}_{m,1}, \\ (L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}})_N &= \widehat{u}_{m,N-1}. \end{cases} \quad (4.3)$$

**Remarque 4.1.** *Notons que la formulation duale usuelle implique* $-\boldsymbol{u_m}$ *à la place de* $\boldsymbol{u_m}$ *dans* (4.1)*. Nous choisissons cette définition par souci de cohérence avec les résultats obtenus en [52].*

D'après le Lemme 2.1, les solutions optimales $\widehat{\mathbf{u}} \in \mathbb{R}^{M\times(N-1)}$ et $\widehat{\mathbf{x}} \in \mathbb{R}^{M\times N}$ du Problème 4.1 dual et du Problème 3.1 primal sont reliées par

$$\begin{cases} (\forall m \in \{1,\dots,M\}) & \widehat{\boldsymbol{x}}_{\boldsymbol{m}} = \boldsymbol{y_m} + L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}}, \\ (\forall n \in \{1,\dots,N-1\}) & \widehat{\mathbf{u}}_n \in -\lambda\partial\|\cdot\|(\widehat{\mathbf{x}}_{n+1} - \widehat{\mathbf{x}}_n). \end{cases} \quad (4.4)$$

A partir de (4.4), nous obtenons directement les conditions nécessaires et suffisantes d'optimalité :

**Proposition 4.1.** *Les solutions du Problème 4.1 dual et du Problème 3.1 primal satisfont les conditions nécessaires et suffisantes suivantes. Pour tout* $m \in \{1,\dots,M\}$*,*

$$\widehat{\boldsymbol{x}}_{\boldsymbol{m}} = \boldsymbol{y_m} + L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}}, \quad (4.5)$$

*et, pour tout* $n \in \{1,\dots,N-1\}$*,*

$$\begin{cases} si \quad \widehat{\mathbf{x}}_n = \widehat{\mathbf{x}}_{n+1} \quad alors \quad \|\widehat{\mathbf{u}}_n\| \leq \lambda, \\ si \quad \widehat{\mathbf{x}}_n \neq \widehat{\mathbf{x}}_{n+1} \quad alors \quad \widehat{\mathbf{u}}_n = -\lambda\frac{\widehat{\mathbf{x}}_{n+1}-\widehat{\mathbf{x}}_n}{\|\widehat{\mathbf{x}}_{n+1}-\widehat{\mathbf{x}}_n\|}. \end{cases} \quad (4.6)$$

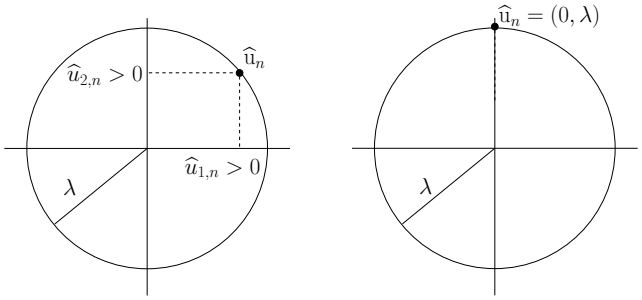



FIGURE 4.2 – **Comparaison entre les discontinuités simultanées et non-simultanées dans l'espace dual.** Gauche : la position $n$ est appropriée pour introduire un changement d'amplitude négatif sur les deux composantes. Droite : la position $n$ est appropriée pour introduire un changement d'amplitude négatif sur la 2ème composante seulement.

La première condition dans (4.6) correspond au cas où toutes les composantes $m \in \{1, \ldots, M\}$ gardent la même valeur à l'instant $n+1$ que celle à l'instant $n$. Cette configuration est illustrée Figure 4.2 (graphe de gauche) pour $M = 2$. La seconde condition modélise les situations où certaines composantes $m$ de $\widehat{\mathbf{x}}$ admettent une discontinuité entre les instants $n$ et $n+1$. Entre autre, la configuration où toutes les composantes de $\widehat{\mathbf{x}}$ ne changent pas simultanément de valeur est particulièrement intéressante (voir Figure 4.2 graphe de droite). En effet, en présence de bruit cette condition est très rarement rencontrée. Par conséquent, dans la suite de ce chapitre nous ne considérerons que le cas où les discontinuités sont simultanées sur toutes les composantes.

**Remarque 4.2.** *La Proposition 4.1 pour $M = 1$ mène aux conditions de KKT usuellement associées au Problème 2.2 ($\ell_1$-TV) :*

$$\begin{cases} si \quad \widehat{x}_n > \widehat{x}_{n+1} \quad alors \quad \widehat{u}_n = +\lambda, \\ si \quad \widehat{x}_n < \widehat{x}_{n+1} \quad alors \quad \widehat{u}_n = -\lambda, \\ si \quad \widehat{x}_n = \widehat{x}_{n+1} \quad alors \quad \widehat{u}_n \in [-\lambda, +\lambda]. \end{cases} \tag{4.7}$$

Par ailleurs, l'algorithme à la volée proposé en [52] et permettant de résoudre le Problème 2.2 ($\ell_1$-TV) est dérivé à partir des Conditions (4.7).

### 4.2.2 Réécriture des conditions de KKT

Contrairement aux Conditions (4.7) pour l'estimation scalaire, les conditions dans le cadre vectoriel dérivées dans la Proposition 4.1 ne sont pas directement utilisables en pratique pour mettre au point un algorithme à la volée car $\widehat{\mathbf{x}}_{n+1} - \widehat{\mathbf{x}}_n$ est *a priori* inconnu à l'instant $n$.

Par conséquent, nous proposons de réécrire la seconde condition de (4.6) par l'intermédiaire de variables auxiliaires non-négatives $(\widehat{\mathbf{z}}_n)_{1 \leq n \leq N-1}$ telles que

$$\begin{cases} \text{si} \quad \widehat{\mathbf{x}}_n = \widehat{\mathbf{x}}_{n+1} \quad \text{alors} \quad \|\widehat{\mathbf{u}}_n\| \leq \lambda, \\ \text{si} \quad \widehat{\mathbf{x}}_n \neq \widehat{\mathbf{x}}_{n+1} \quad \text{alors} \quad \widehat{\mathbf{u}}_n = -\text{sign}(\widehat{\mathbf{x}}_{n+1} - \widehat{\mathbf{x}}_n) \circ \widehat{\mathbf{z}}_n, \end{cases} \tag{4.8}$$

où $\widehat{\mathbf{z}}_n = \lambda \frac{\text{abs}(\widehat{\mathbf{x}}_{n+1} - \widehat{\mathbf{x}}_n)}{\|\widehat{\mathbf{x}}_{n+1} - \widehat{\mathbf{x}}_n\|}$ et où $\circ$ est le produit matriciel de Hadamard. Dès lors, la Proposition 4.1 peut être reformulée composante par composante comme suit :

**Proposition 4.2.** *Les solutions du Problème 3.1 primal et du Problème 4.1 dual satisfont les conditions nécessaires et suffisantes suivantes. Il existe des variables auxiliaires non-négatives*

$(\widehat{\mathbf{z}}_n)_{1\leq n\leq N-1}$ *telles que, pour tout* $m=\{1,\dots,M\}$ *et* $n\in\{1,\dots,N-1\}$,

$$\begin{cases} si \quad \widehat{x}_{m,n} > \widehat{x}_{m,n+1} \quad alors \quad \widehat{u}_{m,n} = +\widehat{z}_{m,n}, \\ si \quad \widehat{x}_{m,n} < \widehat{x}_{m,n+1} \quad alors \quad \widehat{u}_{m,n} = -\widehat{z}_{m,n}, \\ si \quad \widehat{x}_{m,n} = \widehat{x}_{m,n+1} \quad alors \quad \widehat{u}_{m,n} \in [-\widehat{z}_{m,n}, +\widehat{z}_{m,n}], \end{cases} \tag{4.9}$$

*avec* $\|\widehat{\mathbf{z}}_n\| = \lambda$ *et*

$$\widehat{\boldsymbol{x}}_{\boldsymbol{m}} = \boldsymbol{y}_{\boldsymbol{m}} + L^*\widehat{\boldsymbol{u}}_{\boldsymbol{m}}. \tag{4.10}$$

La comparaison entre les équations (4.7) et (4.9) souligne la similarité entre les conditions nécessaires et suffisantes du Problème 2.2 et du Problème 3.1. En effet, les conditions impliquant $\lambda$ dans le cas scalaire impliquent le vecteur auxiliaire $\widehat{\mathbf{z}}$ dans le cas vectoriel. Le fait que $\widehat{\mathbf{z}}$ varie pour chaque paire $(m,n)$ peut s'interpréter dans les procédures « taut string » comme la variation de l'angle de contact entre le tube de rayon $\lambda$ et la « taut string ». C'est justement le manque de connaissance sur $\widehat{\mathbf{z}}$ qui augmente drastiquement la difficulté pour mettre au point un algorithme à la volée.

### 4.2.3 Solution approchée

Si l'on suppose que $\widehat{\mathbf{z}}$ est connu *a priori* de telle manière que, pour tout $n\in\{1,\dots,N-1\}$, $\|\widehat{\mathbf{z}}_n\|=\lambda$, alors le problème primal associé aux Conditions (4.9) s'énonce ainsi

$$\min_{\mathbf{x}=(\boldsymbol{x}_1,\dots,\boldsymbol{x}_M)\in\mathbb{R}^{M\times N}} \sum_{m=1}^{M}\left(\frac{1}{2}\|\boldsymbol{y}_{\boldsymbol{m}}-\boldsymbol{x}_{\boldsymbol{m}}\|^2 + \sum_{n=1}^{N-1}\widehat{z}_{m,n}|(L\boldsymbol{x}_{\boldsymbol{m}})_n|\right) \tag{4.11}$$

et peut être interprété comme $M$ Problèmes 2.2 ($\ell_1$-TV) ayant des paramètres de régularisation variant dans le temps, c-à-d, $(\widehat{\boldsymbol{z}}_m)_{1\leq m\leq M}$.

L'approximation que nous faisons ici consiste à restreindre l'estimation des valeurs de $\widehat{\mathbf{z}}$ à partir d'un ensemble prédéfini $\mathcal{Q}=\{\zeta^{(1)},\dots,\zeta^{(|\mathcal{Q}|)}\}$ choisi de façon à ce que pour tout $q\in\{1,\dots,|\mathcal{Q}|\}$, $\zeta^{(q)}=(\zeta_m^{(q)})_{1\leq m\leq M}\in\mathbb{R}^M$ satisfasse $\|\zeta^{(q)}\|=\lambda$.

La stratégie la plus naive consisterait donc à résoudre $M$ Problèmes 2.2 ($\ell_1$-TV) pour les $|\mathcal{Q}|$ valeurs candidates de $\widehat{\mathbf{z}}$, c'est-à-dire trouver pour chaque $m=\{1,\dots,M\}$ et $q=\{1,\dots,|\mathcal{Q}|\}$,

$$\widehat{\boldsymbol{x}}_{\boldsymbol{m}}^{(q)} = \arg\min_{\boldsymbol{x}_{\boldsymbol{m}}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y}_{\boldsymbol{m}}-\boldsymbol{x}_{\boldsymbol{m}}\|^2 + \zeta_m^{(q)}\|L\boldsymbol{x}_{\boldsymbol{m}}\|_1 \tag{4.12}$$

et mettre au point une méthode pour choisir la solution parmi les $|\mathcal{Q}|$ possibles. Par exemple, choisir celle qui maximise quelque critère de qualité $f$ que ce soit, i.e.,

$$\widehat{\mathbf{x}} = \widehat{\mathbf{x}}^{(q^*)} \quad \text{où} \quad q^* = \arg\max_{1\leq q\leq|\mathcal{Q}|} f(\widehat{\mathbf{x}}^{(q)}). \tag{4.13}$$

Bien que cette méthode pourrait être mise en oeuvre en parallélisant $M$ algorithmes à la volée résolvant le Problème 2.2 ($\ell_1$-TV), cette situation correspondrait au cas où l'estimée de $\mathbf{z}$ est constante, c-à-d, $\tilde{\mathbf{z}}=\zeta^{(q^*)}\mathbb{1}_N$. Par conséquent, les changements de valeurs moyennes seraient traitées indépendamment sur chacune des composantes et aucune parcimonie de groupe ne serait imposée sur $(L\widehat{\boldsymbol{x}}_{\boldsymbol{1}},\dots,L\widehat{\boldsymbol{x}}_{\boldsymbol{M}})^\top$.

Afin de pouvoir bénéficier à la fois d'une implémentation à la volée et de la parcimonie de groupe, nous proposons une solution algorithmique basée sur un estimateur constant par morceaux de $\widehat{\mathbf{z}}$ détaillé dans la prochaine section.

## 4.3 Solution algorithmique à la volée

Dans la suite, nous commençons par étendre au cadre vectoriel l'algorithme à la volée proposé en [52] tout en supposant que $\widehat{\mathbf{z}}$ est connu *a priori*. Cette hypothèse forte, irréaliste en pratique, nous permet de décrire facilement les étapes de l'algorithme. Puis, nous nous sommes intéressés à la question de l'estimation automatique et à la volée de $\widehat{\mathbf{z}}$ parmi les valeurs présentes dans $\mathcal{Q}$. De fait, nous introduisons par conséquent le paramètre $|\mathcal{Q}|$ contrôlant la qualité de l'approximation. Les principales étapes de l'algorithme à la volée sont résumées dans l'Algorithme 6. Tout comme dans [52], la méthode proposée est également basée sur l'encadrement conjoint des solutions primales et duales $\widehat{\mathbf{x}}$ et $\widehat{\mathbf{u}}$ par des bornes inférieures et supérieures mises à jour au fur et à mesure que l'observation $\mathbf{y}$ est parcourue.

**Algorithme 6** Principe de l'algorithme à la volée pour résoudre le Problème 4.1 ($\ell_{1,2}$-TV).

Fixer $n_0 = 1$.
**Tant que** $n_0 < N$ **faire**
  Fixer $n \leftarrow n_0$
  Initialiser les bornes primales et duales
  **Tant que** Règle 1 est satisfaite **faire**
    Fixer $n \leftarrow n + 1$
    **Pour** $m \in \{1, \ldots, M\}$ **faire**
      Mettre à jour les bornes primales et duales
      **Si** Règle 2 n'est pas satisfaite **alors**
        Recalculer les bornes primales et duales
  Estimer la discontinuité $n_{\text{rupt}}$
  Estimer $(\widehat{\mathrm{x}}_j)_{n_0 \leq j \leq n_{\text{rupt}}}$
  Fixer $n_0 \leftarrow n_{\text{rupt}} + 1$
**Sortie :** Solution $\widehat{\mathbf{x}}_{\text{approx}}$

La conception de l'Algorithme 6 repose sur la spécification de la Règle 1 et de la Règle 2 permettant respectivement de détecter une discontinuité et de trouver sa position grâce à la Proposition 4.2..

### 4.3.1 Cas idéal où $\widehat{\mathbf{z}}$ est connu

#### 4.3.1.1 Bornes inférieures et supérieures

D'après la Proposition 4.2, la solution du problème primal, la solution du problème dual et la variable auxiliaire doivent satisfaire, pour tout $n \in \{0, \ldots, N-1\}$

$$\begin{cases} \widehat{\mathbf{u}}_{n+1} = \mathbf{y}_{n+1} + \widehat{\mathbf{u}}_n - \widehat{\mathbf{x}}_{n+1}, \\ \text{abs}(\widehat{\mathbf{u}}_{n+1}) \leq \widehat{\mathbf{z}}_{n+1}, \\ \|\widehat{\mathbf{z}}_{n+1}\| = \lambda. \end{cases} \tag{4.14}$$

avec $\widehat{\mathbf{u}}_0 = \widehat{\mathbf{u}}_N = 0$. Si l'on considère les deux premières conditions, la condition de prolongation $\widehat{\mathbf{x}}_{n+1} = \widehat{\mathbf{x}}_n$ mène à

$$\begin{cases} \mathbf{y}_{n+1} \geq \widehat{\mathbf{x}}_n - \widehat{\mathbf{z}}_{n+1} - \widehat{\mathbf{u}}_n, \\ \mathbf{y}_{n+1} \leq \widehat{\mathbf{x}}_n + \widehat{\mathbf{z}}_{n+1} - \widehat{\mathbf{u}}_n. \end{cases} \tag{4.15}$$

En suivant le même raisonnement dérivé dans [52] pour le cas scalaire, on peut vérifier si la condition de prolongation (4.15) est satisfaite en raisonnant sur en encadrement de $\widehat{\mathbf{u}}_n$ et $\widehat{\mathbf{x}}_n$.

**Définition 4.1** (Bornes locales inférieures et supérieures de $\widehat{\mathbf{u}}_n$ et $\widehat{\mathbf{x}}_n$)**.** *Pour tout $n \in \{1, \dots, N-1\}$, on définit les bornes inférieures et supérieures de $\widehat{\mathbf{x}}_n$, notées respectivement $\underline{\mathbf{x}}_n$ et $\overline{\mathbf{x}}_n$, telles que*

$$\underline{\mathbf{x}}_n \leq \widehat{\mathbf{x}}_n \leq \overline{\mathbf{x}}_n, \tag{4.16}$$

*et l'on note $\underline{\mathbf{u}}_n$ et $\overline{\mathbf{u}}_n$ comme suit :*

$$(\forall m \in \{1, \dots, M\}) \quad \begin{cases} \widehat{u}_{m,n} = \underline{u}_{m,n} & si \quad \widehat{x}_{m,n} = \underline{x}_{m,n}, \\ \widehat{u}_{m,n} = \overline{u}_{m,n} & si \quad \widehat{x}_{m,n} = \overline{x}_{m,n}, \end{cases} \tag{4.17}$$

*où $\underline{\mathbf{u}}_n$ et $\overline{\mathbf{u}}_n$ sont respectivement la bornes supérieure et inférieure de $\widehat{\mathbf{u}}_n$, i.e.*

$$\overline{\mathbf{u}}_n \leq \widehat{\mathbf{u}}_n \leq \underline{\mathbf{u}}_n, \tag{4.18}$$

*Démonstration.* D'après la relation primale-duale (4.4), pour tout $m \in \{1, \dots, M\}$ et $n \in \{1, \dots, N-1\}$,

$$\widehat{u}_{m,n} = y_{m,n} + \widehat{u}_{m,n-1} - \widehat{x}_{m,n}, \tag{4.19}$$

et par définition des bornes inférieures et supérieures de $\widehat{x}_{m,n}$ et $\widehat{u}_{m,n}$, nous obtenons

$$\underline{u}_{m,n} = y_{m,n} + \widehat{u}_{m,n-1} - \underline{x}_{m,n}, \tag{4.20}$$

$$\overline{u}_{m,n} = y_{m,n} + \widehat{u}_{m,n-1} - \overline{x}_{m,n}. \tag{4.21}$$

En soustrayant (4.20) à (4.19) nous obtenons

$$\widehat{u}_{m,n} - \underline{u}_{m,n} = \underline{x}_{m,n} - \widehat{x}_{m,n} < 0 \tag{4.22}$$

en conséquence de (4.16). De la même manière, on montre que $\widehat{u}_{m,n} > \overline{u}_{m,n}$. □

#### 4.3.1.2 Règles de mise à jour & Règle 1

La condition de prolongation $\widehat{\mathbf{x}}_{n+1} = \widehat{\mathbf{x}}_n$, qui a mené à (4.15), peut être réécrite à l'aide des bornes sur $\widehat{\mathbf{x}}_n$ et $\widehat{\mathbf{u}}_n$ comme suit :

$$\begin{cases} \mathbf{y}_{n+1} \geq \underline{\mathbf{x}}_n - \widehat{\mathbf{z}}_{n+1} - \underline{\mathbf{u}}_n, \\ \mathbf{y}_{n+1} \leq \overline{\mathbf{x}}_n + \widehat{\mathbf{z}}_{n+1} - \overline{\mathbf{u}}_n. \end{cases} \tag{4.23}$$

Si cette dernière condition, appelée Règle 1, est vérifiée, alors d'après la relation primale-duale (4.10), nous pouvons mettre à jour les bornes inférieures et supérieures à la position $n+1$ comme suit :

$$\begin{cases} \underline{\mathbf{u}}_{n+1} = \mathbf{y}_{n+1} + \underline{\mathbf{u}}_n - \underline{\mathbf{x}}_n, \\ \overline{\mathbf{u}}_{n+1} = \mathbf{y}_{n+1} + \overline{\mathbf{u}}_n - \overline{\mathbf{x}}_n, \end{cases} \tag{4.24}$$

et

$$\begin{cases} \underline{\mathbf{x}}_{n+1} = \underline{\mathbf{x}}_n, \\ \overline{\mathbf{x}}_{n+1} = \overline{\mathbf{x}}_n. \end{cases} \tag{4.25}$$

**Remarque 4.3.** *De manière équivalente, on peut systématiquement mettre à jour les bornes primales (resp. duales) selon* (4.24) *(resp.* (4.25)*) puis ensuite vérifier si la réécriture suivante de la condition de prolongation* (4.23) *est vérifiée :*

$$\begin{cases} \underline{\mathbf{u}}_{n+1} \geq -\widehat{\mathbf{z}}_{n+1}, \\ \overline{\mathbf{u}}_{n+1} \leq +\widehat{\mathbf{z}}_{n+1}. \end{cases} \tag{4.26}$$

*C'est le choix que nous adopterons dans la suite de ce chapitre.*

#### 4.3.1.3 Prolongation du signal & Règle 2

Si la Règle 1 (i.e. Condition (4.23) ou de manière équivalente (4.26)) est vérifiée, alors l'hypothèse $\widehat{\mathbf{x}}_{n+1} = \widehat{\mathbf{x}}_n$ est valide. Par contre, les bornes supérieures et inférieures doivent peut-être être mises à jour afin de respecter $\widehat{\mathbf{u}}_{n+1} \in [-\widehat{\mathbf{z}}_{n+1}, +\widehat{\mathbf{z}}_{n+1}]$. D'après (4.18), cette condition nécessite de vérifier que la Règle 2 suivante soit valide :

$$\begin{cases} \underline{\mathbf{u}}_{n+1} \leq +\widehat{\mathbf{z}}_{n+1}, \\ \overline{\mathbf{u}}_{n+1} \geq -\widehat{\mathbf{z}}_{n+1}. \end{cases} \tag{4.27}$$

Pour tout $m \in \{1, \ldots, M\}$, trois configurations peuvent être rencontrées :

— Si les deux Conditions (4.27) sont satisfaites, les bornes sont inchangées.

— Si $\underline{u}_{m,n+1} = \underline{u}_{m,n} + y_{m,n+1} - \underline{x}_{m,n} > +\widehat{z}_{m,n+1}$, alors les règles de mises à jour spécifiées en (4.25) ont sous-évaluées la borne supérieure

$$\underline{\nu}_m \equiv \underline{x}_{m,j} \quad (\forall j \in \{n_0, \ldots, n+1\}) \tag{4.28}$$

où $n_0$ est la position de départ du dernier segment. Puisque $\underline{u}_{m,n+1}$ est délimité supérieurement par $+\widehat{z}_{m,n+1}$ et, que pour une telle valeur on peut montrer que

$$\underline{\nu}_m = \underline{x}_{m,n} + \frac{\underline{u}_{m,n+1} - \widehat{z}_{m,n+1}}{n - n_0 + 1}, \tag{4.29}$$

nous proposons d'effectuer les mises à jour suivantes

$$\begin{cases} (\forall j \in \{n_0, \ldots, n+1\}) \quad \underline{x}_{m,j} = \underline{\nu}_m, \\ \underline{u}_{m,n+1} = +\widehat{z}_{m,n+1}. \end{cases} \tag{4.30}$$

— Si $\overline{u}_{m,n+1} < -\widehat{z}_{m,n+1}$, alors la borne inférieure

$$\overline{\nu}_m \equiv \overline{x}_{m,j} \quad (\forall j \in \{n_0, \ldots, n+1\}) \tag{4.31}$$

a été sur-évaluée. De la même manière, puisque $\overline{u}_{m,n+1}$ est borné inférieurement par $-\widehat{z}_{m,n+1}$, nous pouvons montrer que la borne supérieure

$$\overline{\nu}_m = \overline{x}_{m,n} + \frac{\overline{u}_{m,n+1} + \widehat{z}_{m,n+1}}{n - n_0 + 1}, \tag{4.32}$$

permet d'assurer la cohérence des mises à jour suivantes

$$\begin{cases} (\forall j \in \{n_0, \ldots, n+1\}) \quad \overline{x}_{m,j} = \overline{\nu}_m, \\ \overline{u}_{m,n+1} = -\widehat{z}_{m,n+1}. \end{cases} \tag{4.33}$$

*Démonstration.* Pour tout $m \in \{1, \ldots, M\}$ et $n \in \{n_0, \ldots, N-2\}$, si

$$\underline{u}_{m,n+1} = \underline{u}_{m,n} + y_{m,n+1} - \underline{x}_{m,n} > +\widehat{z}_{m,n+1}, \tag{4.34}$$

alors les règles de mise à jour de $\underline{x}_{m,n}$, spécifiées en (4.25), ont sous-évaluées sa valeur $\underline{\nu}_m$.

D'un côté, afin de modifier les bornes inférieures $(\underline{x}_{m,n})_{k_0 \leq j \leq n+1}$, on considère la somme cumulée des observations qui, selon (4.4), mène à

$$\sum_{j=n_0+1}^{n+1} y_{m,j} = u_{m,n+1} - u_{m,n_0} + (n - n_0 + 1) x_{m,n+1}, \tag{4.35}$$

et qui, si $\underline{u}_{m,n+1} = +\widehat{z}_{m,n+1}$, mènerait à

$$\sum_{j=n_0+1}^{n+1} y_{m,j} = \widehat{z}_{m,n+1} - u_{m,n_0} + (n - n_0 + 1)\underline{\nu}_m, \tag{4.36}$$

par définition de $\underline{x}_{m,n+1} = \underline{\nu}_m$.

D'un autre côté, les règles de mise à jour (4.24) et (4.25) ont menées à

$$\underline{u}_{m,n+1} = \underline{u}_{m,n_0} + \sum_{j=n_0+1}^{n+1} y_{m,j} - (n - n_0 + 1)\underline{x}_{m,n}. \tag{4.37}$$

La combinaison de (4.36) et (4.37) conduit à

$$\underline{\nu}_m = \underline{x}_{m,n} + \frac{-\underline{u}_{m,n_0} + \underline{u}_{m,n+1} - \widehat{z}_{m,n+1} + \widehat{u}_{m,n_0}}{n - n_0 + 1}. \tag{4.38}$$

Puisque $\underline{x}_{m,n}$ a été sous-évaluée et que par définition $\widehat{u}_{m,n_0} \leq \underline{u}_{m,n_0}$, nous pouvons ainsi proposer la valeur suivante

$$\underline{\nu}_m = \underline{x}_{m,n} + \frac{\underline{u}_{m,n+1} - \widehat{z}_{m,n+1}}{n - n_0 + 1}, \tag{4.39}$$

afin d'ajuster les bornes inférieures, i.e.,

$$(\forall j \in \{n_0, \ldots, n+1\}) \quad \underline{x}_{m,j} = \underline{\nu}_m. \tag{4.40}$$

En plus, en conséquence de $\widehat{u}_{m,n+1} \in [-\widehat{z}_{m,n+1}, +\widehat{z}_{m,n+1}]$ et d'après l'inégalité (4.18), nous assignons

$$\underline{u}_{m,n+1} = +\widehat{z}_{m,n+1}. \tag{4.41}$$

La démonstration est similaire dans le cas $\overline{u}_{m,n+1} < -\widehat{z}_{m,n+1}$. □

#### 4.3.1.4 Estimer la position de la discontinuité $n_{\text{rupt}}$

Si la Règle 1 n'est pas vérifiée à la position $n$, une discontinuité doit être introduite dans l'intervalle $\{n_0, \ldots, n\}$. Afin d'estimer sa position $n_{\text{rupt}}$, nous avons besoin de distinguer trois cas pour tout $m \in \{1, \ldots, M\}$ :

— Si $\underline{u}_{m,n+1} = \underline{u}_{m,n} + y_{m,n+1} - \underline{x}_{m,n} < -\widehat{z}_{m,n+1}$, alors, puisque $\underline{u}_{m,n}$ est borné, cela implique que $\underline{x}_{m,n}$ est surévalué et par conséquent un changement d'amplitude négatif doit être introduit sur la $m$-ème composante afin de décroître sa valeur. A partir de la Proposition 4.2 et de l'Equation (4.18), l'ensemble des positions $\kappa_m$ appropriées pour introduire une discontinuité sur la $m$-ème composante est

$$\kappa_m = \{j \in \{n_0, \ldots, n\} \quad | \quad \underline{u}_{m,j} = +\widehat{z}_{m,j}\}. \tag{4.42}$$

Ces positions correspondent aux indices $j$ pour lesquels la borne $\underline{u}_{m,j}$ a été mise à jour afin de vérifier la condition $\widehat{u}_{m,j} \in [-\widehat{z}_{m,j}, \widehat{z}_{m,j}]$ (voir paragraphe précédent).

— Si $\overline{u}_{m,n+1} > +\widehat{z}_{m,n+1}$, alors un changement d'amplitude positif doit être introduit sur la $m$-ème composante. L'ensemble des positions $\kappa_m$ appropriées pour introduire une discontinuité sur la $m$-ème composante est

$$\kappa_m = \{j \in \{n_0, \ldots, n\} \quad | \quad \overline{u}_{m,j} = -\widehat{z}_{m,j}\}. \tag{4.43}$$

Ces positions correspondent aux indices $j$ pour lesquels la borne $\overline{u}_{m,j}$ a été mise à jour afin de vérifier la condition $\widehat{u}_{m,j} \in [-\widehat{z}_{m,j}, \widehat{z}_{m,j}]$.

— Sinon, la $m$-ème composante satisfait (4.15) et aucun changement ne doit être introduit. Cependant, puisque l'on considère le cas où les discontinuités ont lieu simultanément sur toutes les composantes, on définit $\kappa_m = \{n_0, \dots, n\}$.

La position de la discontinuité $n_{\text{rupt}}$ correspond à la dernière position pour laquelle on peut introduire le changement d'amplitude adéquat sur chacune des composantes, i.e.,

$$n_{\text{rupt}} = \max_{j \in \cap_{m=1}^{M} \kappa_m} j. \tag{4.44}$$

Une fois que $n_{\text{rupt}}$ a été estimé, nous pouvons définir la valeur $(\widehat{\mathbf{x}}_j)_{n_0 \leq j \leq n_{\text{rupt}}}$. Lorsqu'un changement d'amplitude négatif est détecté sur la $m$-ème composante, on assigne

$$(\forall j \in \{n_0, \dots, n_{\text{rupt}}\}) \quad \widehat{x}_{m,j} = \underline{x}_{m,n+1}, \tag{4.45}$$

en accord avec (4.17). De la même manière, lorsqu'un changement d'amplitude positif est détecté, on assigne

$$(\forall j \in \{n_0, \dots, n_{\text{rupt}}\}) \quad \widehat{x}_{m,j} = \overline{x}_{m,n+1}. \tag{4.46}$$

#### 4.3.1.5 Commencer un nouveau segment

Lorsqu'un segment est créé, on commence la détection d'un nouveau segment à partir de $n_0 = n_{\text{rupt}} + 1$ tant que $n_0 < N$.

D'après (4.4) et par définition des bornes, pour tout $n \in \{1, \dots, N\}$

$$\begin{cases} \underline{\mathbf{x}}_n = \mathbf{y}_n - \underline{\mathbf{u}}_n + \widehat{\mathbf{u}}_{n-1}, \\ \overline{\mathbf{x}}_n = \mathbf{y}_n - \overline{\mathbf{u}}_n + \widehat{\mathbf{u}}_{n-1}. \end{cases} \tag{4.47}$$

En particulier, pour $n = n_0$, combiner (4.9), (4.16), (4.17) et (4.24) nous permet de dériver la procédure d'initialisation suivante

$$\begin{cases} \underline{\mathbf{u}}_{n_0} = +\widehat{\mathbf{z}}_{n_0}, \\ \overline{\mathbf{u}}_{n_0} = -\widehat{\mathbf{z}}_{n_0}, \\ \underline{\mathbf{x}}_{n_0} = \mathbf{y}_{n_0} - \widehat{\mathbf{z}}_{n_0} + \widehat{\mathbf{u}}_{n_0-1}, \\ \overline{\mathbf{x}}_{n_0} = \mathbf{y}_{n_0} + \widehat{\mathbf{z}}_{n_0} + \widehat{\mathbf{u}}_{n_0-1}, \end{cases} \tag{4.48}$$

où la valeur de $\widehat{\mathbf{u}}_{n_0-1}$ est donnée par la Proposition 4.2. De plus, d'après la réécriture de (4.14), $\widehat{\mathbf{u}}_0 = 0$.

### 4.3.2 Estimation du vecteur auxiliaire $\widehat{\mathbf{z}}$

Afin de présenter les étapes de l'algorithme, nous avons jusqu'à présent supposé $\widehat{\mathbf{z}}$ connu *a priori*. Dans cette section, nous montrons comment modifier l'algorithme pour estimer simultanément $\widehat{\mathbf{z}}$ et $\widehat{\mathbf{x}}$.

Pour obtenir une solution approchée à la volée, nous proposons de :

— construire un estimateur constant par morceaux $\widetilde{\mathbf{z}}$ de $\widehat{\mathbf{z}}$,

— considérer seulement les discontinuités conjointes sur toutes les composantes $m \in \{1, \dots, M\}$.

#### 4.3.2.1 Estimateur constant par morceaux de $\widehat{\mathbf{z}}$

Nous proposons un estimateur constant par morceaux dont les valeurs entre chaque discontinuités appartiennent à un ensemble prédéfini $\mathcal{Q}$ introduit à la Section 4.2.3. Pour chaque valeur candidate $\zeta^{(q)}$ où $q \in \{1, \dots, |\mathcal{Q}|\}$, on associe des bornes supérieures et inférieures notées $\underline{\mathbf{u}}_n^{(q)}$, $\overline{\mathbf{u}}_n^{(q)}$, $\underline{\mathbf{x}}_n^{(q)}$, et $\overline{\mathbf{x}}_n^{(q)}$. Elles sont initialisées à chaque position de départ $n_0$ d'un nouveau segment et mises à jour indépendamment pour chaque $q \in \{1, \dots, |\mathcal{Q}|\}$ selon (4.24) et (4.25) jusqu'a ce que la condition de prolongation

$$\begin{cases} \underline{\mathbf{u}}_{n+1}^{(q)} \geq -\zeta^{(q)}, \\ \overline{\mathbf{u}}_{n+1}^{(q)} \leq +\zeta^{(q)}, \end{cases} \tag{4.49}$$

basée sur (4.26), ne soit plus vérifiée. Dans la suite, nous examinons comment modifier l'algorithme décrit à la Section 4.3.1, afin de prendre en compte la sélection automatique de $\widetilde{\mathbf{z}}$ parmi $\mathcal{Q}$. L'algorithme qui en résulte est l'Algorithme 7.

#### 4.3.2.2 Estimer les positions candidates de la discontinuité $n_{\text{rupt}}^{(q)}$

Pour chaque $q \in \{1, \dots, |\mathcal{Q}|\}$, on estime les positions des discontinuités $n_{\text{rupt}}^{(q)}$ tel que nous l'avons précédemment décrit à la Section 4.3.1.4. Toutefois, la principale différence ici est que nous nous restreignons à des discontinuités simultanées sur toutes les composantes. Nous avons déjà détaillé avant la Proposition 4.1 que la création de discontinuités non-simultanées avait une probabilité quasi nulle de se produire. Dès lors, la restriction aux discontinuités simultanées ne va pas impacter la solution.

Par conséquent, pour chaque $q \in \{1, \dots, |\mathcal{Q}|\}$, dès qu'au moins une composante $m \in \{1, \dots, M\}$ ne vérifie pas la condition de prolongation (4.49), une discontinuité doit être créee. Par exemple, si $\underline{u}_{m,n+1}^{(q)} < -\zeta_m^{(q)}$ (resp. $\overline{u}_{m,n+1}^{(q)} > \zeta_m^{(q)}$), alors l'ensemble des positions $\kappa_m^{(q)}$ appropriées pour introduire une discontinuité sur la $m$-ème composante est

$$\kappa_m^{(q)} = \{j \in \{n_0, \dots, n\} \mid \underline{u}_{m,j}^{(q)} = +\zeta_m^{(q)}\} \tag{4.50}$$

$$(\text{resp.} \quad \kappa_m^{(q)} = \{j \in \{n_0, \dots, n\} \mid \overline{u}_{m,j}^{(q)} = -\zeta_m^{(q)}\}). \tag{4.51}$$

Pour les composantes vérifiant (4.49), nous proposons l'heuristique suivante pour déterminer le type de discontinuité et les positions de discontinuité le plus probable : $\forall m_- \neq m$ tel que $\underline{u}_{m_-,n+1}^{(q)} + \overline{u}_{m_-,n+1}^{(q)} < 0$, alors

$$\kappa_{m_-}^{(q)} = \{j \in \{n_0, \dots, n\} \mid \underline{u}_{m_-,j}^{(q)} = +\zeta_m^{(q)}\} \tag{4.52}$$

et, $\forall m_+ \neq m$ tel que $\underline{u}_{m_+,n+1}^{(q)} + \overline{u}_{m_+,n+1}^{(q)} \geq 0$, alors

$$\kappa_{m_+}^{(q)} = \{j \in \{n_0, \dots, n\} \mid \underline{u}_{m_+,j}^{(q)} = +\zeta_m^{(q)}\}. \tag{4.53}$$

Un exemple pour $M = 2$ et pour lequel la seconde composante viole la Condition (4.49) est illustré Figure 4.3. La position de la discontinuité $n_{\text{rupt}}^{(q)}$ et la valeur assignée à $\widehat{\mathbf{x}}^{(q)}$ sur le segment en cours sont toutes deux estimées comme précédemment, voir (4.44), (4.45) et (4.46).

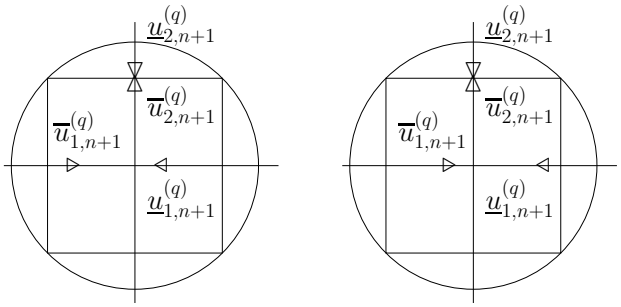



FIGURE 4.3 – **Exemple de configurations menant à la détection d'une discontinuité**. Dans cet exemple $M = 2$, $\zeta_1^{(q)} = \zeta_2^{(q)} = \lambda/\sqrt{2}$. Puisque $\underline{u}^{(q)}_{2,n+1} > \zeta_2^{(q)}$, la condition (4.49) est violée. Le graphique de gauche (resp. droite) illustre la configuration $\overline{u}^{(q)}_{1,n+1} + \underline{u}^{(q)}_{1,n+1} < 0$ (resp. $\overline{u}^{(q)}_{1,n+1} + \underline{u}^{(q)}_{1,n+1} \geq 0$) décrite à la Section 4.3.2.2.

#### 4.3.2.3 Estimer la position de la discontinuité $n_{\text{rupt}}$.

D'après le paragraphe précédent, la procédure d'estimation mène à plusieurs positions candidates pour la discontinuité (au plus $|\mathcal{Q}|$). Parmi ces différentes positions, nous proposons de choisir celle indexée par $q^*$ pour laquelle les bornes $\underline{\mathbf{x}}^{(q^*)}$ et $\overline{\mathbf{x}}^{(q^*)}$ sont le plus proches possibles, i.e.,

$$q^* \in \underset{1\leq q\leq|\mathcal{Q}|}{\text{Argmin}} \left\| \left( \overline{\mathbf{x}}^{(q)}_{n^{(q)}_{\text{rupt}}} - \underline{\mathbf{x}}^{(q)}_{n^{(q)}_{\text{rupt}}} \right) \boldsymbol{\sigma}^{-1} \right\|^2, \tag{4.54}$$

avec

$$\boldsymbol{\sigma} = \text{diag}(\sigma_1, \ldots, \sigma_M), \tag{4.55}$$

où, pour tout $m \in \{1, \ldots, M\}$, $\sigma_m$ est l'écart-type de $\boldsymbol{y}_m$. Le facteur $\boldsymbol{\sigma}^{-1}$ permet d'assurer que toutes les composantes contribuent également dans le critère (4.54) indépendamment de leur ordre de grandeur. Lorsque le minimiseur de (4.54) n'est pas unique, nous choisissons l'indice $q^*$ pour lequel $n^{(q^*)}_{\text{rupt}}$ est le plus élevé. En d'autre termes, nous choisissons l'ensemble de variable auxiliaires permettant de valider la condition de prolongation (4.49) le plus longtemps possible.

Par conséquent, nous obtenons finalement un indice $q^*$ qui nous permet d'estimer $n_{\text{rupt}} = n^{(q^*)}_{\text{rupt}}$ et,

$$(\forall j \in \{n_0, \ldots, n_{\text{rupt}}\}) \quad \tilde{\mathbf{z}}_j = \zeta^{(q^*)},\ \widehat{\mathbf{x}}_j = \widehat{\mathbf{x}}_j^{(q^*)}. \tag{4.56}$$

La position de départ du prochain segment est fixé à $n_0 = n_{\text{rupt}} + 1$, et l'algorithme se répète tant que $n_0 < N$.

#### 4.3.2.4 Commencer un nouveau segment

Considérons la position de départ $n_0$ d'un nouveau segment. Pour tout $q \in \{1, \ldots, |\mathcal{Q}|\}$, l'étape d'initialisation peut se réécrire

$$\begin{cases} \underline{\mathbf{u}}^{(q)}_{n_0} = +\zeta^{(q)}, & \overline{\mathbf{u}}^{(q)}_{n_0} = -\zeta^{(q)}, \\ \underline{\mathbf{x}}^{(q)}_{n_0} = \mathbf{y}_{n_0} - \zeta^{(q)} + \widehat{\mathbf{u}}_{n_0-1}, & \overline{\mathbf{x}}^{(q)}_{n_0} = \mathbf{y}_{n_0} + \zeta^{(q)} + \widehat{\mathbf{u}}_{n_0-1}, \end{cases} \tag{4.57}$$

avec $\widehat{\mathbf{u}}_0 = 0$.

**Remarque 4.4.** *L'étape d'initialisation* (4.57) *dépend implicitement de l'estimation de* $\widehat{\mathbf{z}}$ *faite sur le segment précédent à travers le terme* $\widehat{\mathbf{u}}_{n_0-1}$. *Nos simulations ont montrées que*

**Algorithme 7** Résolution approchée du Problème 3.1 basée sur un algorithme à la volée.

**Entrée :** Ensemble $\mathcal{Q} = \{\zeta^{(1)}, \dots, \zeta^{(|\mathcal{Q}|)}\}$.

Fixer $n_0 = 1$.
**Tant que** $n_0 < N$ **faire**
  **Pour** $q = 1, \dots, |\mathcal{Q}|$ **faire**
    Fixer $n \leftarrow n_0$
    Initialiser les bornes primales et duales selon (4.58)
    **Tant que** (4.49) est satisfaite **faire**
      Fixer $n \leftarrow n + 1$
      **Pour** $m = 1, \dots, M$ **faire**
        Mettre à jour les bornes primales et duales
        **Si** $\underline{u}^{(q)}_{m,n+1} > +\zeta^{(q)}_m$ or $\overline{u}^{(q)}_{m,n+1} < -\zeta^{(q)}_m$ **alors**
          Recalculer les bornes primales et duales
    Estimer les valeurs candidates de la position de la discontinuité $n^{(q)}_{\text{rupt}}$ et de valeur prise sur chaque segment $(\widehat{\mathbf{x}}^{(q)}_j)_{n_0 \leq j \leq n^{(q)}_{\text{rupt}}}$ selon la Section 4.3.2.2
  Estimer la discontinuité $n_{\text{rupt}} \in (n^{(q)}_{\text{rupt}})_{1 \leq q \leq |\mathcal{Q}|}$ selon la Section 4.3.2.3
  Estimer $(\widehat{\mathbf{x}}_j)_{n_0 \leq j \leq n_{\text{rupt}}}$ et $(\tilde{\mathbf{z}}_j)_{n_0 \leq j \leq n_{\text{rupt}}}$ selon (4.56)
  Fixer $n_0 \leftarrow n_{\text{rupt}} + 1$

**Sortie :** Solution $\widehat{\mathbf{x}}$

*l'initialisation* (4.57) *peut mener à une solution inconsistente* $\widehat{\mathbf{x}}$ *dès que* $\widehat{\mathbf{z}}$ *est mal estimé sur un segment. Empiriquement, une meilleure approximation de la solution itérative est obtenue si chaque segment est traité indépendamment, i.e.,*

$$\begin{cases} \underline{\mathbf{u}}^{(q)}_{n_0} = +\zeta^{(q)}, & \overline{\mathbf{u}}^{(q)}_{n_0} = -\zeta^{(q)}, \\ \underline{\mathbf{x}}^{(q)}_{n_0} = \mathbf{y}_{n_0} - \zeta^{(q)}, & \overline{\mathbf{x}}^{(q)}_{n_0} = \mathbf{y}_{n_0} + \zeta^{(q)}. \end{cases} \tag{4.58}$$

## 4.4 Résultats

### 4.4.1 Configuration expérimentale

Dans cette section résultat, sauf mention contraire, nous considérons que les données sont $M$ signaux constants par morceaux $\overline{\mathbf{x}} \in \mathbb{R}^{M \times N}$ (trait plein noir) partageant les mêmes discontinuités, auxquels sont ajoutés des bruits blancs gaussiens centrés : $\mathbf{y} = \overline{\mathbf{x}} + \boldsymbol{b} \in \mathbb{R}^{M \times N}$.

Le signal $\overline{\mathbf{x}}$ est généré comme suit. Premièrement, la longueur de chaque segment est tiré selon une loi normale repliée $\mathcal{N}(12.5, 16.25)$. Puis, pour chaque $m \in \{1, \dots, M\}$, le changement d'amplitude pour chaque discontinuité est tiré indépendamment selon une distribution gaussienne $\mathcal{N}(2, 0.4)$.

Le minimiseur exact du Problème 3.1, noté $\widehat{\mathbf{x}}$, est calculé à partir l'Algorithme 5. Les itérations sont arrêtées dès que la différence relative de critère est plus faible que $10^{-10}$. La solution proposée est quant à elle notée $\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}$ pour indiquer qu'elle a été calculée à partir de l'ensemble prédéfini $\mathcal{Q}$.

Dans un second ensemble de simulations (voir 4.4.4), l'algorithme à la volée proposé sera comparé à une solution à la volée de l'Algorithme 5 itératif ADMM.

### 4.4.2 Conception de $\mathcal{Q}$

Nous proposons de comparer les solutions $\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}$ obtenues pour deux ensembles différents $\mathcal{Q} = \{\zeta^{(1)}, \dots, \zeta^{(|\mathcal{Q}|)}\}$ dans le cadre bivarié (i.e., $M = 2$) pour $N = 10^4$. Pour les deux solu-

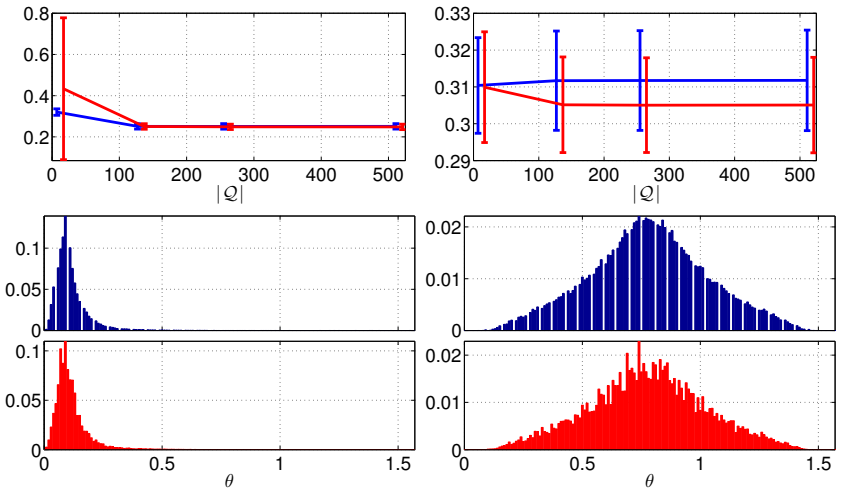



FIGURE 4.4 – **Influence du choix de $\mathcal{Q}$.** Deux conceptions différentes de $\mathcal{Q}$ sont comparées sur 100 réalisations. La première consiste à recouvrir la boule $\ell_2$ de manière homogène alors que la seconde est un recouvrement aléatoire (rouge). Deux configurations expérimentales sont examinées selon que $\boldsymbol{y_1}$ est un ordre de grandeur plus grand que $\boldsymbol{y_2}$ (gauche) ou si $\boldsymbol{y_1}$ et $\boldsymbol{y_2}$ sont du même ordre de grandeur (droite). Haut : $\mathrm{MSE}(\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}}, \widehat{\mathbf{x}})$ en fonction de $|\mathcal{Q}|$. Bas (2ème et 3ème lignes) : distributions de $\theta_{q^*}$ où $q^*$ est l'indice sélectionné par le critère (4.54) pour $|\mathcal{Q}| = 127$.

tions, nous adopterons la paramétrisation suivante

$$(\forall q \in \{1, \ldots, |\mathcal{Q}|\}) \quad \zeta^{(q)} = (\lambda \cos(\theta_q), \lambda \sin(\theta_q)) \tag{4.59}$$

avec $\theta_q \in [0, \pi/2]$ à préciser.

La première solution est obtenue en choisissant un recouvrement homogène de l'ensemble $[0, \pi/2]$. Soit un entier positif $R \in \mathbb{N}_*$, on définit alors $\theta_q = q\pi/2^{R+1}$ et $|\mathcal{Q}| = \sum_{q'=0}^{R-1} 2^{q'}$.

La seconde solution est obtenue pour un ensemble $\mathcal{Q}$ de même taille mais où les valeurs $(\theta_q)_{1 \leq q \leq |\mathcal{Q}|}$ sont tirées uniformément sur l'ensemble $[0, \pi/2]$.

Deux configurations expérimentales sont examinées. Dans la première, $\boldsymbol{y_1}$ est un ordre de grandeur plus grand que $\boldsymbol{y_2}$ (Fig. 4.4, à gauche) alors que dans la seconde, les deux sont du même ordre de grandeur (Fig. 4.4, à droite).

Les performances d'estimation en terme d'erreur quadratique moyenne $\mathrm{MSE}(\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}}, \widehat{\mathbf{x}}) = \widehat{\mathbb{E}}[\frac{1}{N}\|\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}} - \widehat{\mathbf{x}}\|^2]$ (où $\widehat{\mathbb{E}}$ est l'estimateur empirique de la moyenne calculé sur 100 réalisations) sont rapportées Figure. 4.4 (première ligne). Les résultats montrent qu'un recouvrement aléatoire de l'ensemble $[0, \pi/2]$ donne des résultats équivalent à un recouvrement homogène jusqu'à la limite de $|\mathcal{Q}|$ faible.

Sur les 2ème et 3ème lignes, les distributions de $\theta_{q^*}$, où $q^*$ est l'indice sélectionné par le critère (4.54), sont rapportées pour $|\mathcal{Q}| = 127$. Ces histogrammes démontrent l'impact de l'amplitude relative des composantes de $\mathbf{y}$ sur la distribution $\theta_{q^*}$. Si les composantes sont du même ordre de grandeur alors la distribution est symétrique alors qu'elle est asymétrique pour des composantes avec des ordres de grandeur différents.

**Remarque 4.5.** *Si l'on prend par exemple la Figure 4.4 (à droite), il apparait plus sensé de générer $\theta_q$ selon une distribution gaussienne plutôt que selon une distribution uniforme. Par conséquent, si l'on possède des connaissances* a priori *concernant l'amplitude des composantes*

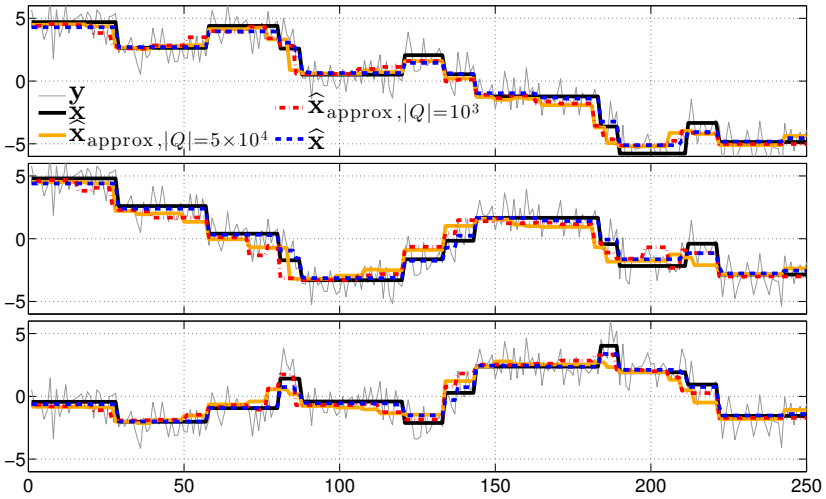



FIGURE 4.5 – **Impact qualitatif de $|\mathcal{Q}|$ sur $\widehat{\mathbf{x}}_{\mathbf{approx},\mathcal{Q}}$.** Pour des raisons de visibilité, seules 3 composantes parmi les $M = 10$ sont affichées. SNR = 4dB. $\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}$ pour $|\mathcal{Q}| = 5 \times 10^4$ est une solution plus satisfaisante que $|\mathcal{Q}| = 10^3$ puisqu'elle a davantage de discontinuités en commun avec $\widehat{\mathbf{x}}$.

*de* $\mathbf{y}$*, celles-ci peuvent être incorporées pour mieux concevoir* $\mathcal{Q}$*. Cela permet également de décroître le temps d'exécution discuté à la Section 4.4.4.*

Dans la suite, nous nous restreignons au cas où $\mathcal{Q}$ est obtenu par recouvrement aléatoire de la boule $\ell_2$ de rayon $\lambda$.

### 4.4.3 Performances hors ligne

Dans cette section, nous nous intéresserons aux performances hors ligne de la solution proposée, c'est-à-dire sans prendre en compte son caractère à la volée. Nous évaluerons les performances pour $M = 10$ et pour deux rapports signal sur bruit : 4dB et 10dB.

**Impact qualitatif de $|\mathcal{Q}|$ sur $\widehat{\mathbf{x}}_{\mathbf{approx},\mathcal{Q}}$.** Pour une réalisation de bruit donné, $\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}$ et $\widehat{\mathbf{x}}$ sont tracées Figure. 4.5 pour $\lambda = 29$, ajusté afin d'obtenir le meilleur rendu visuel. La solution $\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}$ pour $|\mathcal{Q}| = 5 \times 10^4$ (traits pleins orange clair) fourni visuellement une meilleure approximation de $\widehat{\mathbf{x}}$ (traits pointillés bleus) que pour $|\mathcal{Q}| = 10^3$ (traits mixtes rouges).

**Performances d'estimation $\widehat{\mathbf{x}}_{\mathbf{approx},\mathcal{Q}}$ vs. $\widehat{\mathbf{x}}$.** La qualité de l'approximation est davantage quantifiée Figure 4.6 en termes de $\text{MSE}(\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}, \widehat{\mathbf{x}})$ en fonction $\lambda$ pour différents $|\mathcal{Q}|$. Les résultats montrent que la MSE décroît systématiquement lorsque $|\mathcal{Q}|$ augmente. De plus, pour les exemples présentés ici et selon $\lambda$, utiliser $|\mathcal{Q}| \geq 10^4$ n'améliore pas significativement la qualité de la solution, ce qui montre que la sélection de $|\mathcal{Q}|$ ne requiert pas de procédure de réglage compliquée.

**Performances d'estimation $\widehat{\mathbf{x}}$ vs. $\overline{\mathbf{x}}$ et $\widehat{\mathbf{x}}_{\mathbf{approx},\mathcal{Q}}$ vs. $\overline{\mathbf{x}}$.** Comparons à présent la qualité absolue des solutions vis-à-vis du signal non bruité $\overline{\mathbf{x}}$. $\text{MSE}(\widehat{\mathbf{x}}, \overline{\mathbf{x}})$ et $\text{MSE}(\widehat{\mathbf{x}}_{\text{approx},\mathcal{Q}}, \overline{\mathbf{x}})$ pour différents $|\mathcal{Q}|$, sont rapportées Figure 4.7. Les MSEs obtenues sont cohérentes avec le paragraphe précédent. Elles montrent qu'augmenter $|\mathcal{Q}|$ jusqu'à un certain seuil permet de décroître significativement la MSE. Cependant, $\widehat{\mathbf{x}}$ possède une erreur d'estimation plus faible

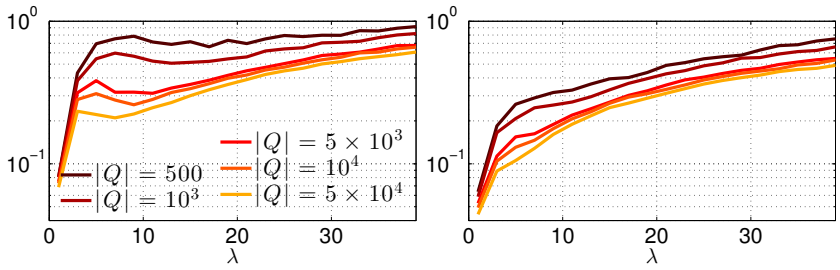



FIGURE 4.6 – **Performances d'estimation $\widehat{\mathbf{x}}_{\mathbf{approx},\mathcal{Q}}$ vs $\widehat{\mathbf{x}}$.** $\mathrm{MSE}(\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}}, \widehat{\mathbf{x}})$ pour différents $|\mathcal{Q}|$. Le SNR est fixé 4dB (resp. 10dB) sur le graphique de gauche (resp. droite).

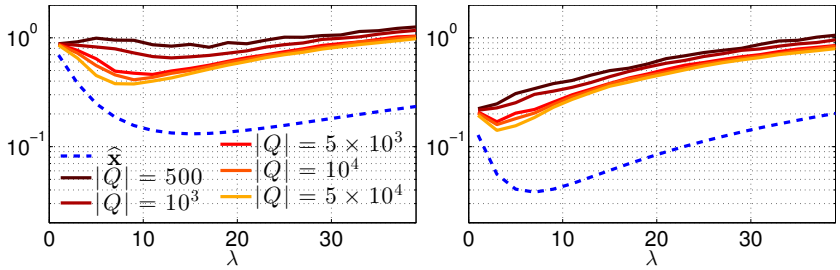



FIGURE 4.7 – **Performances d'estimation $\widehat{\mathbf{x}}$ vs. $\overline{\mathbf{x}}$ and $\widehat{\mathbf{x}}_{\mathbf{approx},Q}$ vs. $\overline{\mathbf{x}}$.** $\mathrm{MSE}(\widehat{\mathbf{x}}, \overline{\mathbf{x}})$ et $\mathrm{MSE}(\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}}, \overline{\mathbf{x}})$ pour différents $|\mathcal{Q}|$. Le SNR est fixé à 4dB (resp. 10dB) sur le graphique de gauche (resp. droite).

que $\widehat{\mathbf{x}}_{\mathrm{approx},Q}$.

### 4.4.4 Performances en ligne

Dans cette section, nous nous intéressons cette fois-ci à la comparaison en ligne de deux solutions. La première est dérivée à partir de l'algorithme à la volée proposé précédemment alors que la seconde est obtenue via l'Algorithme 5 itératif ADMM.

La comparaison est effectuée pour différentes valeurs de $\lambda$ pour un signal $\overline{\mathbf{x}} \in \mathbb{R}^{M \times N}$ ($N = 400$) auquel a été ajouté un bruit gaussien tel que SNR = 3dB. Les performances sont fournies pour $M = 2$ et $M = 5$ composantes.

**Solution proposée en ligne $\widehat{\mathbf{x}}_{\mathrm{online},\mathcal{Q}}$.** Au fur et à mesure que la position $n$ augmente, $\widehat{\mathbf{x}}_{\mathrm{approx},\mathcal{Q}}$ n'est calculé seulement jusqu'à la dernière position $n_0$ et l'algorithme n'a pas encore fourni de solution sur l'ensemble $\{n_0 + 1, \ldots, n\}$. C'est en ce sens que la solution est dite *à la volée*. Cependant, nous pouvons également obtenir une approximation *en ligne* de $\overline{\mathbf{x}}$ jusqu'à $n$, notée $\widehat{\mathbf{x}}_{\mathrm{online},\mathcal{Q}}$, en choisissant d'imposer à la solution des conditions limites en $n$, i.e., $(\forall q \in \{1, \ldots, |\mathcal{Q}|\})\ \widehat{\mathbf{u}}_n^{(q)} = 0$.

**Solution itérative sur une fenêtre glissante $\widehat{\mathbf{x}}_{\mathrm{win},K}$.** Nous considérons une version en ligne naive de l'Algorithme 5 où pour chaque pas de temps $n$, une solution $\widehat{\mathbf{x}}_{\mathrm{win},K}$ est calculée en optimisant sur les $K$ échantillons précédents. Le choix de $K$ est critique. D'un côté, si $K$ est trop faible, alors l'observateur peut manquer la détection de discontinuités. De l'autre côté, si $K$ est trop grand, alors le temps d'exécution devient trop élevé pour obtenir une solution en temps réel. Dans la suite, nous comparons l'influence de trois tailles de fenêtre, respectivement $K = 20$, 50 et 80.

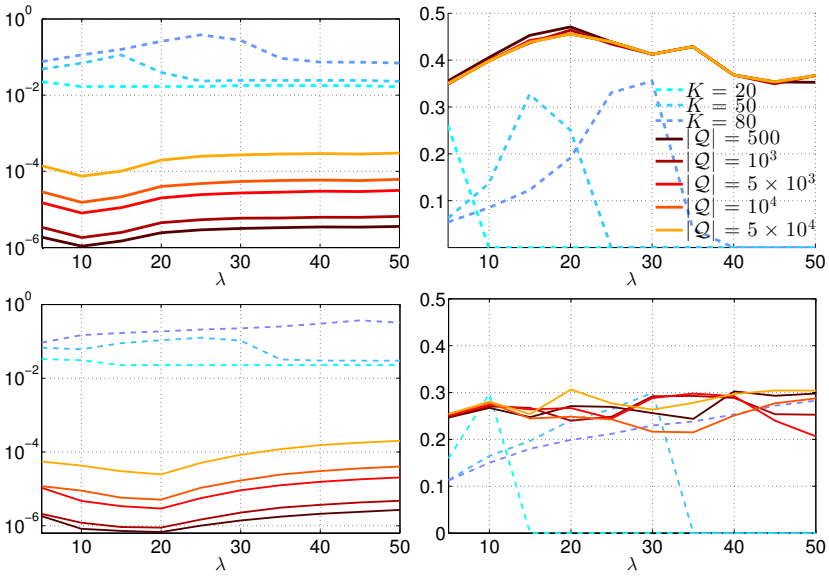



FIGURE 4.8 – **Performances en ligne.** La solution proposée $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$ est affichée en traits pleins alors que la solution ADMM « en ligne » $\widehat{\mathbf{x}}_{\text{win},K}$ est affichée en traits pointillés. Les performances pour $M = 2$ (resp. $M = 5$) sont illustrées sur la figure du haut (resp. bas). Gauche : temps de calcul médian par échantillon (en secondes). Droite : $J(\widehat{\boldsymbol{r}}_{\text{win},K}, \boldsymbol{r})$ et $J(\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}, \boldsymbol{r})$ pour différentes valeurs de $|\mathcal{Q}|$ et $K$.

**Temps d'exécution.** Les comparaisons des temps d'exécutions médian par échantillon (en secondes), calculées sur 10 réalisations de bruit, sont rapportées Figure 4.8 (à gauche) en fonction de $\lambda$. Comme attendu, nous observons que le temps d'exécution augmente avec la taille de $\mathcal{Q}$. Par conséquent, le choix de $|\mathcal{Q}|$ permet de balancer le compromis entre le temps d'exécution et la MSE. Toutefois, le temps d'exécution de $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$ est plusieurs ordres de grandeurs plus faible que celui associé à la solution ADMM en ligne. Il est intéressant de remarquer que les temps de calcul pour $M = 2$ (haut gauche) et $M = 5$ (bas gauche) sont comparables. Nous avons également examiné l'impact d'une initialisation à chaud sur le temps de calcul de la solution ADMM en ligne, et nous avons obtenu que les temps affichés Figure 4.8 ne sont réduits que d'un facteur deux.

Le temps de calcul de $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$ pourrait encore être réduit de plusieurs manières. Premièrement, l'ensemble $\mathcal{Q}$ pourrait être conçu selon les connaissances *a priori* sur la dynamique de chaque composante (voir 4.4.2). Deuxièmement, on pourrait mettre à profit la forme séparable de l'algorithme pour calculer les solutions $\widehat{\mathbf{x}}^{(q)}$ en parallèle pour chaque $q \in \{1, \ldots, |\mathcal{Q}|\}$.

**Précision de la détection des discontinuités.** L'indice de Jaccard (voir Définition 2.7) est utilisé pour mesurer la similarité entre les positions des discontinuités de $\overline{\mathbf{x}}$ et celles obtenues pendant le calcul en ligne de $\widehat{\mathbf{x}}_{\text{win},K}$ et $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$. Pour cela, nous considérons le vecteur d'indicatrices des discontinuités $\overline{\boldsymbol{r}} = (r_n)_{1 \leq n \leq N-1}$ de $\overline{\mathbf{x}}$ (ainsi que $\widehat{\boldsymbol{r}}_{\text{win},K}$ et $\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}$ respectivement associés à $\widehat{\mathbf{x}}_{\text{win},K}$ et $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$) (voir Définition 2.4). Afin d'incorporer un niveau de tolérance sur la localisation des discontinuités, $\overline{\boldsymbol{r}}$, $\widehat{\boldsymbol{r}}_{\text{win},K}$ et $\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}$ sont d'abord convolués avec un filtre passe-bas gaussien à support compact de taille 10 et d'écart-type 3.

$J(\widehat{\boldsymbol{r}}_{\text{win},K}, \overline{\boldsymbol{r}})$ et $J(\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}, \overline{\boldsymbol{r}})$ sont moyennés sur 10 réalisations de bruit et illustrés Figure 4.8 (right plots) en fonction de $\lambda$ pour différents $|\mathcal{Q}|$ et différentes tailles de fenêtre $K$.

Les performances montrent que $J(\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}, \overline{\boldsymbol{r}}) \geq J(\widehat{\boldsymbol{r}}_{\text{win},K}, \overline{\boldsymbol{r}})$ pour pratiquement tous les $\lambda$ et $|\mathcal{Q}|$. Par conséquent, $\widehat{\mathbf{x}}_{\text{online},\mathcal{Q}}$ permet une meilleure détection en ligne des discontinuités de $\overline{\mathbf{x}}$. Les résultats montrent également que $J(\widehat{\boldsymbol{r}}_{\text{online},\mathcal{Q}}, \overline{\boldsymbol{r}})$ ne varie pas significativement avec $|\mathcal{Q}|$ mais décroît légèrement avec $M$. En effet, plus $M$ est élevé, plus la condition de prolongation (4.49) est susceptible d'être violée, créant ainsi davantage de discontinuités.

## 4.5 Conclusion et perspectives

Dans ce chapitre, nous avons développé le premier algorithme $\ell_{1,2}$-TV à la volée pour la détection de discontinuités dans des données vectorielles. Les performances d'estimation de la solution proposée sont comparées à une solution naïve où la solution du Problème 3.1 est calculée dans une fenêtre glissante, et montrent un gain significatif en temps de calcul ainsi qu'une meilleure localisation des discontinuités.

Hormis une examination minutieuse des conditions de KKT du Problème 3.1, l'étape clef de l'algorithme réside dans la mise à jour et le contrôle des bornes inférieures et supérieures de la variable duale à l'intérieur d'un tube de rayon $\lambda$. Un algorithme à la volée est dérivé en découplant les conditions KKT grâce à l'introduction d'un vecteur auxiliaire $\widehat{\mathbf{z}}$, nécessitant d'être estimé, qui fournit l'information sur l'angle de contact entre la « taut string » et le tube. Son estimation affecte fortement la qualité de la solution. Actuellement, son estimation à la volée est accomplie en sélectionnant (selon un critère) une valeur parmi un ensemble prédéfini $\mathcal{Q}$. Nous avons montré que la taille de $\mathcal{Q}$ permet de régler le compromis désiré entre la qualité de la solution et le temps de calcul inhérent à l'application d'intérêt.

De plus, la méthode proposée pourrait également s'étendre à d'autre choix de norme de pénalisation $\ell_{1,p}$ ($p > 1$) dans le membre de droite de (3.12), mais nécessiterait toujours l'estimation de $\widehat{\mathbf{z}}$ à l'intérieur d'une boule $\ell_p$ de rayon $\lambda$. Il serait également intéressant de développer une méthode d'estimation à la volée lorsque l'hypothèse selon laquelle $\widehat{\boldsymbol{z}}$ est constant par morceaux est relâchée.

# Troisième partie

# Autosimilarité multivariée

# Chapitre 5

# Régression non linéaire pour l'analyse d'OfBm

**Sommaire**


**5.1 Autosimilarité multivariée et OfBm . . . . . 97**
5.1.1 Autosimilarité multivariée . . . . . 97
5.1.2 Mouvement Brownien opérateur fractionnaire . . . . . 97
5.1.3 Modèle paramétrique et indéterminations . . . . . 99
**5.2 Analyse en ondelettes des Biv-OfBm . . . . . 99**
5.2.1 Modèle paramétrique des Biv-OfBm . . . . . 99
5.2.2 Spectre en ondelettes . . . . . 101
5.2.2.1 Mélange de lois de puissance . . . . . 101
5.2.2.2 Indétermination supplémentaire . . . . . 102
5.2.2.3 Spectre en ondelettes empirique . . . . . 102
**5.3 Régression non linéaire pour l'identification d'OfBm . . . . . 102**
5.3.1 Formulation variationnelle du problème d'identification d'OfBm . . . 102
5.3.2 Algorithme séparation-évaluation pour les problèmes de minimisation globale . . . . . 104
5.3.3 Algorithme séparation-évaluation pour l'identification de Biv-OfBm 105
5.3.3.1 Relaxation convexe . . . . . 105
5.3.3.2 Encadrement via l'arithmétique par intervalles . . . . . 106
5.3.3.3 Algorithme séparation-évaluation . . . . . 107
**5.4 Performances d'estimation : étude théorique asymptotique . . . 107**
5.4.1 Normalité asymptotique du spectre d'ondelettes . . . . . 108
5.4.2 Consistance de l'estimateur $\widehat{\Theta}_N^M$ . . . . . 109
5.4.3 Normalité asymptotique de l'estimateur $\widehat{\Theta}_N^M$ . . . . . 110
**5.5 Performances d'estimation : étude empirique . . . . . 111**
5.5.1 Configurations expérimentales . . . . . 111
5.5.2 Performances d'estimation . . . . . 111
5.5.3 Coût de calcul . . . . . 113
5.5.4 Nombre d'échantillons vs. précision. . . . . 115
5.5.5 Normalité asymptotique de $\widehat{\Theta}$ . . . . . 115
**5.6 Conclusion et perspectives . . . . . 116**


**Autosimilarité multivariée et OfBm.** En dépit de son importance théorique et pratique, le modèle fBm (voir Définition 1.2) n'arrive pas à fournir un cadre théorique universel pour

l'invariance d'échelle puisque la plupart des contextes d'application modernes impliquent l'enregistrement de séries temporelles multivariées qui nécessitent alors d'être analysées conjointement. Le mouvement Brownien opérateur fractionnaire (OfBm) est une extension multivariée naturelle des fBms. Il a été récemment défini et étudié dans [9, 65, 64, 45] comme le seul processus multivarié gaussien autosimilaire à accroissements stationnaires. Pour un processus $\{X(t)\}_{t\in\mathbb{R}}$ où $X(t) = (X_1(t), \ldots, X_M(t))^\top$, l'autosimililarité multivariée implique que l'exposant de Hurst $H$ fasse maintenant place à une matrice de Hurst

$$\underline{\underline{H}} = W \mathrm{diag}\underline{H} W^{-1} \tag{5.1}$$

où $W$ représente une matrice de mélange $M \times M$ inversible, et $\underline{H}$ est un vecteur de dimension $M$ composé des exposants de Hurst. Les OfBm sont alors caractérisés entièrement par la connaissance de $\underline{H}$, $W$ et la matrice de covariance $\Sigma_X = \mathbb{E}X(1)X(1)^*$.

**Travaux préliminaires.** Les OfBm ont pour l'instant été très rarement utilisés dans les applications, surtout parce que leur utilisation dans le cadre général requiert l'estimation de $M + M^2 + M(M-1)/2$ paramètres de natures très différentes. Toutefois, dans le cas où l'invariance d'échelle se produit composante par composante (cas correspondant à $W$ diagonale), l'identification des OfBm a été minutieusement étudiée [9] et souvent utilisée en applications (cf. e.g., [6, 43]). L'identification a également été récemment achevée pour $W$ non diagonale, mais avec des hypothèses plus restrictives sur $\Sigma_X$ [63]. Encore plus récemment, un estimateur général du vecteur des exposants de Hurst $\underline{H}$ dans le cadre bivarié, i.e., $M = 2$, a été proposé dans [4], mais il requiert des hypothèses supplémentaires pour l'estimation des paramètres $W$ et $\Sigma_X$. La construction d'un paradigme d'estimation multivarié complet reste toujours un problème ouvert dans la littérature.

**Contribution.** La contribution de ce chapitre est double. Premièrement, l'identification complète des OfBm bivariés ($M = 2$, Biv-OfBm) est formulée comme un problème de régression non linéaire dans le domaine des ondelettes. Deuxièmement, une solution algorithmique est établie grâce à une procédure de séparation-évaluation, qui est essentielle due à la nature hautement non convexe du problème d'optimisation proposé.

Les définitions et propriétés des OfBm sont rappelées à la Section 5.1. Une paramétrisation parcimonieuse du processus est également proposée et permet de prévenir de potentielles indéterminations paramétriques des OfBm [65]. Dans la Section 5.2, les propriétés du modèle paramétrique et du spectre en ondelettes des Biv-Ofbms sont explicitement établies et calculées. Cela constitue un cadre mathématique pour la méthode d'estimation proposée. L'identification complète des Biv-OfBm est formulée comme un problème de minimisation dont la solution est développée à partir d'une stratégie de séparation-évaluation (cf. Section 5.3). La consistance et la normalité asymptotique de l'estimateur proposé sont établies théoriquement dans le cas général multivarié (cf. Section 5.4), et évaluées numériquement dans le cadre bivarié à l'aide de simulations de Monte Carlo conduites sur un grand nombre de Biv-OfBm synthétiques (cf. Section 5.5). Les comparaisons avec les estimateurs des exposants de Hurst proposés en [4] sont également rapportées.

**Référence.** Ce chapitre s'appuie sur la contribution [85] et est le résultat d'une collaboration avec G. Didier. Les routines MATLAB pour l'identification et la synthèse des Biv-OfBm sont disponibles à l'adresse suivante `http://perso.ens-lyon.fr/jordan.frecon`.

## 5.1 Autosimilarité multivariée et OfBm

### 5.1.1 Autosimilarité multivariée

Nous avons vu (voir Définition 1.1) que la propriété d'autosimilarité d'un processus univarié $X$ est définie par :

$$\{X(t)\}_{t\in\mathbb{R}} \overset{d.d.f.}{=} \{a^H X(t/a)\}_{t\in\mathbb{R}} \quad (\forall a>0). \tag{5.2}$$

Par analogie, si $X=(X_1,\ldots,X_M)^\top$ désigne à présent un processus multivarié, l'extension vectorielle de (5.2) prend la forme suivante :

**Définition 5.1** (Autosimilarité multivariée)**.** *Soit un $X=(X_1,\ldots,X_M)^\top$ un processus multivarié. $X$ est dit autosimilaire s'il vérifie la relation suivante :*

$$\{X(t)\}_{t\in\mathbb{R}} \overset{d.d.f.}{=} \{a^{\underline{\underline{H}}} X(t/a)\}_{t\in\mathbb{R}} \quad (\forall a>0), \tag{5.3}$$

*où l'exposant d'échelle est la matrice de Hurst*

$$\underline{\underline{H}} = W\, diag\underline{H} W^{-1} \tag{5.4}$$

*et où*

$$a^{\underline{\underline{H}}} := \sum_{k=0}^{+\infty} \log^k(a)\underline{\underline{H}}^k/k!. \tag{5.5}$$

*Dans l'expression (5.4) ci-dessus, $\underline{H}=(H_1,\ldots,H_M)\in]0,1[^M$ désigne le vecteur des exposants de Hurst et $W\in\mathbb{R}^{M\times M}$ est une matrice de mélange.*

### 5.1.2 Mouvement Brownien opérateur fractionnaire

Considérons $M$ mouvements Brownien fractionnaires $X=(X_1,\ldots,X_M)^\top$ d'exposants de Hurst respectifs $\underline{H}=(H_1,\ldots H_M)$ et tels que (voir Définition 1.2)

$$(\forall m\in\{1,\ldots,M\}) \quad \Sigma_{X_m}\equiv \mathbb{E}X_m(1)X_m(1)=\sigma_m^2. \tag{5.6}$$

Une première extension multivariée du modèle fBm consiste à considérer que les $M$ composantes puissent être corrélées. On parle de mouvement Brownien fractionnaire multivarié (mfBm) [10]. Dans ce cas, la matrice de covariance

$$\Sigma_X \equiv \mathbb{E}X(1)X^*(1) = (\sigma_{\mathrm{x}_m}\sigma_{\mathrm{x}_n}\rho_{\mathrm{x}_m,\mathrm{x}_n})_{1\leq m_\leq M, 1\leq n\leq M} \tag{5.7}$$

où $\rho_{\mathrm{x}_m,\mathrm{x}_n}$ désigne la corrélation entre les composantes $X_m$ et $X_n$. Il a été montré dans [10, 64] que le processus $X$ est correctement défini (i.e., la matrice de covariance $\mathbb{E}X(t)X^*(s)$ est toujours définie positive) sous certaines conditions sur $\underline{H}$ et $\Sigma_X$, rappelées dans la proposition suivante :

**Proposition 5.1** (Condition d'existence des mfBms [10, 64].)**.** *Soit un fBm multivarié $X=(X_1,\ldots,X_M)^\top$ caractérisé par les exposants de Hurst $\underline{H}=(H_1,\ldots,H_M)$, les variances $(\sigma^2_{\mathrm{x}_1},\ldots,\sigma^2_{\mathrm{x}_M})$ et les corrélations inter-composantes $(\rho_{\mathrm{x}_n,\mathrm{x}_m})_{1\leq m\leq M,1\leq n\leq M}$. Alors $X$ est correctement défini si et seulement si*

$$\det\Big(\frac{\sigma_{\mathrm{x}_m}\sigma_{\mathrm{x}_n}}{2\pi}\rho_{\mathrm{x}_n,\mathrm{x}_m}\sin\Big(\frac{\pi}{2}(H_m+H_n)\Big)\Gamma(H_m+H_n+1)\Big)_{\substack{1\leq m\leq M\\ 1\leq n\leq M}} > 0. \tag{5.8}$$

Lorsque $X$ est bien défini, la propriété d'autosimilarité (voir Définition 5.1) se simplifie en :

$$\{X_1(at), \ldots, X_M(at)\}_{t\in\mathbb{R}} \overset{d.d.f.}{=} \{a^{H_1}X_1(t), \ldots, a^{H_M}X_M(t)\}_{t\in\mathbb{R}},\ \forall a > 0. \tag{5.9}$$

Dans ce cas, la matrice de mélange $W$ est diagonale et $X$ est invariant d'échelle composante par composante. L'estimation des paramètres peut alors être conduite indépendamment sur chacune des auto-composantes et composantes croisées (cf. [9, 45] pour une étude théorique des performances d'estimation, ou [43] pour méthode d'estimation basée sur les ondelettes et appliqué sur des données réelles).

Cependant, dans le cas général l'autosimilarité multivariée peut être couplée à travers les composantes. Les définitions les plus générales des OfBm ont été formulées en [9, 65, 64, 45] comme étant le seul processus Gaussien multivarié autosimilaire à accroissements stationnaire :

**Définition 5.2** (Mouvement Brownien opérateur fractionnaire (OfBm))**.** *Soit $W \in \mathbb{R}^{M\times M}$ une matrice inversible et $\{X(t)\}_{t\in\mathbb{R}}$ un fBm multivarié avec pour exposants de Hurst $\underline{H} = (H_1, \ldots, H_M) \in ]0,1[^M$. On appelle OfBm, le processus $\{Y(t)\}_{t\in\mathbb{R}}$ défini par*

$$\{Y(t)\}_{t\in\mathbb{R}} = \{WX(t)\}_{t\in\mathbb{R}}. \tag{5.10}$$

*Alors $Y$ est un processus gaussien autosimilaire à accroissements stationnaires avec pour matrice de Hurst $\underline{\underline{H}} = W\,\mathrm{diag}\underline{H}W^{-1}$ entièrement caractérisé par sa fonction de covariance*

$$\mathbb{E}Y(t)Y(s)^* = W\mathbb{E}X(t)X(s)^*W^*, \tag{5.11}$$

*avec* $(\forall m, n \in \{1, \ldots, M\})$

$$\mathbb{E}X_m(t)X_n(s)^* = \frac{\sigma_{\mathrm{x_m}}\sigma_{\mathrm{x_n}}\rho_{\mathrm{x_m},\mathrm{x_n}}}{2}\left(|t|^{H_m+H_n} + |s|^{H_m+H_n} - |t-s|^{H_m+H_n}\right), \tag{5.12}$$

*où* $\sigma_{\mathrm{x_m}}\sigma_{\mathrm{x_n}}\rho_{\mathrm{x_m},\mathrm{x_n}} = \mathbb{E}X_m(1)X_n(1)^*$.

Lorsque $W$ n'est pas diagonale, le comportement en échelle sur chacune des composantes de $Y$ consiste en une somme de lois de puissances. Par conséquent, la conception d'estimateurs dans le cadre multivarié ne peut pas reposer sur une extension directe des procédures univariées usuelles.

Pour finir, notons que les propositions et définitions présentées ici correspondent à une sous-famille de fBm et d'OfBm. Dans la suite de ce manuscrit nous considérons donc que les hypothèses suivantes sont satisfaites.

**Proposition 5.2** (Hypothèses sur l'OfBm $\{Y(t)\}_{t\in\mathbb{R}}$)**.**

*(OFBM1) $\{Y(t)\}_{t\in\mathbb{R}}$ est un OfBm à valeurs dans $\mathbb{R}^M$ ayant pour paramètre de Hurst $\underline{\underline{H}}$, non nécessairement diagonalisable, dont les valeurs propres satisfont $0 < \Re(H_m) < 1$, $(\forall m \in \{1, \ldots, M\})$.*

*(OFBM2) $\mathbb{E}Y(t)Y(t)^*$, $t \neq 0$, est une matrice de plein rang.*

*(OFBM3) $\{Y(t)\}_{t\in\mathbb{R}}$ est un processus stochastique réversible dans le temps.*

### 5.1.3 Modèle paramétrique et indéterminations

Puisque $Y = WX$ est inversible, nous pouvons montrer que

$$\Sigma_{\mathrm{y}}(t) \equiv \mathbb{E}Y(t)Y(t)^* = W\mathbb{E}X(t)X(t)^*W^* \equiv W\Sigma_{\mathrm{x}}(t)W^* \tag{5.13}$$

révèle trois formes d'indétermination dans le modèle paramétrique introduit à la Définition 5.2. Ces dernières sont rassemblées dans la proposition suivante.

**Proposition 5.3** (Indéterminations dans le modèle paramétrique des OfBm)**.** *Soit $Y = WX$ un OfBm défini en Définition 5.2. Alors $Y$ possède les indéterminations suivantes :*

1. *Soient $T_X = diag(\sigma_{\mathrm{x}_1}, \ldots, \sigma_{\mathrm{x}_M})$ et $\Sigma_X = T_X C_X T_X^*$ la matrice de covariance de $X$ où $C_X \equiv (\rho_{\mathrm{x_m,x_n}})_{1\leq m\leq M, 1\leq n\leq M}$. Alors,*

$$Y = WX = W'X', \tag{5.14}$$

   *où $W' = WT_X$ et $X' = T_X^{-1}X$.*

2. *Soit $\Pi \in \{0,1\}^{M\times M}$ une matrice de permutation, i.e., qui ne contient qu'un seul élément non-nul et valant 1 par colonne et par ligne. Alors,*

$$Y = WX = W'X', \tag{5.15}$$

   *où $W' = W\Pi$ et $X' = \Pi^T X$.*

3. *Soient $S$ une matrice diagonale ayant pour entrées $\pm 1$ et $X' = SX$. Alors,*

$$Y = WX = W'X', \tag{5.16}$$

   *où $W' \equiv WS$.*

Afin de lever les trois indéterminations mentionnées précédemment, nous adopterons dans toute la suite les trois conventions suivantes :

**Proposition 5.4** (Conventions sur $W$ et $H$)**.**

1. *Les colonnes de $W$ sont normalisées à 1.*
2. *Les exposants de Hurst vérifient $H_1 \leq ... \leq H_M$.*
3. *Les entrées diagonales de $W$ sont positives.*

## 5.2 Analyse en ondelettes des Biv-OfBm

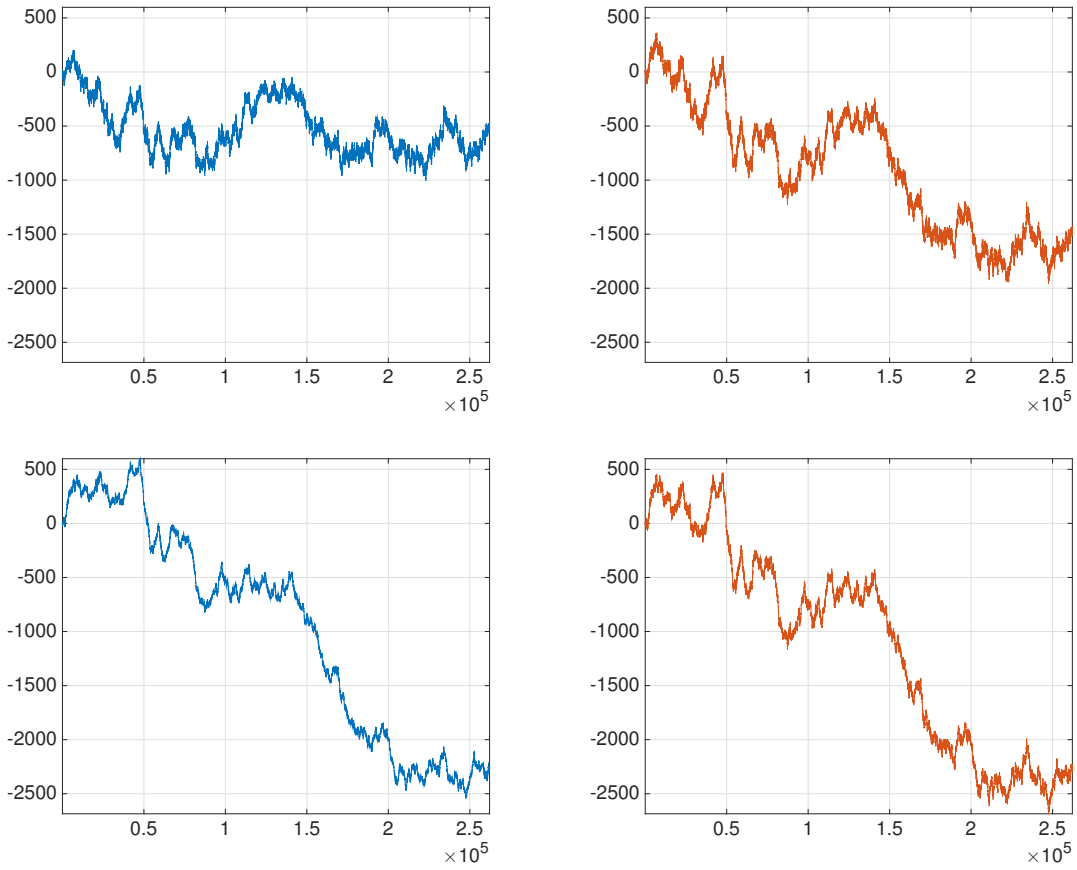
Dans cette section, nous commençons par particulariser les définitions et conventions précédentes au cas bivarié, i.e., $M = 2$. Puis, nous présentons un modèle paramétrique du spectre en ondelettes en guise d'extension bivariée de l'Exemple 1.3.

### 5.2.1 Modèle paramétrique des Biv-OfBm

Lorsque $M = 2$, la condition d'existence des mfBms (voir Proposition 5.1) se réduit à la condition suivante :

**Proposition 5.5** (Condition d'existence des Biv-fBms [10, 64])**.** *Soit un Biv-fBm $X = (X_1, X_2)^\top$ caractérisé par les exposants de Hurst $\underline{H} = (H_1, H_2) \in ]0,1[^2$ et la corrélation inter-composantes $\rho_{\mathrm{x}} \equiv \rho_{\mathrm{x}_1,\mathrm{x}_2}$. Alors $X$ est correctement défini si et seulement si*

$$g(H_1, H_2, \rho_{\mathrm{x}}) \equiv \Gamma(2H_1+1)\Gamma(2H_2+1)\sin(\pi H_1)\sin(\pi H_2) - \rho_{\mathrm{x}}^2\Gamma(H_1+H_2+1)^2\sin^2(\pi(H_1+H_2)/2) > 0. \tag{5.17}$$




(a) Biv-fBm $X$ (b) Biv-OfBm $Y = WX$

FIGURE 5.1 – **Exemple de réalisation d'OfBm.** (a) Biv-fbm $X$ de paramètres ($H_1 = 0.4, H_2 = 0.6, \sigma_{\mathrm{x}_1} = 2, \sigma_{\mathrm{x}_2} = 1, \rho_{\mathrm{x}} = 0.5$). (b) Biv-OfBm $Y = WX$ avec $W$ défini pour $\beta = 0.6$ et $\gamma = -0.2$ (cf. (5.19)).

L'ensemble des régions pour lesquelles $(H_1, H_2, \rho_{\mathrm{x}})$ satisfait (5.17) est illustré par la Figure 5.4 (gauche). Un exemple de Biv-fBm est également représenté par la Figure 5.1 (a) pour $(H_1 = 0.4, H_2 = 0.4, \sigma_{\mathrm{x}_1} = 2, \sigma_{\mathrm{x}_2} = 1, \rho_{\mathrm{x}} = 0.5)$.

A présent, nous arrivons à la définition générique à 7 paramètres des Biv-OfBm que nous utiliserons dans la suite de ce manuscrit :

**Proposition 5.6** (Modèle paramétrique des Biv-OfBm)**.** *Soit $\{X(t)\}_{t\in\mathbb{R}}$ un Biv-fBm avec pour exposant de Hurst $\underline{H} = (H_1, H_2)$, où $0 < H_1 \leq H_2 < 1$, et tel que*

$$\mathbb{E}X(1)X(1)^* = \begin{pmatrix} \sigma^2_{\mathrm{x}_1} & \sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\rho_{\mathrm{x}} \\ \sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\rho_{\mathrm{x}} & \sigma^2_{\mathrm{x}_2} \end{pmatrix}, \tag{5.18}$$

*et soit une matrice de mélange définie par*

$$W = \begin{pmatrix} \frac{1}{\sqrt{1+\gamma^2}} & \frac{\beta}{\sqrt{1+\beta^2}} \\ \frac{-\gamma}{\sqrt{1+\gamma^2}} & \frac{1}{\sqrt{1+\beta^2}} \end{pmatrix} \quad \text{où} \quad \begin{cases} \beta \in [-1, 1], \\ \gamma \in [-1, 1]. \end{cases} \tag{5.19}$$

*Alors le Biv-OfBm $\{Y(t) \equiv WX(t)\}_{t\in\mathbb{R}}$ introduit à la Définition 5.2 est entièrement ca-*

*ractérisé par les 7 paramètres suivants*

$$\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma). \tag{5.20}$$

Un exemple de Biv-OfBm est illustré par la Figure 5.1 (b) pour $\Theta = (H_1 = 0.4, H_2 = 0.4, \sigma_{\mathrm{x}_1} = 2, \sigma_{\mathrm{x}_2} = 1, \rho_{\mathrm{x}} = 0.5, \beta = 0.6, \gamma = -0.2)$. Nous remarquons qu'il est difficile de reconnaitre après le mélange (cf. Figure 5.1 (b)) les composantes du Biv-fBm avant le mélange (cf. Figure 5.1 (a)).

### 5.2.2 Spectre en ondelettes

Nous proposons de nous intéresser ici à la fonction de partition de $Y$ à l'ordre 2 (voir Définition 1.7) définie à partir des coefficients d'ondelettes, également appelée *spectre en ondelette*, que nous noterons

$$E(2^j) \equiv \mathbb{E} D_{\mathrm{y}}(j,\cdot) D_{\mathrm{y}}(j,\cdot)^* \tag{5.21}$$

où $D_{\mathrm{y}}$ désigne les coefficients d'ondelettes bivariés de $Y$ :

**Définition 5.3** (Coefficients d'ondelettes bivariés)**.** *Les coefficients d'ondelettes discrets bivariés de $\{Y(t) = (Y_1(t), Y_2(t))\}_{t\in\mathbb{R}}$ à l'échelle $2^j$ et à la position $2^j k$ sont définis par*

$$(D_{\mathrm{y}}(j,k)) \equiv (D_{\mathrm{y}_1}(j,k), D_{\mathrm{y}_2}(j,k)), \tag{5.22}$$

*où $D_{\mathrm{y}_1}(j,k)$ et $D_{\mathrm{y}_2}(j,k)$ sont respectivements les coefficients d'ondelettes de $Y_1$ et $Y_2$ (voir Définition 1.9). Pour une introduction détaillée aux transformées en ondelettes, nous invitons le lecteur à se référer à, e.g., [139].*

Dans la suite, nous supposerons que l'ondelette mère $\psi_0$ choisie pour le calcul des coefficients d'ondelettes vérifie les conditions suivantes :

**Proposition 5.7** (Hypothèses sur l'ondelette mère $\psi_0 \in L^1(\mathbb{R})$)**.**

*(W1)* $N_\psi \geq 2$.

*(W2)* $\mathrm{supp}(\psi_0)$ *est un intervalle compact.*

*(W3)* $\sup_{t\in\mathbb{R}} |\psi_0(t)|(1+|t|)^\alpha < \infty$ *pour* $\alpha > 1$.

Les propriétés des coefficients d'ondelettes des OfBm ont été étudiées en détail dans le cadre multivarié dans [4]. Dans la suite cette section, nous rappellerons seulement quelques propriétés élémentaires et développerons le nécessaire pour l'identification complète des Biv-OfBm.

#### 5.2.2.1 Mélange de lois de puissance

A partir de (5.9) et $Y = WX$, nous pouvons montrer que le spectre en ondelette s'écrit [4, 85] :

$$\mathbb{E} D_{\mathrm{y}}(j,\cdot) D_{\mathrm{y}}(j,\cdot)^* = W 2^{j(\underline{H} + \mathrm{Id}/2)} E_0 2^{j(\underline{H}^* + \mathrm{Id}/2)} W^* \tag{5.23}$$

où

$$E_0 \equiv \mathbb{E} D_{\mathrm{x}}(0,\cdot) D_{\mathrm{x}}(0,\cdot)^* = \begin{pmatrix} \sigma_{\mathrm{x}_1}^2 \eta_{H_1} & \rho_{\mathrm{x}} \sigma_{\mathrm{x}_1} \sigma_{\mathrm{x}_2} \eta_{\frac{H_1+H_2}{2}} \\ \rho_{\mathrm{x}} \sigma_{\mathrm{x}_1} \sigma_{\mathrm{x}_2} \eta_{\frac{H_1+H_2}{2}} & \sigma_{\mathrm{x}_2}^2 \eta_{H_2} \end{pmatrix} \tag{5.24}$$

et

$$\eta_h = -\frac{1}{2} \int_{\mathbb{R}} |u|^{2h} \mathrm{d}u \int_{\mathbb{R}} \psi_0(v) \psi_0(v-u)^* \mathrm{d}v > 0. \tag{5.25}$$

A partir de la paramétrisation des Biv-OfBm introduite à la Proposition 5.6, le spectre en ondelettes possède la forme explicite suivante :

**Proposition 5.8** (Modèle paramétrique du spectre en ondelettes des Biv-OfBm)**.** *Soit un Biv-OfBm* $\{Y(t)\}_{t\in\mathbb{R}}$ *entièrement paramétré par* $\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma)$ *(c.f. Proposition 5.6), alors son spectre en ondelettes à l'échelle* $2^j$ *possède la forme suivante*

$$\mathbb{E}D_{\mathrm{y}}(j,\cdot)D_{\mathrm{y}}(j,\cdot)^* \equiv E(2^j,\Theta) = \begin{pmatrix} E_{11}(2^j,\Theta) & E_{12}(2^j,\Theta) \\ E_{12}(2^j,\Theta) & E_{22}(2^j,\Theta) \end{pmatrix}, \tag{5.26}$$

*avec*

$$\begin{aligned} E_{11}(2^j,\Theta) = \; & (1+\gamma^2)^{-1} && \sigma^2_{\mathrm{x}_1}\eta_{H_1}2^{j(2H_1+1)} \\ & + 2\beta(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2} && \rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1)} \\ & + \beta^2(1+\beta^2)^{-1} && \sigma^2_{\mathrm{x}_2}\eta_{H_2}2^{j(2H_2+1)}, \end{aligned} \tag{5.27}$$

$$\begin{aligned} E_{12}(2^j,\Theta) = \; & -\gamma(1+\gamma^2)^{-1} && \sigma^2_{\mathrm{x}_1}\eta_{H_1}2^{j(2H_1+1)} \\ & + (1-\beta\gamma)(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2} && \rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1)} \\ & + \beta(1+\beta^2)^{-1} && \sigma^2_{\mathrm{x}_2}\eta_{H_2}2^{j(2H_2+1)}, \end{aligned} \tag{5.28}$$

$$\begin{aligned} E_{22}(2^j,\Theta) = \; & \gamma^2(1+\gamma^2)^{-1} && \sigma^2_{\mathrm{x}_1}\eta_{H_1}2^{j(2H_1+1)} \\ & - 2\gamma(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2} && \rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\eta_{\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1)} \\ & + (1+\beta^2)^{-1} && \sigma^2_{\mathrm{x}_2}\eta_{H_2}2^{j(2H_2+1)}. \end{aligned} \tag{5.29}$$

#### 5.2.2.2 Indétermination supplémentaire

Les équations (5.27), (5.28) et (5.29) révèlent que $|E_{11}(2^j,\Theta)|$, $|E_{12}(2^j,\Theta)|$ et $|E_{22}(2^j,\Theta)|$ sont invariants par la transformation $(\beta,\gamma,\rho_{\mathrm{x}}) \to -(\beta,\gamma,\rho_{\mathrm{x}})$. Par conséquent, dans la suite nous restreindrons la définition de $\rho_{\mathrm{x}}$ à $\rho_{\mathrm{x}} \geq 0$.

#### 5.2.2.3 Spectre en ondelettes empirique

Pour une réalisation $Y \in \mathbb{R}^N$ donnée, nous ferons usage de l'estimateur de la variance d'échantillon de $\mathbb{E}D_{\mathrm{y}}(j,\cdot)D_{\mathrm{y}}(j,\cdot)^*$ :

$$S(2^j) = \frac{1}{N_j}\sum_{k=1}^{N_j} D_{\mathrm{y}}(j,k)D_{\mathrm{y}}(j,k)^*, \quad N_j = \frac{N}{2^j}, \tag{5.30}$$

où $N_j$ désigne le nombre de coefficients d'ondelettes à l'échelle $2^j$. La Figure 5.2 illustre que $S(2^j)$ est un estimateur satisfaisant de $\mathbb{E}D_{\mathrm{y}}(j,\cdot)D_{\mathrm{y}}(j,\cdot)^*$.

## 5.3 Régression non linéaire pour l'identification d'OfBm

### 5.3.1 Formulation variationnelle du problème d'identification d'OfBm

Estimer le vecteur $\Theta$ des paramètres de Biv-OfBm est un problème difficile puisque les entrées du spectre en ondelettes (ou de Fourier) sont un mélange de lois de puissance (cf. Eqs. (5.27)-(5.29)). Par conséquent, nous pouvons écarter l'extension directe des techniques univariées usuelles car elles ne permettraient d'estimer seulement l'exposant de Hurst dominant [205], i.e., $H_2$. Pour cette raison, nous proposons de formuler l'identification complète de Biv-OfBm (i.e., l'estimation de $\Theta$) comme le problème de minimisation suivant :

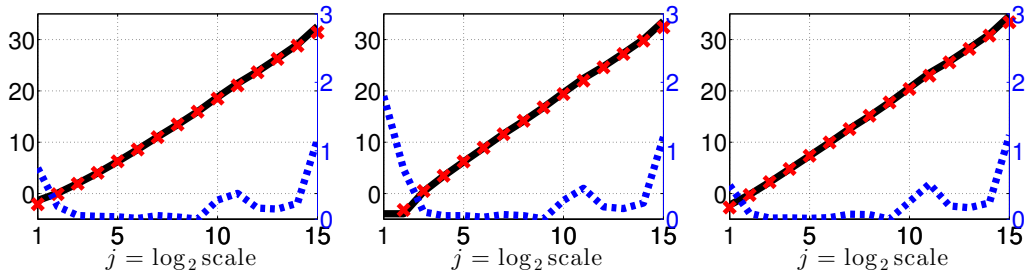



FIGURE 5.2 – **Spectre en ondelettes.** Superposition de $\log_2 |E_{p,p'}(2^j, \Theta)|$ (rouge, '+') et $\log_2 |S_{p,p'}(2^j)|$ (noir) (avec $(p, p') = (1,1), (1,2), (2,2)$ de gauche à droite) pour une réalisation de Biv-OfBm avec $\Theta = (H_1 = 0.4, H_2 = 0.8, \rho_{\mathrm{x}} = 0.1, \sigma_{\mathrm{x}_1} = 1, \sigma_{\mathrm{x}_2} = 1, \beta = 0.5, \gamma = 0.5)$. La différence absolue entre les données et le modèle est représentée en bleu pointillé.

**Problème 5.1** (Régression non linéaire pour l'identification de Biv-OfBm)**.** *Soient une réalisation d'un Biv-OfBm* $Y = (Y_1, \ldots, Y_M)^\top \in \mathbb{R}^{M \times N}$ *et* $(j_1, j_2)$ *tels que* $1 \leq j_1 < j_2 \leq \log_2 N$*. Le problème d'identification de Biv-OfBm est formulé comme le problème de régression non linéaire des coefficients d'ondelettes suivant*[1]

$$\widehat{\Theta}_N^M = \underset{\Theta \in \mathcal{Q}_0}{\operatorname{argmin}}\, C_N(\Theta), \tag{5.31}$$

*avec*

$$C_N(\Theta) \equiv \sum_{\substack{i_1, i_2 = 1 \\ i_1 \leq i_2}}^{2} \sum_{j=j_1}^{j_2} \big(\log_2 |S_{i_1,i_2}(2^j)| - \log_2 |E_{i_1,i_2}(2^j, \Theta)|\big)^2, \tag{5.32}$$

*où* $S(2^j)$ *et* $E(2^j, \Theta)$ *désignent respectivement l'estimateur empirique* (5.30) *et le modèle paramétrique* (5.27)-(5.29) *du spectre en ondelettes de* $Y$ *à l'échelle* $2^j$*.*

**Remarque 5.1.** *L'utilisation de* $\log_2$ *permet d'assurer que toutes les échelles* $2^j$ $(\forall j \in \{j_1, \ldots, j_2\})$ *contribuent également au critère* $C_N$*.*

L'espace de recherche $\mathcal{Q}_0$ permet d'incorporer les informations *a priori* sur le vecteur paramétrique $\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma)$ sous formes de contraintes. Nous résumons ci-dessous l'ensemble des contraintes et conventions :

— Condition $g(H_1, H_2, \rho_{\mathrm{x}}) > 0$ (cf. Proposition 5.5)

— Condition $\rho_{\mathrm{x}} \in [0, 1]$ (cf. Section 5.2.2.2)

— Conditions $0 < H_1 \leq H_2 < 1$ et $(\beta, \gamma) \in [-1, 1]^2$ (cf. Proposition 5.6)

**Remarque 5.2.** *La contrainte* $g(H_1, H_2, \rho_{\mathrm{x}}) > 0$ *implique que* $0 < H_1 < 1$ *et* $0 < H_2 < 1$*.*

Dans un souci de faisabilité, nous restreignons davantage $(\sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}) \in [0, \sigma_{\max}]^2$, avec $\sigma_{\max} = \sqrt{\widehat{\sigma}^2_{\mathrm{y}_1} + \widehat{\sigma}^2_{\mathrm{y}_2}}$, où $\widehat{\sigma}^2_{\mathrm{y_m}}$ désigne l'estimateur de la variance des accroissements de $Y_m$. Nous aboutissons finalement à l'espace de recherche suivant :

$$\begin{aligned} \mathcal{Q}_0 = \Big\{ \Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma) \in \mathbb{R}^7 \,| \\ \Theta \in [0,1]^3 \times [0, \sigma_{\max}]^2 \times [-1,1]^2, \\ g(H_1, H_2, \rho_{\mathrm{x}}) > 0, H_1 \leq H_2 \Big\}. \end{aligned} \tag{5.33}$$

1. L'exposant $\cdot^M$ a été ajouté afin de faire référence au $M$-estimateur dont nous apporterons des détails théoriques dans la Section 5.4.

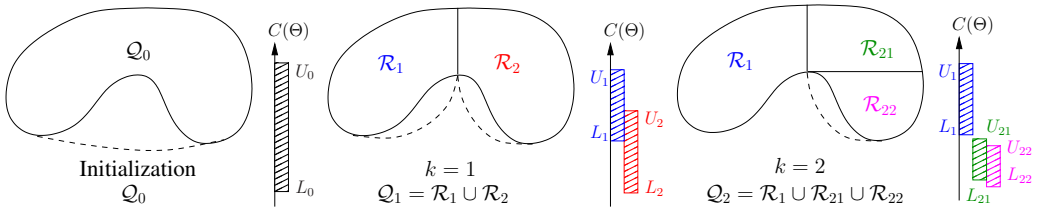



FIGURE 5.3 – **Illustration de l'algorithme séparation-évaluation.**

Le Problème 5.1 est un problème de minimisation compliqué pour deux raisons. La première, car il consiste à démélanger plusieurs lois de puissances, ce qui mène à une fonctionnelle à minimiser $C_N(\Theta)$ hautement non convexe. La deuxième, car les paramètres à estimer dans $\Theta$ (exposants d'échelle, paramètres de mélange, variances et corrélations) sont de natures très différentes. Nous détaillons dans la section suivante une méthode originale pour trouver le minimum global du Problème 5.1 à l'aide d'un algorithme *séparation-évaluation* (« branch-bound »).

### 5.3.2 Algorithme séparation-évaluation pour les problèmes de minimisation globale

Les algorithmes séparation-évaluation sont essentiellement des méthodes d'énumération intelligentes. Ils permettent de résoudre une large variété de problèmes d'optimisation non convexes sous contraintes [109, 120, 152, 178].

Dans le contexte du problème d'identification d'OfBm, nous cherchons le minimiseur global du Problème 5.1. Pour se faire, l'algorithme consiste à partitionner (« branching ») l'espace de recherche $\mathcal{Q}_0$ en des sous-régions de plus en plus petites, à borner le minimum global de $C_N$ sur chacune des ces sous-régions, puis à identifier celles susceptibles de contenir le minimum global. Cette procédure peut être reformulée en 4 étapes répétées jusqu'à ce qu'un critère d'arrêt soit atteint :
- *Sélection.* Choisir une région $\mathcal{R}$ de l'espace de recherche et la relaxer en un ensemble convexe fermé (i.e., un intervalle) tel qu'illustré en pointillés sur la Figure 5.3.
- *Séparation.* Diviser $\mathcal{R}$ en deux plus petites régions $\mathcal{R}_a$ et $\mathcal{R}_b$ telles que $\mathcal{R} = \mathcal{R}_a \cup \mathcal{R}_b$ et $\mathcal{R}_a \cap \mathcal{R}_b = \emptyset$.
- *Encadrement.* Calculer les bornes inférieures et supérieures de $C_N$ sur $\mathcal{R}_a$ et $\mathcal{R}_b$. Les bornes supérieures sont obtenues en évaluant $C_N$ n'importe où dans la région d'intérêt. Les bornes inférieures sont calculées en ayant recours aux techniques d'arithmétique par intervalles [153, 117, 154]), qui combinent des opérations élémentaires afin d'obtenir une borne inférieure grossière d'une fonction, ici $C_N$ sur n'importe quel intervalle.
- *Elaguement.* L'élaguement est guidé par trois mécanismes : éliminer les régions ne satisfaisant pas les contraintes (*infaisabilité*) ; éliminer les régions dont la borne inférieure est plus grande que la plus petite borne supérieure sur toute autre région de $\mathcal{Q}_0$ puisqu'elles ne peuvent pas contenir le minimum global (*borne*) ; éliminer les régions dont la taille, selon tous les paramètres, a atteint la précision fixée par l'utilisateur (*taille*).

Le principe de l'algorithme est illustré Figure 5.3. A l'étape d'initialisation ($k = 0$, gauche), $\mathcal{Q}_0$ est relaxé en un ensemble convexe (pointillés) sur lequel sont calculées les bornes $(L_0, U_0)$ du minimum global de $C_N$, i.e., $L_0 \leq \min_{\Theta \in \mathcal{Q}_0} C_N(\Theta) \leq U_0$. A la première itération ($k = 1$, centre) $\mathcal{Q}_0$, est scindée en $\mathcal{R}_1$ et $\mathcal{R}_2$, et les bornes $(L_1, U_1)$ et $(L_2, U_2)$ sont calculées sur leur relaxation convexe respective. A la deuxième itération ($k = 2$, droite), la région $\mathcal{R}_2$, apparaissant la plus prometteuse car $L_2 < L_1$ (stratégie de recherche « best-first »), est

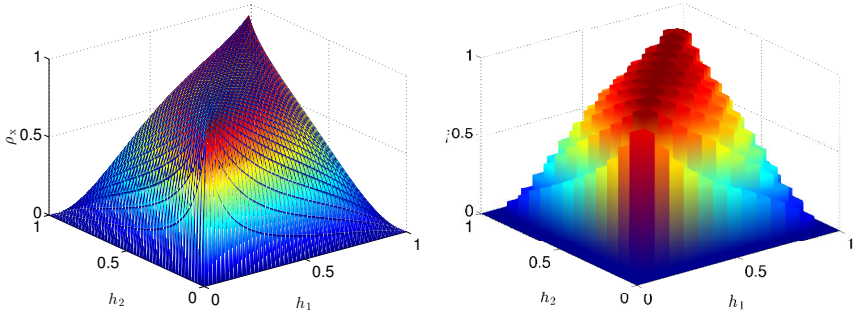



FIGURE 5.4 – **Illustration de la contrainte de faisabilité des OfBm.** Ensemble $\left\{(H_1, H_2, \rho_x) \in [0,1]^3 \mid g(H_1, H_2, \rho_x) > 0\right\}$ (gauche) et son approximation intérieure $\mathcal{P} \equiv \cup_{i=1}^{\Delta^2} \mathcal{P}_i$ (droite) pour $\Delta = 50$.

scindée en $\mathcal{R}_{21}$ et $\mathcal{R}_{22}$. De la même manière, des bornes $(L_{21}, U_{21})$ et $(L_{22}, U_{22})$ sont calculées sur leur relaxation convexe respective. On observe que puisque $L_1 > \min(U_{21}, U_{22})$, alors $\mathcal{R}_1$ ne peut pas contenir le minimum global et peut être éliminée avec certitude de l'espace de recherche. Le raisonnement est identique pour les itérations suivantes.

Les performances de l'algorithme de séparation-évaluation dépendent fortement de deux étapes clefs : la relaxation convexe et l'encadrement du minimum de $C_N$. Dans la suite, nous nous intéressons à ces deux étapes particularisées à la résolution du Problème 5.1.

### 5.3.3 Algorithme séparation-évaluation pour l'identification de Biv-OfBm

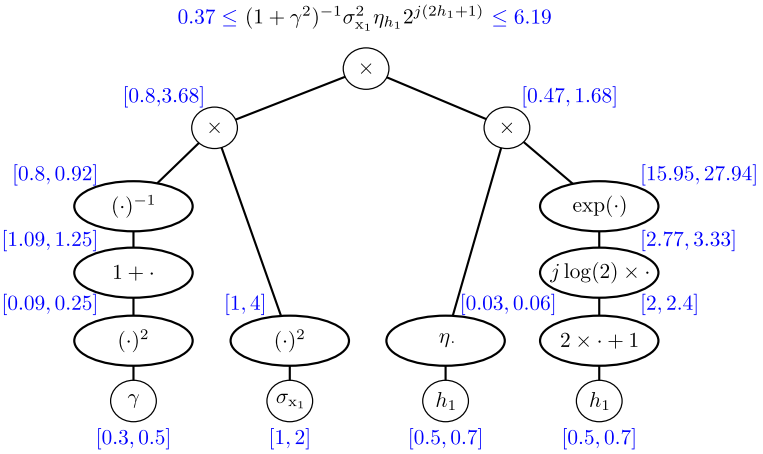
#### 5.3.3.1 Relaxation convexe

Par nature, les techniques d'encadrement basées sur *l'arithmétique par intervalles* s'appliquent seulement à des *intervalles*, c'est-à-dire à des ensembles convexes fermés. Par conséquent, l'utilisation de cette technique dans une procédure de séparation-évaluation, nécessite de relaxer chaque sous-région $\mathcal{R}$ en un ensemble convexe. En particulier, ceci est vrai dès l'étape d'initialisation où nous devons fournir une relaxation convexe de l'espace de recherche $\mathcal{Q}_0$, illustrée Figure 5.3 (gauche) en traits pointillés. Cependant, dans le cas présent, nous ne pouvons pas relaxer la contrainte non convexe $g(H_1, H_2, \rho_x) > 0$ car elle assure la faisabilité de la solution (cf. Proposition 5.5). A la place, nous proposons d'approcher $\mathcal{Q}_0$ par une relaxation convexe intérieure détaillée dans la proposition suivante :

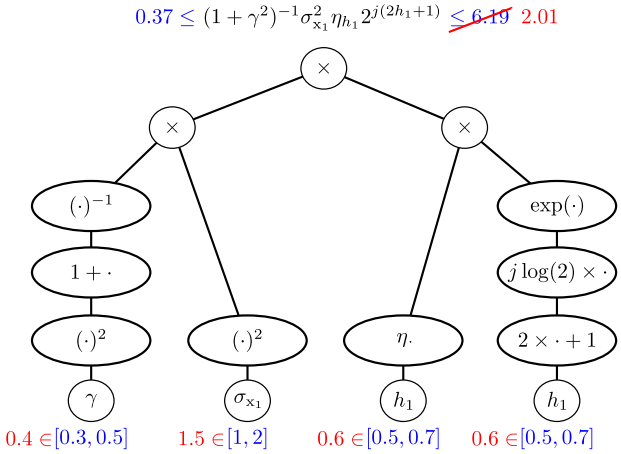
**Proposition 5.9** (Approximation $\mathcal{S}_0$ de l'espace de recherche $\mathcal{Q}_0$)**.** *Soient $\mathcal{Q}_0$ l'ensemble défini en* (5.33) *et $\Delta \in \mathbb{N}^*$. Alors $\mathcal{Q}_0$ peut être approché par une relaxation convexe intérieure $\mathcal{S}_0$ composé de l'union de $\Delta^2$ ensembles convexes $\mathcal{C}_i$, i.e., $\left(\mathcal{S}_0 \equiv \cup_{i=1}^{\Delta^2} \mathcal{C}_i\right) \subset \mathcal{Q}_0$ et $\cap_{i=1}^{\Delta^2} \mathcal{C}_i = \emptyset$, définis par*

$$\mathcal{C}_i = \Big\{\Theta = (H_1, H_2, \rho_x, \sigma_{x_1}, \sigma_{x_2}, \beta, \gamma) \in \mathbb{R}^7 \mid \Theta \in \mathcal{P}_i \times [0, \sigma_{\max}]^2 \times [-1,1]^2, H_1 \leq H_2\Big\}, \tag{5.34}$$

*et où $\left(\mathcal{P} \equiv \cup_{i=1}^{\Delta^2} \mathcal{P}_i\right) \subset \left\{(H_1, H_2, \rho_x) \in [0,1]^3 \mid g(H_1, H_2, \rho_x) > 0\right\}$ est obtenu en divisant $(H_1, H_2) \in [0,1]^2$ en carrés $\mathcal{T}_i$ à partir du pas de discrétisation $\Delta^{-1}$ et en définissant $\mathcal{P}_i = \mathcal{T}_i \times [0, \rho_i]$ où $\rho_i$ est la plus grande valeur telle que $(\forall (H_1, H_2) \in \mathcal{T}_i)$, $g(H_1, H_2, \rho_i) > 0$.*




(a) Décomposition en arbre de calcul (noir) et obtention des bornes par propagation des intervalles (bleu).




(b) Raffinement de la borne supérieure (rouge).

FIGURE 5.5 – **Illustration de la méthode d'encadrement par l'arithmétique par intervalles.**

La relaxation convexe intérieure proposée à l'avantage de nécessairement satisfaire la contrainte $g(H_1, H_2, \rho_x) > 0$, ce qui est un bénéfice majeur en pratique car les régions infaisables ne seront pas explorées par l'algorithme.

La Figure 5.4 illustre la qualité de l'approximation $\mathcal{P}$ pour $\Delta = 50$ que nous utiliserons dans la suite.

#### 5.3.3.2 Encadrement via l'arithmétique par intervalles

Les techniques basées sur l'arithmétique par intervalles sont utilisées en pratique pour calculer la borne inférieure d'une fonctionnelle $C_N$ sur un ensemble convexe $\mathcal{R}$ [153, 117, 154]. Elles reposent sur la décomposition explicite de $C_N$ en plusieurs fonctions élémentaires telles la somme, le produit, l'inverse, le carré, le logarithme, l'exponentielle, ..., appelée *arbre de calcul*.

Par souci de lisibilité, nous ne détaillerons pas l'arbre de calcul associé au critère $C_N$ mais seulement celui associé au premier terme de (5.27), illustré sur la Figure 5.5 (noir). Les feuilles de l'arbre (la ligne du bas en noir) consistent en une ou plusieurs occurences des variables présentes dans $\Theta$. Chaque noeud du graphe correspond à une fonction élémentaire

s'appliquant à son ou ses fils. Par conséquent, $(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\eta_{H_1}2^{j(2H_1+1)}$ peut être formé en parcourant l'arbre depuis les feuilles jusqu'à la racine.

Supposons à présent que nous cherchons à borner $(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\eta_{H_1}2^{j(2H_1+1)}$ sur un intervalle $\mathcal{R}$ où $\gamma \in [0.3, 0.5]$, $\sigma_{\mathrm{x}_1} \in [1,2]$ et $H_1 \in [0.5, 0.7]$. De la même manière, un encadrement peut être obtenu en propageant les intervalles depuis les feuilles jusqu'à la racine (voir Figure 5.5a en bleu). La littérature sur l'arithmétique par intervalles fournit un encadrement pour la plupart des fonctions élémentaires dont certaines sont rappelées ci-dessous :

**Exemple 5.1** (Opérations usuelles en arithmétique par intervalles)**.**

- $[\underline{x},\overline{x}] + [\underline{y},\overline{y}] = [\underline{x}+\underline{y}, \overline{x}+\overline{y}]$
- $[\underline{x},\overline{x}] - [\underline{y},\overline{y}] = [\underline{x}-\overline{y}, \overline{x}-\underline{y}]$.
- $[\underline{x},\overline{x}] \times [\underline{y},\overline{y}] = [\min\{\underline{x}\times\underline{y}, \underline{x}\times\overline{y}, \overline{x}\times\underline{y}, \overline{x}\times\overline{y}\}, \max\{\underline{x}\times\underline{y}, \underline{x}\times\overline{y}, \overline{x}\times\underline{y}, \overline{x}\times\overline{y}\}]$
- $[\underline{x},\overline{x}] \div [\underline{y},\overline{y}] = [\underline{x},\overline{x}] \times [\frac{1}{\overline{y}}, \frac{1}{\underline{y}}]$, *si* $0 \notin [\underline{y},\overline{y}]$
- $\log([\underline{x},\overline{x}]) = [\log(\underline{x}), \log(\overline{x})]$, *si* $0 \notin [\underline{x},\overline{x}]$
- $\exp([\underline{x},\overline{x}]) = [\exp(\underline{x}), \exp(\overline{x})]$

La seule fonction non-élémentaire présente dans la décomposition de $C_N$ est $h \mapsto \eta_h$, définie in (5.25) et illustrée Figure 5.6, pour l'ondelette orthogonales de Daubechies la moins asymétrique $\psi_0$ avec $N_\psi = 2$. A partir de l'étude de sa monotonicité, nous pouvons dériver l'encadrement empirique suivant :

**Proposition 5.10** (Fonction $\eta$ en arithmétique par intervalles)**.** *Soient un ensemble* $[\underline{h},\overline{h}] \subseteq [0,1]$ *et la fonction* $h \mapsto \eta_h$ *défini en* (5.25) *pour l'ondelette orthogonale de Daubechies la moins asymétrique* $\psi_0$ *avec* $N_\psi = 2$ *moments nuls. Alors*

$$\eta_{[\underline{h},\overline{h}]} = \begin{cases} [\eta_{\underline{h}}, \eta_{\overline{h}}], & \text{si } \overline{h} < 0.3, \\ [\eta_{\overline{h}}, \eta_{\underline{h}}], & \text{si } \underline{h} \geq 0.3, \\ [\min(\eta_{\overline{h}}, \eta_{\underline{h}}), 0.071], & \text{autrement.} \end{cases} \tag{5.35}$$

La propagation des intervalles permet d'obtenir un encadrement grossier du minimum global de $C_N$ sur chaque région $\mathcal{R}$. Notons qu'une simple évaluation de $C_N$ pour n'importe quel valeur $\Theta \in \mathcal{R}$, et en particulier au milieu de $\mathcal{R}$, produit également une borne supérieure du minimum global. En pratique, cela permet de raffiner l'estimation obtenue par propagation d'intervalles comme illustré Figure 5.5b (rouge).

#### 5.3.3.3 Algorithme séparation-évaluation

L'identification complète des Biv-Ofbm est obtenue via la séquence des opérations de l'Algorithme 8.

## 5.4 Performances d'estimation : étude théorique asymptotique

Dans cette section, les propriétés asymptotiques de la solution exacte $\widehat{\Theta}_N^M$ (voir (5.31)) sont étudiées théoriquement dans le cadre général des OfBm multivariés [9, 65, 64]. En d'autre mots, les résultats qui seront énoncés ne sont pas restreints au cadre bivarié de la Section 5.1. Nous généralisons donc la somme dans la fonction objective (5.32) aux indices $i_1, i_2 = 1, \ldots, M \geq 2$.

L'ensemble des énoncés présentés dans cette section sont dus à G. Didier. Les démonstrations peuvent être retrouvées dans l'Annexe C.1.

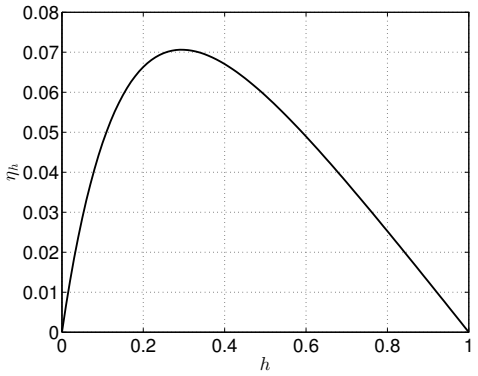



FIGURE 5.6 – **Illustration de la fonction $h \mapsto \eta_h$.** Elle est ici représentée pour l'ondelette orthogonale de Daubechies la moins asymétrique $\psi_0$ avec $N_\psi = 2$ moments nuls.

**Algorithme 8** séparation-évaluation pour l'identification de Biv-OfBm

**Entrée :** Spectre en ondelette empirique $S(2^j)$, $j = j_1, \ldots, j_2$.
Modèle paramétrique du spectre en ondelette $E(2^j, \Theta)$.
Précision maximale $\delta$ sur chacun des paramètres de $\Theta$.
Paramètre de relaxation convexe intérieure $\Delta$.

1: Fixer $\widehat{\mathcal{S}} = \emptyset$, $\mathcal{S}_0 = \cup_{i=1}^{\Delta^2} \mathcal{C}_i$ et $k = 0$.
2: Calculer les bornes inférieures $l_i$ de $C_N$ sur $\mathcal{C}_i$ ($\forall i = 1, \ldots, \Delta^2$).
3: Calculer les bornes supérieures $u_i$ de $C_N$ sur $\mathcal{C}_i$ ($\forall i = 1, \ldots, \Delta^2$).
4: Fixer $U = \min(u_1, \ldots, u_{\Delta^2})$ et $L = \min(l_1, \ldots, l_{\Delta^2})$.
5: **Tant que** $\mathcal{S}_k \neq \emptyset$ **faire**
6: (Sélection) Sélectionner la région $\mathcal{R} \subset \mathcal{S}_k$ avec la borne inférieure la plus faible $L$ (stratégie de recherche « best-first »).
7: (Séparation) Diviser $\mathcal{R}$ en $\mathcal{R}_a$ et $\mathcal{R}_b$ tels que $\mathcal{R} = \mathcal{R}_a \cup \mathcal{R}_b$ et $\mathcal{R}_a \cap \mathcal{R}_b = \emptyset$, selon son plus long côté, où la longueur d'un côté est défini relativement à la précision maximum $\delta$ fixée par l'utilisateur.
8: (Encadrement inférieur) Calculer les bornes inférieures de $C_N$ sur $\mathcal{R}_a$ et $\mathcal{R}_b$, à l'aide de l'arithmétique par intervalles.
9: (Encadrement supérieur) Calculer la borne supérieure de $C_N$ sur $\mathcal{R}_a$ (resp. $\mathcal{R}_b$), en évaluant $C_N(\Theta)$ pour $\Theta$ choisi au centre de $\mathcal{R}_a$ (resp. $\mathcal{R}_b$).
10: (Branchement) Mettre à jour le partitionnement $\mathcal{S}_{k+1} = (\mathcal{S}_k \backslash \mathcal{R}) \cup \mathcal{R}_a \cup \mathcal{R}_b$,
11: (Elaguement) Supprimer d'éventuelles régions $\mathcal{R}^*$ de $\mathcal{S}_{k+1}$ soit par *non optimalité*, *infaisablité* ou par *taille*, i.e. $\mathcal{S}_{k+1} \leftarrow \mathcal{S}_{k+1} \backslash \mathcal{R}^*$. Rajouter dans $\widehat{\mathcal{S}}$ les régions de $\mathcal{S}_{k+1}$ qui ont été supprimées par *taille* ; Supprimer les régions $\widehat{\mathcal{S}}$ par *non-optimalité*.
12: Fixer $k \leftarrow k + 1$

**Sortie :** Liste des régions candidates $\widehat{\mathcal{S}}$ obtenue pour la précision $\delta$.
Région $\widehat{\mathcal{S}}_{\min} \in \widehat{\mathcal{S}}$ avec la borne inférieure la plus faible.
L'estimateur à la précision $\delta$ est le milieu de $\widehat{\mathcal{S}}_{\min}$ et est noté

$$\widehat{\Theta}_N^{M,BB}. \tag{5.36}$$

### 5.4.1 Normalité asymptotique du spectre d'ondelettes

Le comportement asymptotique de l'estimateur $\widehat{\Theta}_N^M$ tire partie de la normalité asymptotique (à échelle fixe) de la variance des ondelettes. Sous les hypothèses (OFBM1-3) et (W1-3), la normalité asymptotique du spectre d'ondelettes peut être établie à partir d'un argument pratiquement identique à celui utilisé en [4, Théorème 3.1]. Pour le confort du lecteur, nous

rappelons l'affirmation ci-dessous :

**Théorème 5.1.** *Supposons que les hypothèses (OFBM1-3) et (W1-3) soient vérifiées. Soit* $F \in \mathcal{S}(\frac{P(P+1)}{2}m, \mathbb{R})$ [2] *la matrice de covariance asymptotique énoncée en [4, Proposition 3.3]. Alors,*

$$\Big(\sqrt{K_j}(\mathrm{vec}_{\mathcal{S}} S(2^j) - \mathrm{vec}_{\mathcal{S}} E(2^j, \Theta))\Big)_{j=j_1,\dots,j_2} \overset{d.d.f}{\to} \mathbf{Z}, \tag{5.37}$$

*lorsque* $N \to \infty$*, où* $j_1 < \ldots < j_2$ *et* $m = j_2 - j_1 + 1$ *et* $\mathbf{Z} \overset{d.d.f}{=} \mathcal{N}_{\frac{P(P+1)}{2} \times m}(0, F)$*, où* $\mathrm{vec}_{\mathcal{S}}$ *est l'opération qui rassemble en un vecteur la partie triangulaire haute d'une matrice symétrique :*

$$\mathrm{vec}_{\mathcal{S}}(S) = (s_{11}, \dots, s_{1P}; \dots; s_{P-1,P-1}, s_{P-1,P}; s_{P,P})^*. \tag{5.38}$$

### 5.4.2 Consistance de l'estimateur $\widehat{\Theta}_N^M$

Soit $\overline{\Theta}$ le vecteur des vrais paramètres. Afin de prouver la consistance de l'estimateur $\widehat{\Theta}_N^M$, nous avons besoin d'introduire les hypothèses suivantes sur la paramétrisation $\Theta$ :

**Proposition 5.11** (Hypothèses sur le vecteur paramétrique $\Theta$)**.**

*(P1) L'espace de recherche* $\Xi \subseteq \mathcal{Q}_0$ *(voir* (5.33)*) est un ensemble compact à dimension finie et*

$$\overline{\Theta} \in \mathrm{int}\,\Xi. \tag{5.39}$$

*(P2) Pour un* $j^* \in \{j_1, \dots, j_2\}$*,*

$$\Theta \neq \Theta' \Rightarrow |E_{i_1^*, i_2^*}(2^{j^*}, \Theta)| \neq |E_{i_1^*, i_2^*}(2^{j^*}, \Theta')| \tag{5.40}$$

*pour un couple* $(i_1^*, i_2^*)$*.*

*(P3)* $(\forall i_1, i_2 \in \{1, \dots, P\}), (\forall j \in \{j_1, \dots, j_2\})$*,*

$$E_{i_1,i_2}(2^j, \overline{\Theta}) \neq 0. \tag{5.41}$$

*(P4) L'application*

$$\Theta \mapsto E(2^j, \Theta) \tag{5.42}$$

*est trois fois continûment différentiable sur* $\mathrm{int}\,\Xi$*.*

Sous l'hypothèse (P3), les fonctions $\log_2 |E_{i_1,i_2}(2^j, \Theta)|$, $(\forall i_1, i_2 \in \{1, \dots, M\})$, sont bien définies. Ce fait et le Théorème 5.1 impliquent alors que les fonctions $\log_2 |S_{i_1,i_2}(2^j)|$, $(\forall i_1, i_2 \in \{1, \dots, M\})$, sont bien définies avec probabilité tendant vers 1.

Pour sa part, l'hypothèse (P2) implique que la valeur absolue de chaque entrée de la matrice cible $E(2^{j^*}, \Theta)$ est (paramétriquement) identifiable, c'est-à-dire, qu'il existe une fonction injective $\Xi \ni \Theta \mapsto |E(2^{j^*}, \Theta)|$.

La fonction objectif $\Theta \mapsto C_N(\Theta)$ est une fonction de $N$, tout comme $S(2^j)$, $(\forall j \in \{j_1, \dots, j_2\})$. Puisque $C_N$ est continue et que $\Xi$ est compact alors, pour tout $N$, un minimum $\widehat{\Theta}_N$ est atteint asymptotiquement, d'où la formation de la séquence suivante

$$\{\widehat{\Theta}_N^M\}_{N \in \mathbb{N}}. \tag{5.43}$$

Toute séquence (5.43) définie un $M$-estimateur de $\overline{\Theta}$, e.g., [203, Chapitre 5]. Le théorème suivant montre que la séquence d'estimateur (5.43) est consistant.

2. $\mathcal{S}(n, \mathbb{R})$ est l'espace des matrices symétriques réelles de taille $n \times n$.

**Théorème 5.2** (Consistance de l'estimateur $\widehat{\Theta}_N^M$)**.** *Supposons que les hypothèses (OFBM1-3), (W1-3) et (P1-4) soient vérifiées. Alors, la séquence des minima* (5.43) *est un estimateur consistant, i.e.,*

$$\widehat{\Theta}_N^M \to \overline{\Theta} \quad \text{en probabilité.} \tag{5.44}$$

**Remarque 5.3.** *L'unicité de $\widehat{\Theta}_N^M$ pour un $N$ donné n'est pas assurée par les conditions (P1-4). Cependant, elle n'est pas requise parmi les hypothèses du Théoreme 5.2.*

### 5.4.3 Normalité asymptotique de l'estimateur $\widehat{\Theta}_N^M$

Contrairement à la consistance de l'estimateur, la démonstration de sa normalité asymptotique requiert une hypothèse supplémentaire énoncé dans la suite

**Proposition 5.12** (Hypothèse additionnelle sur le vecteur paramétrique $\Theta$)**.**

$$\textit{(P5)} \qquad \det\Big(\sum_{j=j_1}^{j_2}\sum_{1\le i_1\le i_2\le M}\Lambda_{i_1,i_2}(2^j,\overline{\Theta})\Lambda_{i_1,i_2}(2^j,\overline{\Theta})^*\Big)>0, \tag{5.45}$$

*où l'on défini le vecteur $\Lambda_{i_1,i_2}(2^j,\Theta)^* = \nabla_\Theta \log_2 |E_{i_1,i_2}(2^j,\Theta)|$.*

**Théorème 5.3.** *Supposons que les hypothèses (OFBM1-3), (W1-3) et (P5) soient vérifiées. Soit $\{\widehat{\Theta}_N^M\}_{N\in\mathbb{N}}$ une séquence consistante de minima respectivement associé à $\{C_N\}_{N\in\mathbb{N}}$. Alors*

$$\sqrt{N}(\widehat{\Theta}_N^M - \overline{\Theta}) \overset{d.d.f}{\to} \mathbf{W}, \quad N\to\infty, \ \textit{avec} \tag{5.46}$$

$$\mathbf{W} \overset{d.d.f}{=} \Big(\sum_{j=j_1}^{j_2}\sum_{1\le i_1\le i_2\le M}\Lambda_{i_1,i_2}(2^j,\Theta)\Lambda_{i_1,i_2}(2^j,\Theta)^*\Big)^{-1} \cdot\Big(\sum_{j_1=1}^{j_2}\frac{2^{j/2}}{\log 2}\sum_{1\le i_1\le i_2\le M} Z_{i_1,i_2}(2^j)\frac{\Lambda_{i_1,i_2}(2^j,\Theta)}{E_{i_1,i_2}(2^j,\overline{\Theta})}\Big), \tag{5.47}$$

*où $\mathbf{Z} = (Z_{i_1,i_2}(2^j))_j$ est un vecteur aléatoire dont la distribution est obtenue dans la limite faible* (5.37)*.*

Le résultat suivant est un corollaire du Théorème 5.2 et du Théorème 5.3.

**Corollaire 5.1.** *Supposons que les hypothèses (OFBM1-3), (W1-3) et (P1-5) soient vérifiées. Soit $\{\widehat{\Theta}_N^M\}_{N\in\mathbb{N}}$ une séquence consistante de minima respectivement associé à $\{C_N\}_{N\in\mathbb{N}}$. Soit également $\{\widehat{\Theta}_N^{M,BB}\}_{N\in\mathbb{N}}$ un estimateur de la forme* (5.36) *satisfaisant*

$$\|\widehat{\Theta}_N^M - \widehat{\Theta}_N^{M,BB}\| \le \frac{C}{N^{1/2+\varepsilon}} \quad \text{a.s.} \tag{5.48}$$

*pour certaines constantes $C>0$ et $\varepsilon>0$. Alors,*

$$\sqrt{N}(\widehat{\Theta}_N^{M,BB} - \overline{\Theta}) \overset{d.d.f}{\to} \mathbf{W}, \quad N\to\infty,$$

*où le vecteur aléatoire $\mathbf{W}$ est donné en* (5.47)*.*

**Remarque 5.4.** *La Condition* (5.48) *est facilement satisfaite en pratique, puisque sur un ensemble compact et pour un faible coût de calcul, l'algorithme séparation-évaluation garantit de fournir une solution à une distance contrôlée du vrai minimum.*

**Remarque 5.5.** *La condition technique* (5.45) *devrait être satisfaite dans de nombreux cas d'intérêts discutés dans l'Annexe C.3.*

## 5.5 Performances d'estimation : étude empirique

### 5.5.1 Configurations expérimentales

Des simulations de Monte Carlo ont été réalisées pour quantifier les performances à taille finie de l'estimateur $\widehat{\Theta}_N^{M,BB}$.

**Configurations expérimentales.** Afin d'examiner l'influence de $\overline{\Theta}$ sur les performances d'estimation, 9 différentes valeurs de $\overline{\Theta}$ ont été utilisées. Elles sont obtenues en variant la corrélation entre les composantes ($\rho_{\mathrm{x}} = 0.1$, 0.45 et 0.8) et les paramètres de mélange (aucun mélange, $\beta = \gamma = 0$ ; mélange orthogonal, $\beta = \gamma = 0.5$, et $\beta = -\gamma = 0.5$, appelé mélange anti-orthogonal). Trois tailles d'observations (court, $N = 2^{10}$, moyen, $N = 2^{14}$ et long, $N = 2^{18}$) sont examinées. Les résultats sont rapportés ici pour $(H_1, H_2) = (0.4, 0.8)$. Des conclusions similaires sont obtenues pour différents choix de $(H_1, H_2)$.

Pour chaque configuration $\overline{\Theta}$, les performances d'estimation sont évaluées à partir de diagrammes en boîte calculés sur 100 copies indépendantes d'OfBm. La synthèse des OfBm est réalisée à l'aide de la toolbox multivarié mise au point dans [104, 105], cf. www.hermir.org. Le coût de calcul est quant à lui quantifié en terme de pourcentage du nombre d'itérations que nécessiterait un algorithme de recherche exhaustif sur grille. L'analyse en ondelettes est menée à partir des ondelettes orthogonales de Daubechies les moins asymétriques [139]. Toutes les échelles sont utilisées pour calculer $C_N : j_1 = 1 \leq j \leq j_2 = \log_2 N - N_\psi - 1$. Les résultats sont rapportés pour $N_\psi = 2$. Il a été observé qu'augmenter $N_\psi$ n'améliorait pas les performances d'estimation.

L'algorithme de séparation-évaluation proposé est mis en oeuvre pour $\mathcal{S}_0 = \cup_{i=1}^{\Delta^2} \mathcal{C}_i$ avec $\Delta = 50$. L'impact de la précision $\delta$ a également été étudié et est discuté dans la Section 5.5.4.

**Etat de l'art.** Les performances de $\widehat{\Theta}_N^{M,BB}$, re-labellisé $\widehat{\Theta}^M$, sont comparées contre deux autres procédures d'estimation.

Dans la première, les exposants d'échelle $(H_1, H_2)$ sont estimés à l'aide de l'estimateur univarié du paramètre d'Hurst appliqué indépendamment sur $\log_2 S_{11}(2^j)$ et $\log_2 S_{22}(2^j)$ [205]. L'estimateur univarié de $H_1$ (resp. $H_2$) est obtenu en prenant le minimum (resp. maximum) des coefficients de régression linéaire de $\log_2 S_{11}(2^j)$ et $\log_2 S_{22}(2^j)$ contre $j \in J^{\mathrm{fs}}$ (resp. $j \in J^{\mathrm{cs}}$), avec $J^{\mathrm{fs}} = \{j_1, \ldots, \lfloor \frac{j_1+j_2}{2} \rfloor\}$ (petites échelles) et $J^{\mathrm{cs}} = \{\lfloor \frac{j_1+j_2}{2} \rfloor + 1, \ldots, j_2\}$ (grandes échelles). On note $(\widehat{H}_1^{\mathrm{U}}, \widehat{H}_2^{\mathrm{U}})$ les estimées correspondantes.

Dans la deuxième, les paramètres $(H_1, H_2, \beta)$ sont estimés à l'aide d'un estimateur multivarié semi-paramétrique $(\widehat{H}_1^{\mathrm{W}}, \widehat{H}_2^{\mathrm{W}}, \widehat{\beta}^{\mathrm{W}})$ proposé en [4], qui repose sur la structure propre multi-échelle de $S(2^j)$.

Les statistiques de $\widehat{\Theta}^{\mathrm{M}}$, $(\widehat{H}_1^{\mathrm{W}}, \widehat{H}_2^{\mathrm{W}}, \widehat{\beta}^{\mathrm{W}})$ et $(\widehat{H}_1^{\mathrm{U}}, \widehat{H}_2^{\mathrm{U}})$ sont comparées Figure 5.8 à Figure 5.11 en jaune, bleu et magenta respectivement.

Bien que la paramétrisation complète des Biv-OfBm requiert un vecteur paramètre à 7 dimensions $\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x}_1}, \sigma_{\mathrm{x}_2}, \beta, \gamma)$, pour simplifier l'exposé nous nous intéresserons aux 5 paramètres les plus intéressants $(H_1, H_2, \rho_{\mathrm{x}}, \beta, \gamma)$. Cela fait suite à la littérature univariée qui se concentre sur l'estimation de $H$ pour les fBms, tout en négligeant le paramètre le moins intéressant $\sigma^2$.

### 5.5.2 Performances d'estimation

**Estimation de l'exposant d'échelle dominant $H_2$ (Figure 5.7).** Comme attendu, toutes les méthodes produisent une estimation précise du plus grand exposant $H_2$. Alors que toutes les méthodes montrent des performances comparables pour un grand nombre d'échantillons,

il est intéressant de remarquer que $\widehat{H}_2^{\mathrm{M}}$ affiche de meilleures performances (biais et dispersion plus faibles) pour un faible nombre d'échantillons que $\widehat{H}_2^{\mathrm{W}}$ et $\widehat{H}_2^{\mathrm{U}}$. L'impact de la corrélation $\rho_{\mathrm{x}}$ et des paramètres du mélange $(\beta, \gamma)$ sur les performances est faible.

**Estimation de l'exposant d'échelle non dominant $H_1$ (Fig. 5.8).** Estimer le plus faible exposant d'échelle $H_1$ est intrinsèquement plus difficile à cause du mélange de lois de puissance qui masque la valeur de l'exposant de Hurst non dominant. Comme attendu, l'analyse univariée n'arrive pas à estimer correctement $H_1$ (excepté quand il n'y a pas de mélange $\beta = \gamma = 0$). Alors que $\widehat{H}_1^{\mathrm{M}}$ et $\widehat{H}_2^{\mathrm{W}}$ montrent essentiellement les mêmes performances pour un grand nombre d'échantillon, il est intéressant de remarquer que $\widehat{H}_1^{\mathrm{M}}$ affiche des performances largement supérieures (biais et dispersion plus faible) pour un faible nombre d'échantillons. Cependant, on observe que $\widehat{H}_1^{\mathrm{M}}$ est biaisé pour $\beta = -\gamma = 0.5$ et $\rho_{\mathrm{x}} = 0.1$, ce qui montre que $\widehat{H}_1^{\mathrm{M}}$ est davantage affecté par la conjonction de faible corrélation et de mélange anti-orthogonal que $\widehat{H}_2^{\mathrm{M}}$.

**Estimation du paramètre de mélange $\beta$ (Fig. 5.9).** Un avantage significatif de $\widehat{\beta}^{\mathrm{M}}$ est qu'il est robuste à un faible nombre d'échantillons, alors que $\widehat{\beta}^{\mathrm{W}}$ ne l'est pas. Alors que les paramètres du mélange ne semblent pas avoir un impact sur les performances de $\widehat{\beta}^{\mathrm{M}}$, une faible corrélation $\rho_{\mathrm{x}}$ détériore ses performances. Des expériences additionnelles montrent que les performances d'estimation de $\widehat{\beta}^{\mathrm{M}}$ sont robustes à une diminution de $H_2 - H_1$, alors que les performances de $\widehat{\beta}^{\mathrm{W}}$ sont drastiquement détériorées pour $H_2 - H_1 \to 0$.

**Estimation du paramètre de mélange $\gamma$ (Fig. 5.10).** Les performances de $\widehat{\gamma}^{\mathrm{M}}$ sont très satisfaisantes, bien qu'on observe qu'elles sont affectées par une faible corrélation.

**Estimation de la corrélation $\rho_{\mathrm{x}}$ (Fig. 5.11).** Le paramètre $\rho_{\mathrm{x}}$ semble être le plus difficile à estimer, avec un biais significatif pour une faible corrélation et un mélange anti-orthogonal. Ce résultat est cohérent avec [9].

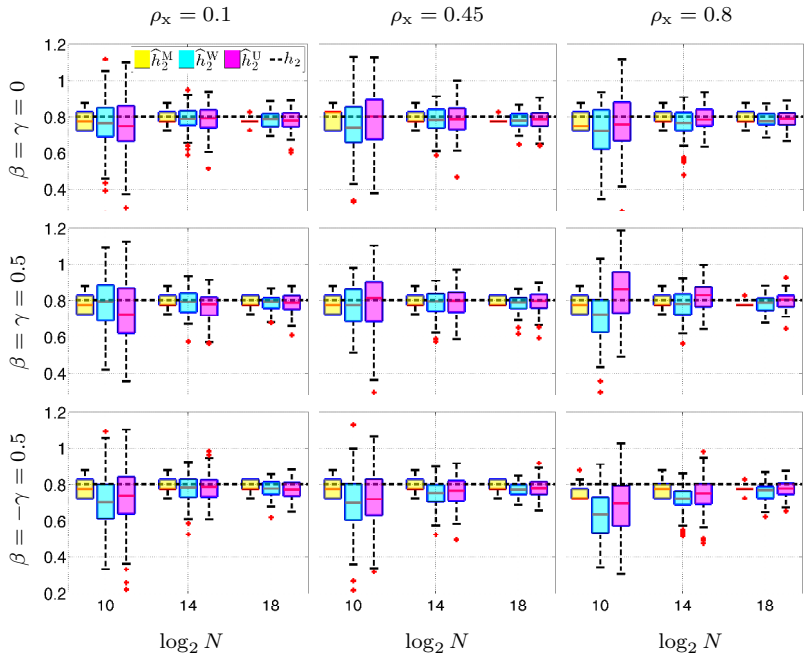



FIGURE 5.7 – **Performances d'estimation de $H_2$** en fonction de $\log_2 N$.

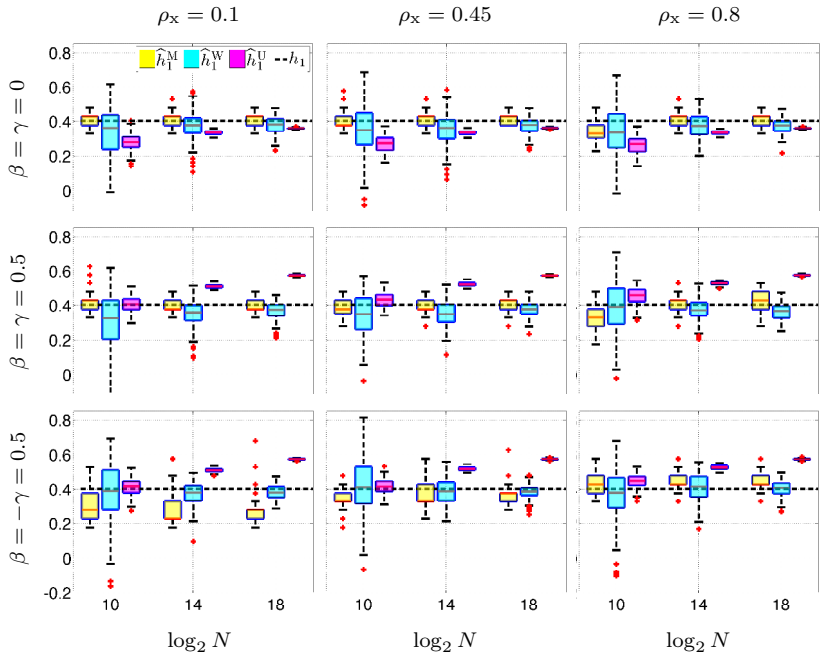



FIGURE 5.8 – **Performances d'estimation de** $H_1$ en fonction de $\log_2 N$.

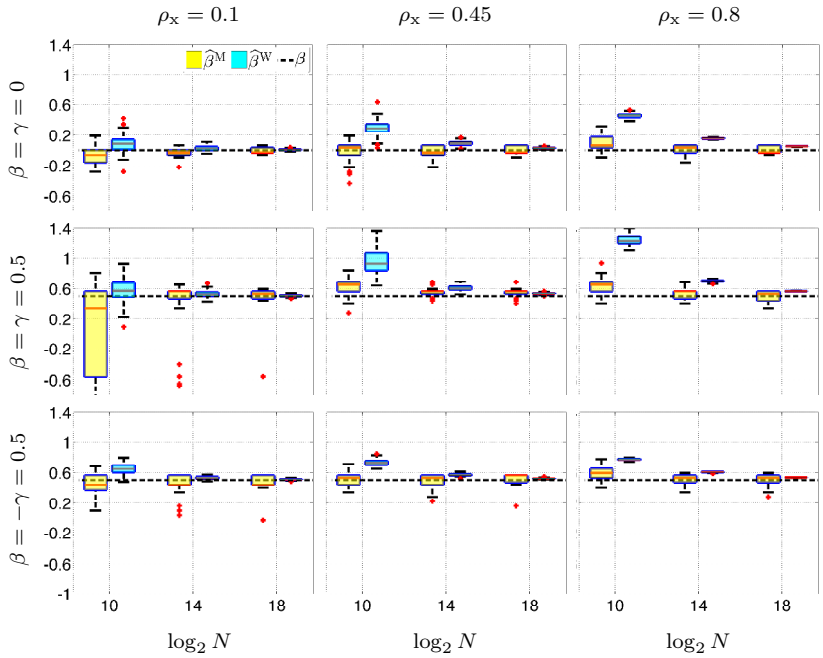



FIGURE 5.9 – **Performances d'estimation de** $\beta$ en fonction de $\log_2 N$.

### 5.5.3 Coût de calcul

Le coût de calcul de l'algorithme d'identification proposé et décrit à la Section 5.3.3.3 est reporté Figure 5.12 (figures du haut) en fonction de $\log_2 N$ pour chaque configuration de paramètres. Le coût de calcul est significativement plus faible que celui d'un algorithme de recherche exhaustif sur grille. La Figure 5.12 montre également clairement que plus la corrélation entre les composantes est faible, plus la minimisation de la fonctionnelle $C_N$ est facile, ce qui indique que le terme croisé joue un rôle significatif dans l'identification des

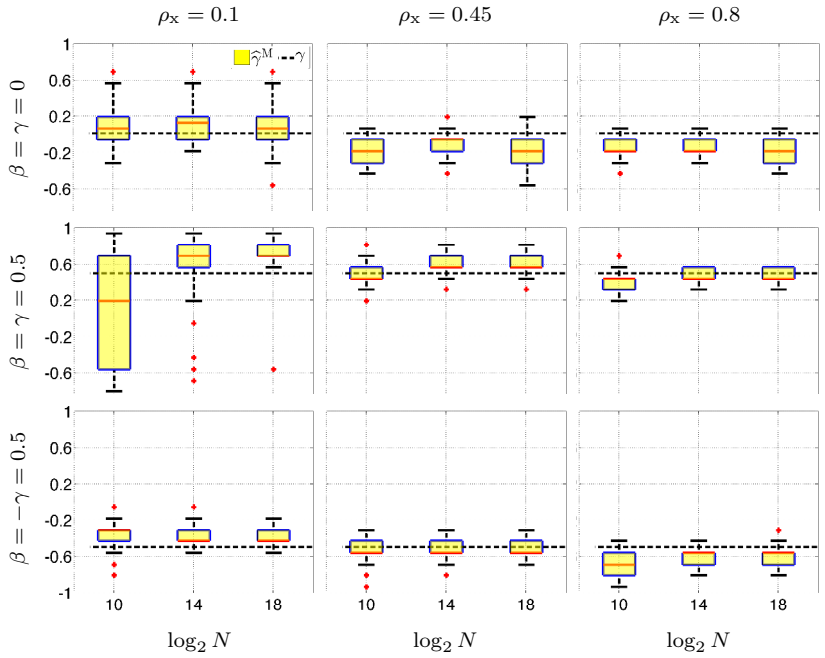



FIGURE 5.10 – **Performances d'estimation de** $\gamma$ en fonction de $\log_2 N$.

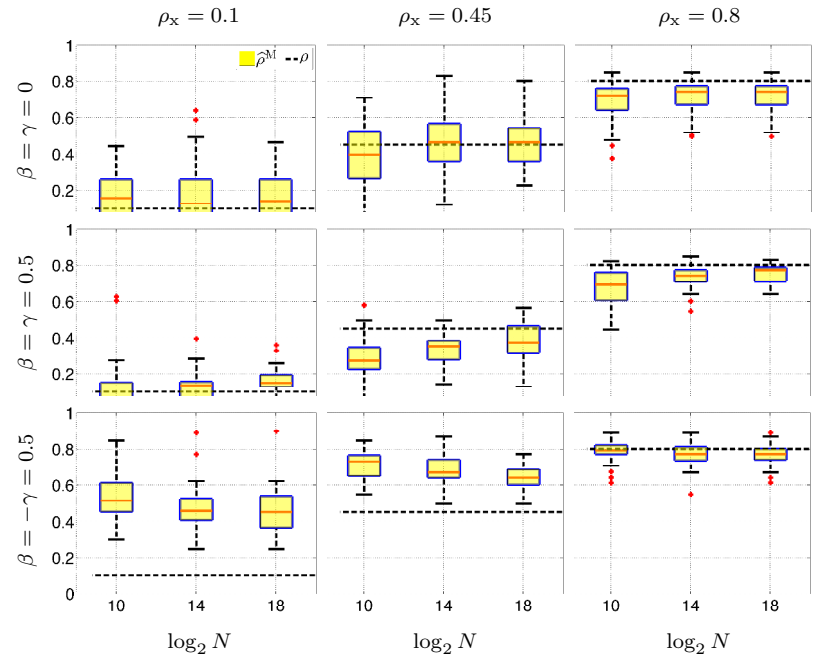



FIGURE 5.11 – **Performances d'estimation de** $\rho_x$ en fonction de $\log_2 N$.

Biv-OfBm. Bien que surprenant à première vue, la nette décroissance du coût de calcul avec l'augmentation du nombre d'échantillons peut être interprétée comme le fait qu'il est évidement plus facile de démélanger trois lois de puissance lorsqu'un grand nombre d'échelles $2^j$ est disponible, c'est-à-dire pour un grand nombre d'échantillons. Il convient de noter que le mélange orthogonal, qui semblait intuitivement être le plus facile, apparaît être celui le plus coûteux en terme de nombre d'itérations requis pour minimiser $C_N$. Sans surprise, le coût de calcul augmente quand la précision $\delta$ requise sur les estimées augmente ($\delta \to 0$) Figure 5.13 (gauche).

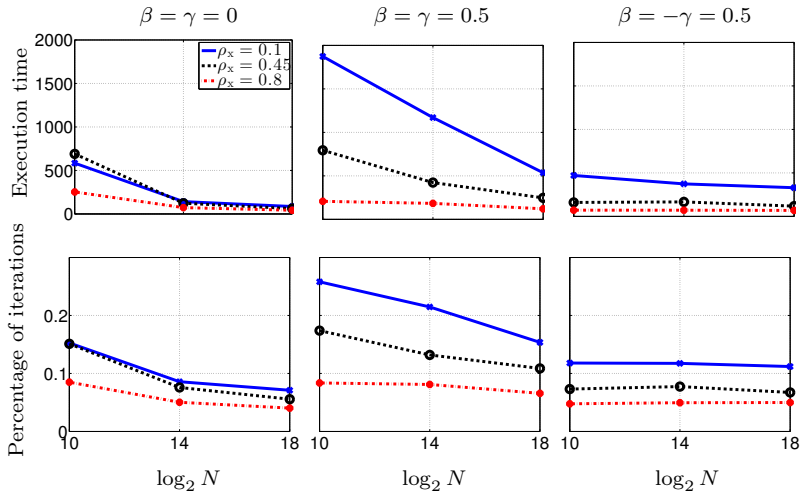



FIGURE 5.12 – **Coût de calcul en fonction de** $\log_2 N$**.** Haut : temps d'exécution moyen (en secondes). Bas : pourcentage d'itérations effectuées (comparé à un algorithme exhaustif sur une grille de même précision $\delta$).

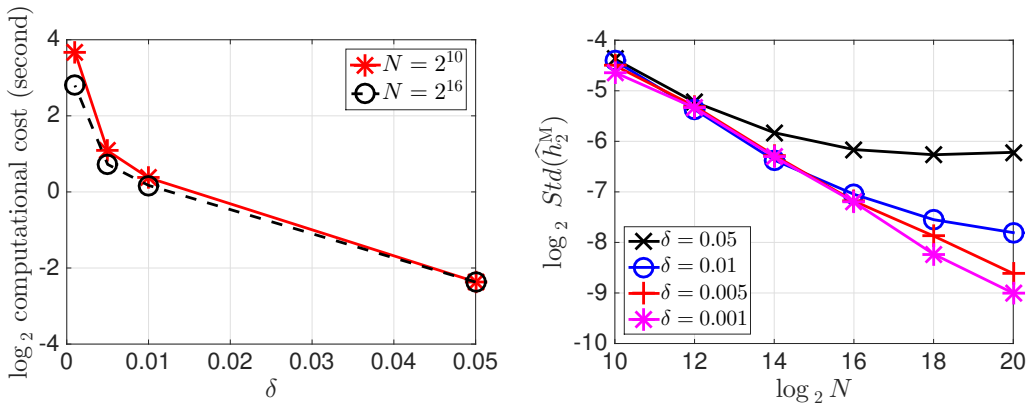



FIGURE 5.13 – **Précision.** Coût de calcul en secondes (gauche) et dispersion des estimées de $H_2$ (droite), en fonction de la précision $\delta$.

### 5.5.4 Nombre d'échantillons vs. précision.

Les Figures 5.7 à 5.11 suggèrent qu'augmenter la taille des échantillons $N$ améliore les performances de $(\widehat{H}_1^{\mathrm{U}}, \widehat{H}_2^{\mathrm{U}})$ et $(\widehat{H}_1^{\mathrm{W}}, \widehat{H}_2^{\mathrm{W}}, \widehat{\beta}^{\mathrm{W}})$, comme le montre la décroissance du biais et de la variance médiane avec $N$. La Figure 5.13 (right) indique que $N$ a davantage d'impact sur $\widehat{\Theta}^{\mathrm{M}}$. Tant que la dispersion des estimées reste plus grande que la précision choisie $\delta$ (sélectionnée indépendamment de $N$), on observe la décroissance attendue du biais et de la dispersion lorsque $N$ augmente. Puisque la minimisation est arrêtée lorsque la précision choisie $\delta$ est atteinte, augmenter $N$ sans décroitre $\delta$ n'apporte pas de gain de performance, ce qui montre que la précision doit décroitre lorsque la taille de l'échantillon augmente (empiriquement comme $N^{-1/4}$) pour améliorer les performances.

### 5.5.5 Normalité asymptotique de $\widehat{\Theta}$

Pour une étude rigoureuse de la normalité, $\Theta$ est restreint à $(H_1,\ H_2)$, alors que les autres paramètres sont fixés *a priori* et supposé connus. Les moyennes sont obtenues sur $10^3$ réalisations indépendantes de Biv-OfBm pour chaque nombre d'échantillon. La Figure 5.14

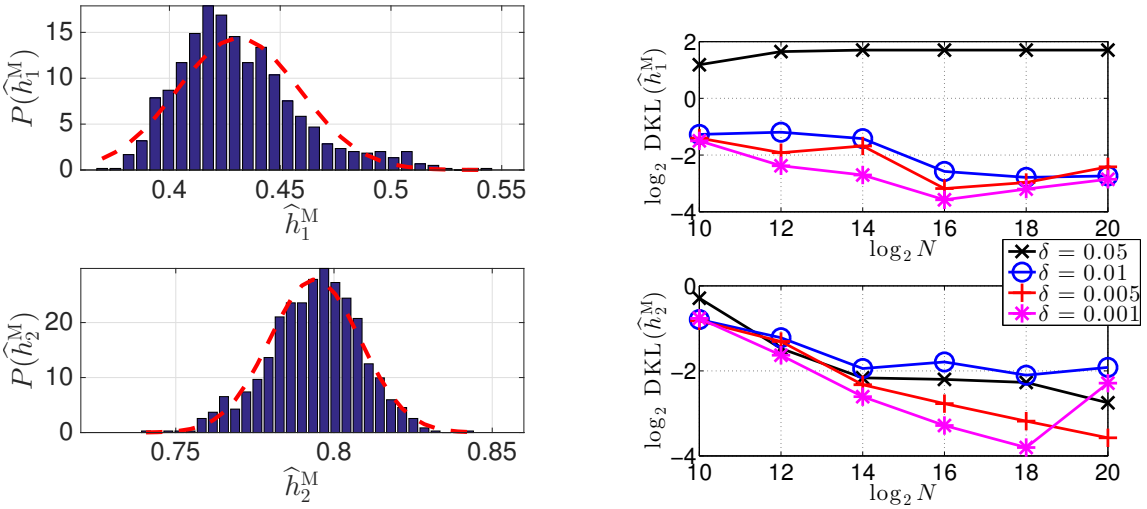



FIGURE 5.14 – **Normalité asymptotique.** Paramètres $N = 2^{20}$ et $\delta = 10^{-3}$. Gauche : distributions empiriques de $\widehat{H}_1^{\mathrm{M}}$, $\widehat{H}_2^{\mathrm{M}}$ et leur meilleur ajustement gaussien. Droite : divergence de Kullback-Leibler entre les distributions empiriques de $\widehat{H}_1^{\mathrm{M}}$, $\widehat{H}_2^{\mathrm{M}}$ et leur meilleur ajustement gaussien, en fonctions de la taille des observations $N$.

(gauche) compare visuellement les distributions empiriques de $\widehat{H}_1^{\mathrm{M}}$, $\widehat{H}_2^{\mathrm{M}}$ à leur meilleur ajustement gaussien. La Figure 5.14 (droite) montre la divergence de Kullback-Leibler entre les distributions empiriques de $\widehat{H}_1^{\mathrm{M}}$, $\widehat{H}_2^{\mathrm{M}}$ et leur meilleur ajustement gaussien, en fonction du nombre d'échantillons $N$. La Figure 5.14 confirme la normalité asymptotique des estimées $\widehat{H}_1^{\mathrm{M}}$, $\widehat{H}_2^{\mathrm{M}}$ comme il a été prédit théoriquement à la Section 5.4. Il montre d'autre part que plus la précision requise $\delta$ est grand, plus la convergence vers la normalité est rapide. De plus, la normalité asymptotique est atteinte pour un nombre bien plus faible d'échantillons pour le plus grand exposant d'échelle $\widehat{H}_2^{\mathrm{M}}$ que pour le plus faible $\widehat{H}_1^{\mathrm{M}}$.

## 5.6 Conclusion et perspectives

A notre connaissance, ce chapitre propose le premier procédé d'identification complète des Biv-OfBm. Son originalité réside dans la formulation du problème d'identification comme un problème de régression non linéaire ainsi que dans la proposition d'une procédure séparation-évaluation pour fournir des solutions efficaces et élégantes au problème d'optimisation non convexe correspondant.

La consistance et la normalité de l'estimateur ont été démontrées théoriquement dans un cadre multivarié général par G. Didier. Les performances d'estimation sont évaluées en dimension finie par des simulations de Monte Carlo et s'avèrent satisfaisantes pour tous les paramètres. Les paramètres $\gamma$ (mélange) et $\rho_{\mathrm{x}}$ (corrélation entre les composantes) restent les plus difficiles à estimer, bien qu'aucune autre procédure d'estimation n'ait été proposée dans la littérature. Cependant, inclure l'estimation de $\gamma$ et $\rho_{\mathrm{x}}$ dans la formulation du problème de régression non linéaire permet de surpasser la méthode de l'état de l'art, $(\widehat{H}_1^{\mathrm{W}}, \widehat{H}_2^{\mathrm{W}}, \widehat{\beta}^{\mathrm{W}})$, pour estimer $H_1$, $H_2$ et $\beta$, au prix d'un coût de calcul lourdement augmenté. La procédure de séparation-évaluation proposée montre toutefois un gain significatif en coût de calcul comparé à une recherche exhaustive sur grille. Les performances d'estimation sont satisfaisantes et assez contrôlées pour être mises en application sur des données réelles.

La procédure d'identification proposée ainsi que ses bonnes performances d'estimation ouvrent la voie vers des tests d'hypothèses permettant de tester l'absence de mélange (i.e., $W$ diagonal) ou de corrélation entre les composantes (i.e., $\rho_{\mathrm{x}} \equiv 0$).

# Chapitre 6

# Application : détection d'anomalies dans le trafic internet

**Sommaire**


**6.1 Identifiation paramétrique de Biv-OfGns . . . . . 121**
6.1.1 Bruit Gaussien opérateur fractionnaire . . . . . 121
6.1.2 Intégration fractionnaire et analyse en ondelettes . . . . . 121
6.1.3 Spectre en ondelettes paramétrique de Biv-OfGn . . . . . 121
6.1.4 Procédure d'identification . . . . . 122
6.1.5 Performances . . . . . 123
**6.2 Application aux données internet . . . . . 123**
6.2.1 Données MAWI et projections aléatoires . . . . . 123
6.2.2 Rôle du paramètre d'intégration fractionnaire . . . . . 124
6.2.3 Résultats . . . . . 125
**6.3 Conclusion et perspectives . . . . . 128**


**Surveillance du trafic internet.** Le contrôle et la modélisation du trafic internet constituent une tâche cruciale pour l'ingénierie et la conception des réseaux, l'allocation de ressources, l'optimisation des performances et des services, et pour l'évaluation de la sécurité. En particulier, la détection de trafic anormal a suscité une attention et un effort de recherche considérable puisqu'un trafic malicieux peut avoir des conséquences dramatiques à la fois pour l'utilisateur et pour l'opérateur. Cependant, dans le contexte du trafic internet, la détection d'anomalies est un défi de taille dû à la nature très hétérogène des réseaux, applications, protocoles et comportements des utilisateurs. De sucroît, les trafics normaux montrent une large diversité, et leur définition reste une question difficile. Les comportements anormaux peuvent correspondre à un niveau encore plus élevé d'hétérogénéité, allant des dénis de services distribués (≪ Distributed Denial of Service ≫, DDoS) et des balayages de ports jusqu'à de potentielles occurrences de comportements encore jamais rencontrés (cf. e.g., [20, 127, 148] pour une revue détaillée). De plus, la nature des données disponibles peut également dépendre du réseau ou de l'opérateur. Les données peuvent être analysées soit en paquets IP (≪ Internet Protocol ≫) sous forme de quintuplets [1] systématiquement horodatés et dont le volume est exprimé en bytes, soit en enregistrements de granularité plus élevée (applications, protocoles, ...). Le présent chapitre analyse les séries temporelles composées du nombre de paquets IP

1. Le quintuplet standard d'un paquet contient 5 champs d'entête : l'adresse IP, le numéro de port de la source et de la destination, et le protocole IP (TCP, UDP or ICMP).

$Y_{Pkt,\Delta}(t)$ (ou de bytes $Y_{Byt,\Delta}(t)$) dans des intervalles de temps consécutifs, de taille $\Delta$.

**Etat de l'art : détection d'anomalies.** La littérature dédiée à la détection d'anomalies dans le trafic internet est considérable (cf. e.g., [20, 127, 148, 62, 162] pour des revues) et une description exhaustive n'est pas le centre d'intérêt de ce chapitre. L'attention est portée ici sur la détection d'anomalies à partir de séries temporelles agrégées, ce qui a l'avantage de respecter la vie privée des utilisateurs, contrairement aux techniques reposant sur l'examination de la charge utile des paquets IP. A cause de la nature constamment changeante du trafic internet, extraire une référence pour le trafic normal contre laquelle les trafics anormaux peuvent être comparés est une tâche pratiquement impossible. A l'instar, les outils de projection aléatoire, également appelés *procédures de hachage* ou « sketching », sont utilisés pour la construction automatique de trafic auto-référencé [123, 62, 27].

La détection d'anomalies repose souvent sur une modélisation statistique du trafic normal. Parmi les diverses tentatives, l'autosimilarité et le mouvement Brownien fractionnaire (voir Définition 1.2) ont montré qu'il était possible de modéliser pertinemment et de façon robuste la statistique du trafic internet à travers une large variété de réseaux, trafics et depuis les premiers âges d'internet jusqu'aux collections de données les plus récentes [132, 164, 214, 2, 83]. Le paramètre de Hurst $H$ quantifiant la relation entre les différentes échelles est souvent trouvé empiriquement $H \in [0.8, 1]$.

Cependant, jusqu'à présent, dans la plupart des procédures de détection d'anomalies reposant sur l'autosimilarité, l'analyse et la modélisation sont restées univariées : les séries temporelles aggrégées du nombre de paquets et de bytes ont été analysées indépendamment [132, 2, 62, 27, 99]. Il est également souvent très débattu et controversé de décider si l'analyse d'autosimilarité doit être conduite sur les séries temporelles du nombre de paquets ou de bytes. Souvent, les paramètres $H$ mesurés sur chaque type de données diffèrent, ce qui remet en question soit la validité du célèbre mécanisme de file d'attente qui relie l'autosimilarité à la nature à queue lourde de la distribution des objets distribués sur internet (cf. e.g., [57, 197, 28]), soit qui soulève des questions telles que : est-ce que les mécanismes s'appliquent aux paquets, aux bytes, ou aux deux ? Si oui, pourquoi $H$ devrait-il être différent entre les paquets et les bytes ?

**Contribution.** Afin d'adresser ces questions, la disponibilité jointe des séries temporelles du nombre de paquets et de bytes a rarement été exploitée au sein d'une analyse bivariée de l'autosimilarité, voire *a contrario* [189] et [27] pour des comparaisons préliminaires entre les exposants d'autosimilarité mesurés indépendemment sur $Y_{Pkt,\Delta}(t)$ et $Y_{Byt,\Delta}(t)$. Nous proposons de modéliser conjointement l'autosimilarité des paquets et des bytes par un Biv-OfGn (cf. [64, 9]) dont l'étude des paramètres permet de concevoir une procédure de détection d'anomalies. Afin d'adapter le modèle aux données internet, nous proposons d'incorporer un paramètre d'intégration fractionnaire dans l'analyse du spectre en ondelettes. Les performances d'estimations sont évaluées à partir de simulations de Monte Carlo conduites sur des OfGns synthétiques mimant les propriétés du trafic internet. Une procédure de détection d'anomalies, construite à partir des paramètres Biv-OfGn des séries temporelles des paquets et des bytes, est ensuite appliquée à un large ensemble de collections récentes et de qualité du trafic internet, le repertoire MAWI [1], décrit à la Section 6.2.1. Les résultats sont discutés en Section 6.2.3.

**Références.** Ce chapitre s'appuie sur la contribution [86].

## 6.1 Identifiation paramétrique de Biv-OfGns

### 6.1.1 Bruit Gaussien opérateur fractionnaire

Soit $\{X(t) = (X_1(t), X_2(t))^\top\}_{t\in\mathbb{R}}$ un Biv-fBm . Par analogie avec les fBms, on définit le processus de ses accroissements $X^\delta(t) = X(t+1) - X(t)$ qui peuvent être assimilés à une dérivée du processus à la résolution 1. Ils sont appelés bruits gaussiens fractionnaires multivariés (≪ multivariate fractional Gaussian noise ≫, mfGn). Ce sont des processus gaussiens, centrés et stationnaires entièrement décrits par leur matrice de covariance : $(\forall m, n \in \{1, 2\})$ et $(\forall t, s \in \mathbb{R})$

$$\mathbb{E}X_m^\delta(t)X_n^\delta(s)^* = \frac{1}{2}\sigma_{\mathrm{x_m}}\sigma_{\mathrm{x_n}}\rho_{\mathrm{x_m,x_n}}\big(|t-s+1|^{H_m+H_n} + |t-s-1|^{H_m+H_n} - 2|t-s|^{H_m+H_n}\big). \quad (6.1)$$

Si l'on considère à présent une matrice de mélange $W$, les accroissements du Biv-OfBm $Y = WX$ (voir Définition 5.2), i.e., $Y^\delta = WX^\delta$, définissent un Biv-OfGn entièrement caractérisé par la connaissance de la matrice de covariance (6.1) et de $W$.

### 6.1.2 Intégration fractionnaire et analyse en ondelettes

Contrairement à la définition classique des coefficients d'ondelettes bivariés (voir Définition 5.3), nous considérons cette fois ci la définition :

**Définition 6.1** (Coefficients d'ondelettes bivariés intégrés fractionnairement)**.** *Les coefficients d'ondelettes discrets bivariés de $\{Y(t) = (Y_1(t), Y_2(t))\}_{t\in\mathbb{R}}$ à l'échelle $j$ et à la position $k$ sont définis par*

$$(D_{\mathrm{y}}(j,k)) \equiv (D_{\mathrm{y_1}}(j,k), D_{\mathrm{y_2}}(j,k)) \quad (6.2)$$

*où, pour tout $m \in \{1, 2\}$, $D_{\mathrm{y_m}}(j,k)$ est le coefficient d'ondelette de $Y_m$ à l'échelle $j$ et à la position $k$, défini par*

$$D_{\mathrm{y_p}}(j,k) = \int_{\mathbb{R}} \psi_{j,k}(t)Y_p(t)\mathrm{d}t, \quad (6.3)$$

$$\text{où } \{\psi_{j,k}(t) = 2^{-j(1/2-\mu)}\psi_0(2^{-j}t - k)\}_{(j,k)\in\mathbb{Z}^2} \quad (6.4)$$

La seule différence réside dans l'introduction d'un paramètre additionnel $\mu$ dans (6.4) qui joue le rôle de paramètre d'intégration fractionnaire [213] et dont le rôle crucial en pratique est détaillé à la Section 6.2.2.

### 6.1.3 Spectre en ondelettes paramétrique de Biv-OfGn

De manière similaire à la Section 5.2.2.1 pour les Biv-OfBms, nous pouvons montrer que le spectre en ondelettes des Biv-OfGns s'écrit [4, 85] :

**Proposition 6.1** (Modèle paramétrique du spectre en ondelettes de Biv-OfGn)**.** *Soit un Biv-OfGn $\{Y^\delta(t)\}_{t\in\mathbb{R}}$ entièrement paramétré par $\Theta = (H_1, H_2, \rho_{\mathrm{x}}, \sigma_{\mathrm{x_1}}, \sigma_{\mathrm{x_2}}, \beta, \gamma)$, alors son spectre en ondelettes à l'échelle $2^j$ possède la forme suivante*

$$\mathbb{E}D_{\mathrm{y}^\delta}(j,k)D_{\mathrm{y}^\delta}(j,k)^* \equiv E^\delta(2^j,\Theta) = \begin{pmatrix} E_{11}^\delta(2^j,\Theta) & E_{12}^\delta(2^j,\Theta) \\ E_{12}^\delta(2^j,\Theta) & E_{22}^\delta(2^j,\Theta) \end{pmatrix}, \quad (6.5)$$

*avec*

$$(E_{11}^{\delta}(\Theta))_j = (1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\tilde{\eta}_{j,H_1}2^{j(2H_1+1+2\mu)} + 2\beta(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\tilde{\eta}_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + \beta^2(1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\tilde{\eta}_{j,H_2}2^{j(2H_2+1+2\mu)}, \quad (6.6)$$

$$(E_{12}^{\delta}(\Theta))_j = -\gamma(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\tilde{\eta}_{j,H_1}2^{j(2H_1+1+2\mu)} + (1-\beta\gamma)(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\tilde{\eta}_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + \beta(1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\tilde{\eta}_{j,H_2}2^{j(2H_2+1+2\mu)}, \quad (6.7)$$

$$(E_{22}^{\delta}(\Theta))_j = \gamma^2(1+\gamma^2)^{-1}\sigma_{\mathrm{x}_1}^2\tilde{\eta}_{j,H_1}2^{j(2H_1+1+2\mu)} - 2\gamma(1+\beta^2)^{-1/2}(1+\gamma^2)^{-1/2}\rho_{\mathrm{x}}\sigma_{\mathrm{x}_1}\sigma_{\mathrm{x}_2}\tilde{\eta}_{j,\frac{H_1+H_2}{2}}2^{j(H_1+H_2+1+2\mu)} + (1+\beta^2)^{-1}\sigma_{\mathrm{x}_2}^2\tilde{\eta}_{j,H_2}2^{j(2H_2+1+2\mu)}. \quad (6.8)$$

*et où*

$$\tilde{\eta}_{j,H} = \frac{1}{2}\int_{\mathbb{R}^2}\left(|u+2^{-j}|^{2H}+|u-2^{-j}|^{2H}-2|u|^{2H}\right)\psi_0(v)\psi_0(v-u)^*\mathrm{d}u\mathrm{d}v \quad (6.9)$$

Contrairement aux Biv-OfBms, le modèle paramétrique du spectre en ondelettes des Biv-OfGns comporte le terme $\tilde{\eta}_{j,H}$, dépendant de l'échelle $2^j$, à la place de $\eta_H$, défini en (5.25).

Pour une réalisation $Y^{\delta} \in \mathbb{R}^N$ d'un Biv-OfGn, nous rappelons également ci-dessous l'estimateur de la variance d'échantillon de $\mathbb{E}D_{\mathrm{y}^{\delta}}(j,\cdot)D_{\mathrm{y}^{\delta}}(j,\cdot)^*$ :

$$S^{\delta}(2^j) = \frac{1}{K_j}\sum_{k=1}^{K_j} D_{\mathrm{y}^{\delta}}(j,k)D_{\mathrm{y}^{\delta}}(j,k)^*, \quad K_j = \frac{N}{2^j}. \quad (6.10)$$

### 6.1.4 Procédure d'identification

Tout comme dans le cadre de l'étude des Biv-OfBms (voir Section 5.3.1), la procédure d'identification de Biv-OfGn est formulée par l'estimation de $\Theta$ à partir du problème de régression non linéaire suivant :

**Problème 6.1** (Régression non linéaire pour l'identification de Biv-OfGn)**.** *Soit $Y^{\delta} \in \mathbb{R}^N$ une réalisation d'un Biv-OfGn et $(j_1, j_2)$ tels que $1 \leq j_1 < j_2 \leq \log_2 N$. Le problème d'identification de Biv-OfGn est formulé comme le problème de régression non linéaire des coefficients d'ondelettes suivant*

$$\widehat{\Theta}_N^M = \underset{\Theta\in\mathcal{Q}_0}{\operatorname{argmin}}\, C_N(\Theta), \quad (6.11)$$

*avec*

$$C_N(\Theta) \equiv \sum_{\substack{i_1,i_2=1\\ i_1\leq i_2}}^{P=2}\sum_{j=j_1}^{j_2}\Big(\log_2|S_{i_1,i_2}^{\delta}(2^j)| - \log_2|E_{i_1,i_2}^{\delta}(2^j,\Theta)|\Big)^2, \quad (6.12)$$

*où $S^{\delta}(2^j)$ et $E^{\delta}(2^j,\Theta)$ désignent respectivement l'estimateur empirique* (6.10) *et le modèle paramétrique* (6.6)-(6.8) *du spectre en ondelettes de $Y^{\delta}$ à l'échelle $2^j$, et où l'espace de recherche $\mathcal{Q}_0$ est défini en* (5.33)*.*

La résolution du Problème 6.1 reprend quasiment la méthodologie proposée à la Section 5.3 pour la résolution du Problème 5.1 et dont les itérations sont indiquées dans l'Algorithme 8. Il convient seulement de remplacer les quantités $E(2^j, \Theta)$ et $S(2^j)$ par $E^\delta(2^j, \Theta)$ et $S^\delta(2^j)$.

Notons que la présence du terme $\tilde{\eta}_{j,H}$ dépendant de l'échelle $2^j$ conduit à des bornes inférieures plus grossières que dans le cas des Biv-OfBms. Ceci ce traduit en pratique par un nombre d'itérations plus élevé pour obtenir des performances d'estimation similaires. C'est le prix à payer pour bénéficier d'un modèle plus flexible que celui des OfBms.

### 6.1.5 Performances

**Configurations expérimentales.** Les performances d'estimation sont évaluées à partir de simulations de Monte Carlo. La procédure est appliquée à des copies indépendantes de Biv-OfGns synthétiques, réalisées par nos soins, dont la taille $N$ et les paramètres $\overline{\Theta}$ miment ceux observés et espérés pour le trafic internet. En outre, nous considérons un nombre d'échantillons $N = 3600$ particulièrement faible pour démélanger trois lois de puissances, des exposants de Hurst élevés, i.e., $1 > \overline{H}_2 \gtrsim \overline{H}_1 \gtrsim 0.7$, et une forte corrélation $\overline{\rho} = 0.8$.

Nous avons précédemment fait l'hypothèse que $|H_2 - H_1| > 0$ ou $(\beta \neq 0, \gamma \neq 0)$ pouvaient être des signatures d'éventuelles anomalies. Nous proposons donc de considérer diverses configurations selon que $(H_2 = H_1 = 0.7)$ (Figure 6.1, haut) ou $(H_2 = 0.9, H_1 = 0.7)$ (Figure 6.1, bas) et selon que $(\beta = \gamma = 0$, i.e., aucun mélange) (Figure 6.1, gauche) ou non (Figure 6.1, droite).

**Performances d'estimation.** Les performances d'estimation sont quantifiées Figure 6.1 à partir de diagrammes de dispersion. Dans l'ensemble, $H_1$ et $H_2$ sont toujours correctement estimés. Il convient de noter qu'estimer l'exposant non-dominant $H_1$ est naturellement plus difficile lorsque $H_2 > H_1$ (Figure 6.1, bas). Toutefois, l'estimation est suffisamment de bonne qualité pour pouvoir discriminer les cas $H_1 = H_2$ et $H_1 < H_2$.

Cependant, les paramètres de mélange $(\beta, \gamma)$ et la corrélation $\rho$ restent difficiles à estimer pour de courtes séries temporelles. Par conséquent, il peut être difficile de tester en pratique la présence de mélange, i.e., $\beta \neq 0$ et $\gamma \neq 0$.

## 6.2 Application aux données internet

### 6.2.1 Données MAWI et projections aléatoires

**Base de données MAWI.** L'archive MAWI [1, 42] répertorie un large ensemble de traces du trafic internet collectées récemment au sein du réseau d'infrastructure *WIDE* (AS2500) reliant les universités et les instituts de recherche japonais à internet. Les quintuplets des paquets IP collectés chaque jour entre 14h00 et 14h15 (heure normale du Japon) sont horodatés, anonymisés et rendus disponibles publiquement. Chaque trace contient approximativement entre 100 à 150 millions de paquets IP.

**Projections aléatoires.** Tel que discuté en introduction, une question majeure dans la détection d'anomalies consiste à définir et à calculer une référence normale du trafic à laquelle se comparer. Cependant, à cause de la nature constamment changeante du trafic internet, il n'est pas pertinent de se comparer à toute autre trace collectée un autre jour ou à partir d'un autre réseau [62, 27]. A la place, l'utilisation des projections aléatoires (également appelés « sketchs ») [123] a montré être une procédure pertinente pour la construction auto-référencée de trafics normaux.

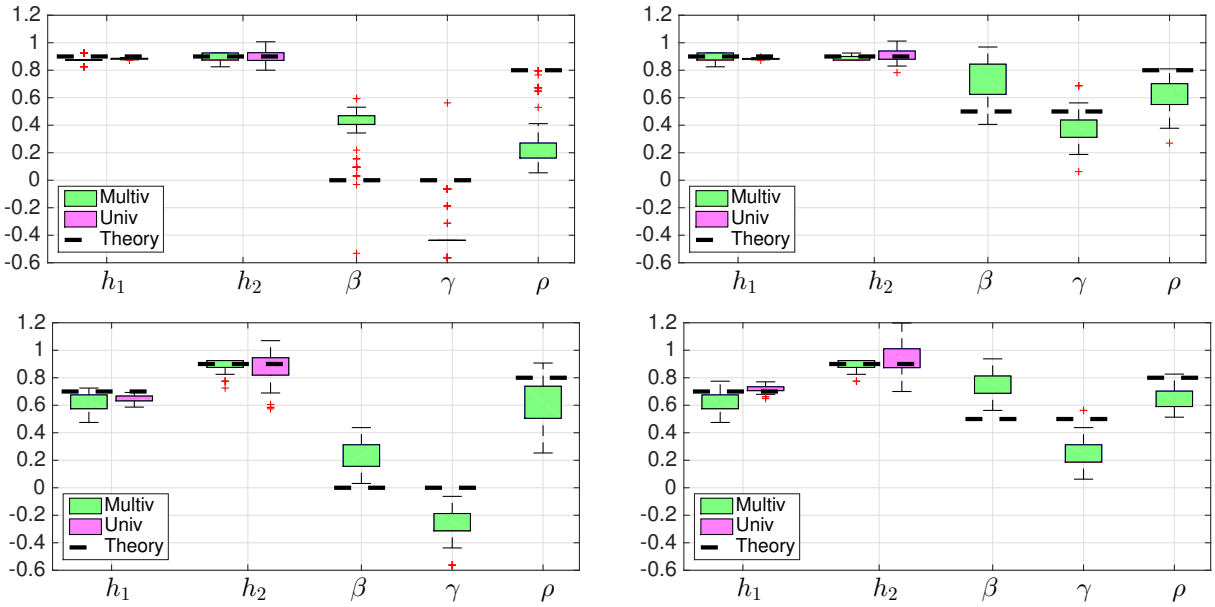



FIGURE 6.1 – **Performances d'estimation sur des OfGns synthétiques** modélisant quatre configurations mimant potentiellement le trafic internet : $(\overline{H}_1, \overline{H}_2) = (0.9, 0.9)$ ou $(0.7, 0.9)$ et $(\overline{\beta}, \overline{\gamma}) = (0, 0)$ ou $(0.5, 0.5)$.

Le principe des projections aléatoires consiste à assigner aléatoirement chaque paquet IP à l'une des $M$ entrées d'une table de hachage selon un des ses attributs, ici l'adresse IP source [198]. Par conséquent, tous les paquets de n'importe quel flux avec la même adresse IP source seront tous alloués aléatoirement à la même entrée. Le trafic est alors découpé en $M$ sous-trafics.

Lorsque le trafic ne contient aucune anomalie, on s'attend à ce que tous les sketchs soient statistiquement équivalents. Quand le trafic contient quelques anomalies, associées à la même adresse IP source, tous les paquets malicieux correspondants sont associés au même sketch et les autres sketchs ne contiennent aucune anomalie. La comparaison entre les $M$ sketchs peut ensuite être utilisée pour concevoir une référence statistique pour un trafic normal, et permet par conséquent de pouvoir détecter des anomalies.

### 6.2.2 Rôle du paramètre d'intégration fractionnaire

Alors que dans le modèle théorique des Biv-OfGns les exposants vérifient $0 < H_1 \leq H_2 < 1$, en pratique les $H$ estimés pour un trafic normal peuvent prendre de grandes valeurs proches de 1 [62, 27]. Ceci soulève d'importantes questions en pratique quant à l'utilisation de la procédure séparation-évaluation décrite à la Section 5.3, puisque certaines données issues du monde réel peuvent aléatoirement et par accident mener à des valeurs $H \notin [0, 1]$. Ce comportement est illustré par les estimées univariées de $H$ rapportées (Figure 6.3), puisqu'une estimation univariée basée sur la régression linéaire des coefficients d'ondelettes ne nécessite pas d'imposer $H \in [0, 1]$ (cf. e.g., [2, 62, 27]).

Afin de contourner ce problème, nous proposons dans ce chapitre l'introduction d'un paramètre additionnel d'intégration fractionnaire $\mu$ dans la définition des coefficients d'ondelettes (6.4). Il en découle que $\mu$ apparait également dans le modèle paramètre du spectre en ondelettes (5.27)-(5.29). Plutôt que choisir un seul coefficient $\mu$ (usuellement choisi $\mu = 1$), nous proposons de découpler ses valeurs comme suit

**Proposition 6.2** (Intégration fractionnaire de Biv-OfGn). *Soit $Y^\delta \in \mathbb{R}^N$ une réalisation d'un Biv-OfGn. Soit $S^\delta$ son spectre en ondelette empirique calculé à partir des coefficients en ondelettes $D_{y^\delta}$ définis en* (6.4) *pour le paramètre d'intégration fractionnaire $\mu_W$ et soit son modèle paramétrique $E^\delta$ défini en* (6.6)-(6.8) *pour le paramètre d'intégration fractionnaire $\mu_B$. Alors les estimées $\widehat{H}_1$ et $\widehat{H}_2$ obtenues en résolvant le Problème 6.1 sont translatées par la quantité $-(\mu_B - \mu_W)$.*

Cette opération revient à intégrer fractionnairement les données pour assurer que $H_1$ et $H_2$ se situent bien au centre de l'intervalle $[0, 1]$. Il suffit ensuite de translater les estimées de $H_1$ et $H_2$ par la quantité $\mu_B - \mu_W$. Etant donné que nous nous intéresserons principalement à la quantité $H_2 - H_1$, cela n'aura aucun impact sur nos résultats et conclusions. Nous envisageons la possibilité de découler les paramètres $\mu_W$ et $\mu_B$ comme une astuce pratique et efficace pour adapter le modèle aux données issues du monde réel.

Dans la suite, les coefficients d'ondelettes sont calculés pour $\mu_W = 1/2$ alors que le spectre en ondelettes est évalué pour $\mu_B = 1$.

### 6.2.3 Résultats

**Configuration expérimentale.** Les résultats présentés dans cette étude correspondent à quatre exemples typiques et récents du trafic internet collectés en quatre jours différents en 2008, 2013, 2014 et 2015 respectivement. Des tables de hachage à $M = 16$ sketchs sont utilisées. Les sketchs sont agrégés à $\Delta_0 = 0.25$s, comme il est bien documenté que l'autosimilarité dans le trafic internet se développe depuis l'échelle de la seconde jusqu'aux heures [2, 62, 27]. Des examples de sketchs agrégés sont illustrés sur la Figure 6.2 pour le nombre de paquets (1ère ligne) et de bytes (5ème ligne). L'analyse en ondelettes est conduite en utilisant des ondelettes orthogonales de Daubechies les moins asymétriques avec $N_\psi = 2$ moments nuls [139]. La minimisation de (6.11) est conduite en utilisant les échelles $j_1 = 3$ à $j_2 = 8$ correspondant respectivement aux échelles de temps allant de 2s à 1min (les données disponibles sont limités à 15min d'enregistrement). Bien que l'identification complète des Biv-OfGns requiert 7 paramètres, nous nous intéresserons ici aux 5 paramètres les plus intéressants, c'est-à-dire $\overline{\Theta} = (\overline{H}_1, \overline{H}_2, \overline{\beta}, \overline{\gamma}, \overline{\rho}_{\mathrm{x}})$ et nous considérerons $\overline{\sigma}_{\mathrm{x}_1} = \overline{\sigma}_{\mathrm{x}_2} = 1$ connu par normalisation *a priori* des données. L'estimation de $\overline{\Theta}$ est faite sur chaque sketch de chaque ensemble de données à partir de la procédure de séparation-évaluation.

**Métadonnées et vérité terrain.** Les métadonnées des anomalies présentes sur le trafic internet MAWI sont obtenues par des experts à l'aide d'une procédure informatique, le MawiLab [84], qui inspecte de façon systématique et automatique le contenu du trafic internet à l'aide de plusieurs estimateurs d'anomalies de l'état de l'art. Il convient de mentionner que ces métadonnées ne constituent pas la vérité terrain mais sont seulement des indications à comparer avec les résultats de la procédure de détection proposée.

**Modélisation statistique du trafic internet par un Biv-OfGn.** La Figure 6.2 illustre *a posteriori* que la statistique jointe du nombre de paquets et de bytes estimée empiriquement en utilisant $S^\delta_{p,p'}(j)$, est bien modélisé par $E^\delta_{p,p'}(\widehat{\Theta})$, avec les paramètres $\widehat{\Theta}$ estimés à partir de la procédure de séparation-évaluation. Elle montre également que les Biv-OfGns modélisent de manière aussi efficace les sketchs avec ou sans anomalies, mais avec des paramètres estimés $\widehat{\Theta}$ clairement différents. La Figure 6.2 montre que la paramétrisation à l'aide de 5 paramètres au lieu de 7 est satisfaisante.

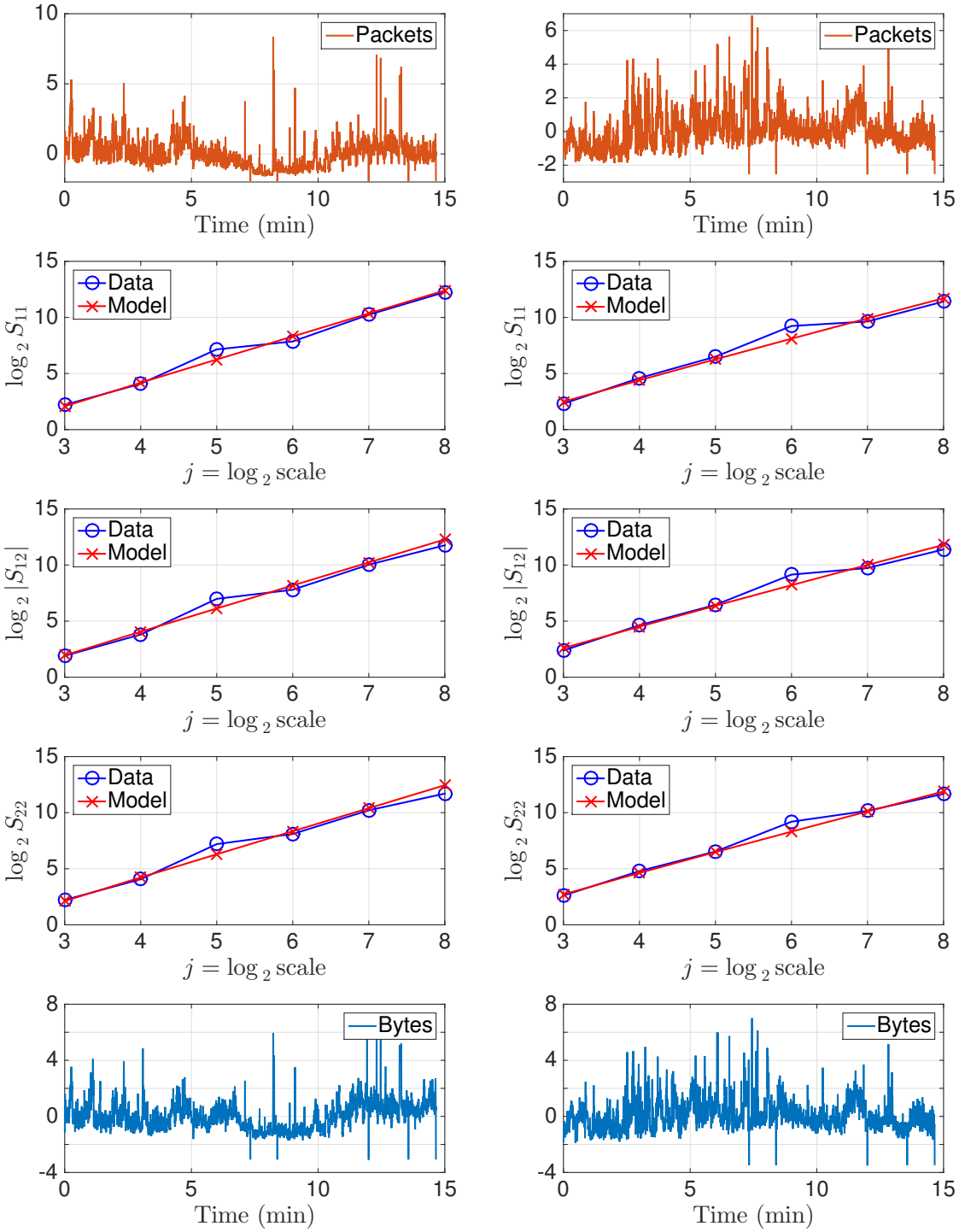
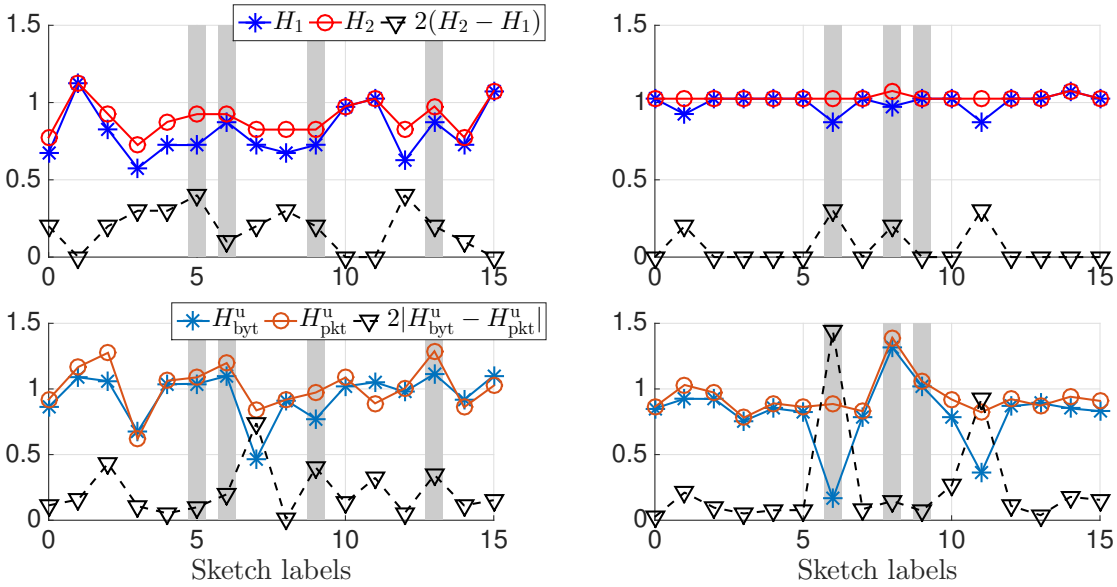
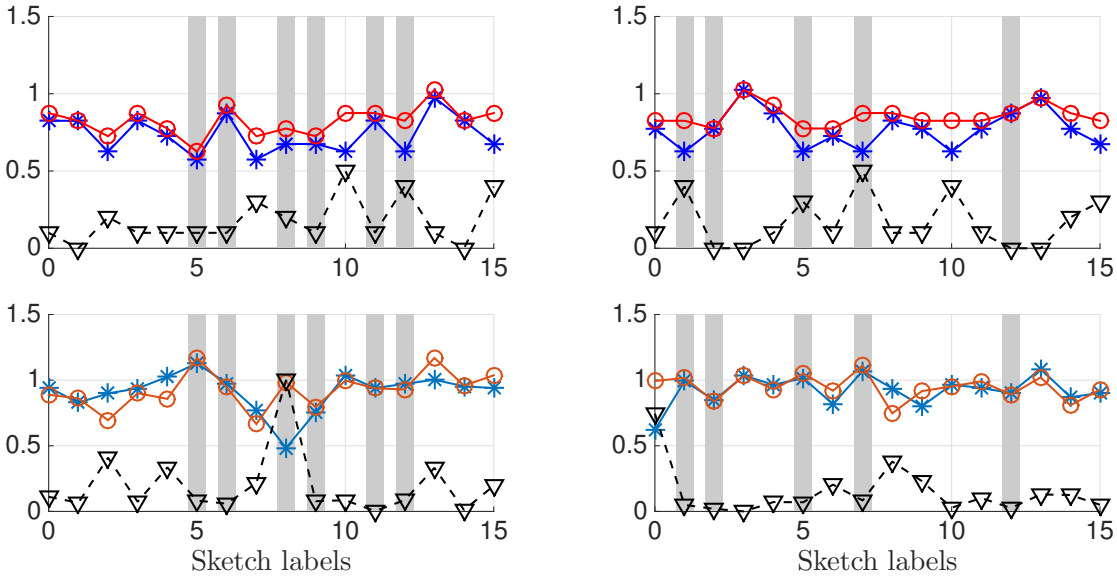



FIGURE 6.2 – **Modèle Biv-OfGn pour les séries temporelles du nombre agrégé de paquets et de bytes reçus.** 1ère et 5e lignes : séries temporelles normalisées du nombre de paquets et de bytes reçus ; 2ème et 4ème lignes : comparaisons entre $\log_2 S_{p,p'}$ (bleu) vs. $\log_2 E_{p,p'}(\hat{\Theta})$ (rouge) pour $(p, p') = (1, 1)$, $(1, 2)$ et $(2, 2)$. Sketch sans (gauche) ou avec (droite) anomalies.




(a) 2008 (b) 2013




(a) 2014 (b) 2015

FIGURE 6.3 – **Détection d'anomalies à partir de $H_2 - H_1$.** Les lignes verticales grises indiquent les sketchs marqués par les experts comme contenant le plus grand nombre de paquets anormaux pour chaque données collectées en 2008 (a), 2013 (b), 2014 (c) et 2015 (c). Pour chaque groupe de figure nous comparons l'analyse bivariée contre l'analyse univariée. Haut : analyse bivariée, $H_2$, $H_1$ et $|H_2 - H_1|$ (multiplié par un facteur 2 par souci de lisibilité) en fonctions du numéro de chaque sketch. Bas : analyse univariée, $H^U_{Byt}$, $H^U_{Pkt}$ and $|H^U_{Byt} - H^U_{Pkt}|$ (multiplié par un facteur 2 par souci de lisibilité) en fonctions du numéro de chaque sketch.

**Détection d'anomalies.** Les paramètres $\widehat{\Theta}$ estimés pour chaque sketch et chaque trace sont comparés avec les métadonnées des anomalies obtenues par le MawiLab. Les premières inspections, non reportées ici, montrent que la corrélation $\widehat{\rho}$ et les paramètres de mélange $(\widehat{\beta}, \widehat{\gamma})$ estimés ne semblent pas corréler correctement avec les occurrences d'anomalies. Cela peut être expliqué en partie par la pauvre qualité d'estimation de ces paramètres, tels qu'observée dans les simulations de Monte Carlo dans une configuration expérimentale semblable. Cette analyse montre cependant clairement que les paramètres de mélange $\beta$ et $\gamma$ dévient significativement de 0, ce qui démontre le besoin d'une estimation bivariée pour une estimation non biaisée des paramètres $\overline{H}_1$ et $\overline{H}_2$. La comparaison entre les lignes du haut et du bas de la Figure 6.3(a)-(d) souligne une différence significative entre les estimées univariées $(\widehat{H}^U_{\text{byt}}, \widehat{H}^U_{\text{pkt}})$ et bivariées $(\widehat{H}_1, \widehat{H}_2)$, ce qui illustre clairement les limitations et les biais d'une analyse univariée.

L'inspection de $\widehat{H}_1$ et $\widehat{H}_2$ estimés par l'algorithme de séparation-évaluation tend à montrer que de grandes deviations $|\widehat{H}_2 - \widehat{H}_1|$ de 0, concordent avec un nombre significatif des sketchs marqués par le MawiLab comme contenant un grand nombre de paquets anormaux.

Une inspection minutieuse de la Figure 6.3 montre que l'accord entre les déviations de 0 et les occurences d'anomalies n'est pas parfait, cela peut avoir deux origines : Comme il l'a été mentionné précédemment, les métadonnées ne sont pas la vérité terrain et certaines anomalies peuvent ne pas avoir été détectées par la procédure du MawiLab. Alternativement, certaines anomalies relevées par le MawiLab peuvent ne pas être caractérisées par $|\overline{H}_2 - \overline{H}_1| \gg 0$, qui peut en retour être la signature de sous-classes spécifiques d'anomalies. Ces désaccords nécessitent une inspection plus poussée de la part des experts et fera l'objet d'une prochaine étude. Toutefois, l'inspection des experts indique que :

— L'anomalie *Trinoculaire*, spécifique à une expérience de réseau informatique to a computer network experiment en cours d'exécution sur le réseau MAWI [177], est systématiquement détectée.

— 96% des attaques par déni de service ont été détectées en 2013.

— 100% des anomalies « Heavy Hitter » ont été détectées en 2014.

Les résultats de cette étude sont très encourageants et cohérents avec le mécanisme de file d'attente reliant l'autosimilarité à la distribution à queue lourde des objets sur internet proposée en [57, 197, 28] qui prédit des exposants de Hurst identiques pour les paquets et les bytes. Ils sont également cohérents avec de précédents résultats empiriques reposant sur l'analyse univariée des paquets et des bytes qui tend à réconforter l'idée que $H_{\text{byt}} \simeq H_{\text{pkt}}$ pour un trafic normal, alors que des déviations de $|H_{\text{byt}} - H_{\text{pkt}}|$ de 0 peuvent indiquer la présence d'anomalies [62, 27].

## 6.3 Conclusion et perspectives

Ce chapitre promeut la modélisation jointe de l'autosimilarité du nombre de paquets et de bytes reçus par un Biv-OfGn dont les paramètres sont estimés via la résolution d'un problème de régression non linéaire par un algorithme de séparation-évaluation. Des simulations numériques conduites sur des copies indépendantes de Biv-OfGns synthétiques dont les paramètres miment ceux observés en pratique sur le trafic internet montrent que la procédure proposée délivre des performances d'estimation de $H_1$ et $H_2$ satisfaisantes, alors que les paramètres $\rho$ et $(\beta, \gamma)$ sont plus difficiles à estimer dans de telles configurations. Les routines MATLAB permettant à la fois l'identification et la synthèse de Biv-OfGns sont disponibles en ligne à l'adresse `http://perso.ens-lyon.fr/jordan.frecon`.

Nous avons également montré que les propriétés d'invariances d'échelle observées dans le trafic internet, à la fois pour les séries temporelles du nombre de paquets et de bytes, sont pertinemment modélisées par des Biv-OfGns. Les Biv-OfGns sont d'ailleurs tout aussi pertinents pour modéliser les trafic normaux ou anormaux. Une procédure de détection d'anomalies, dérivée à partir de l'examination des paramètres $\Theta$ du modèle, apporte l'évidence que certains types d'anomalies sont caractérisés par une déviation $|H_2 - H_1| \gg 0$.

Les résultats sont suffisamment satisfaisants pour envisager d'appliquer systématiquement la procédure d'identification proposée à plus grande échelle, avec un ensemble de métadonnées enrichi, afin de ne pas permettre seulement la détection d'anomalies mais également d'être en mesure de les classifier à partir de leurs propriétés d'invariance d'échelle.

# Quatrième partie

# Développements méthodologiques autour de la TV

# Chapitre 7

# Processus de Poisson homogènes par morceaux

**Sommaire**


**7.1 Présentation du problème . . . . . 134**
7.1.1 Modèle . . . . . 134
7.1.2 Impact de l'aggrégation . . . . . 135
**7.2 Statistique Poissonienne vs. Gaussienne . . . . . 136**
7.2.1 Divergence de Kullback-Leibler . . . . . 136
7.2.2 Transformée d'Anscombe . . . . . 137
**7.3 Méthodologies . . . . . 139**
7.3.1 Méthode du seuil . . . . . 139
7.3.2 Méthodes d'estimation par approche TV . . . . . 140
7.3.2.1 Estimées par approche TV . . . . . 140
7.3.2.2 Réestimation à 2 niveaux . . . . . 141
7.3.3 Méthode de segmentation par approche TV . . . . . 142
**7.4 Comparaison des différentes méthodes . . . . . 142**
7.4.1 Configuration expérimentale . . . . . 142
7.4.2 Quantification des performances . . . . . 144
**7.5 Conclusion et perspectives . . . . . 146**


**Fluorescence intermittente.** Nous proposons de nous intéresser au phénomène de fluorescence intermittente [161, 108] où un émetteur nanoscopique (e.g., boîte quantique) alterne aléatoirement entre deux états d'émission : l'un fortement émissif (appelé ON) et l'autre faiblement émissif (appelé OFF). Une caractéristique particulière de ce phénomène concerne le comportement en lois de puissance de la distribution de la durée des états ON et OFF. Alors que de nombreux efforts théoriques continuent d'être investis afin de comprendre la raison de ce comportement (cf., e.g., les références mentionnées dans [108]), les performances d'estimation et le coût de calcul prohibitif des méthodes d'estimation actuelles demeurent un frein à l'interprétation et à la validation des exposants de lois de puissance mesurés expérimentalement. Le présent chapitre se place dans ce contexte et examine l'apport de plusieurs stratégies visant à détecter en amont les états ON des états OFF.

**Détection de discontinuités d'intensité.** Jusqu'à présent, les méthodes d'estimation de discontinuités et de segmentation que nous avons présentées au Chapitre 2 reposaient sur

l'hypothèse que l'observation $\boldsymbol{y} \sim \mathcal{N}(\overline{\boldsymbol{x}}, \sigma^2 \mathbb{1}_N)$ où $\overline{\boldsymbol{x}}$ est constant par morceaux. Dans ce chapitre, nous nous intéressons au cas où l'observation $Y$ est une réalisation d'un processus de Poisson, i.e., $Y \sim \mathcal{P}(\overline{X})$, dont l'intensité $\overline{X}$ alterne aléatoirement entre $\overline{\mu}_{\mathrm{ON}}$ et $\overline{\mu}_{\mathrm{OFF}}$. La source de bruit dans l'observation $Y$ est ici seulement due à la nature stochastique du processus de comptage et nous ne tiendrons pas compte de la présence de source externe de bruit où $Y \sim \mathcal{P}(\overline{Y}) + B$. Le problème que nous considérons ici consiste à estimer les discontinuités de $\overline{X}$ à partir du compte du nombre d'occurrences de $Y$ dans des intervalles de temps réguliers $\Delta$.

**Travaux préliminaires.** Une méthode courante largement utilisée dans la pratique repose sur la sélection d'un seuil permettant ainsi de discriminer l'état ON de l'état OFF. Cependant, les performances d'estimation peuvent largement varier selon l'intensité et le choix de $\Delta$ [56].

Ce problème a également reçu l'intérêt de la communauté statistique par le développement d'approches fréquentistes et bayésiennes [220, 222, 111, 183]. Notamment, une stratégie en deux temps a été proposée dans [211] où les discontinuités sont estimées par un test de rapport de vraisemblance généralisé, puis les niveaux d'intensité similaires sont regroupés par un algorithme de classification hiérarchique. Récemment, l'emploi d'un modèle hiérarchique Bayésien dans [68] a permis d'estimer conjointement les positions des discontinuités, le nombre de niveaux et les valeurs des intensités. Cependant, cette approche est très coûteuse en temps de calcul et son utilisation peut s'avérer prohibitive pour l'analyse de données de grande taille.

Ce problème peut également être formulé par des approches TV dont la solution peut être obtenue efficacement par des méthodes proximales [175, 173, 174, 31, 149]. En contrepartie, les performances d'estimation dépendent fortement du choix du paramètre de régularisation.

**Contribution.** Nous proposons d'évaluer les performances des approches TV pour la détection des discontinuités d'intensités et l'estimation de la statistique des temps ON et OFF. Après avoir détaillé le modèle dans la Section 7.1, nous discuterons dans la Section 7.2 des conséquences engendrées par la statistique de Poisson des observations et présenterons comment sont adaptées les solutions algorithmiques présentées au Chapitre 2 ou comment « gaussianniser » les données pour s'y rapporter. La présentation des différentes stratégies fait l'objet de la Section 7.3 et les performances d'estimation sont comparées dans la Section 7.4 sur des données synthétiques.

## 7.1 Présentation du problème

### 7.1.1 Modèle

**Processus de Poisson à 2 niveaux intensité.** Nous considérons le problème où un processus de Poisson alterne entre deux états d'émission que nous noterons ON et OFF.

Pendant l'état ON (resp. OFF), le processus est hautement (resp. pauvrement) émissif avec pour densité d'émission $\overline{X} \equiv \overline{\mu}_{\mathrm{ON}}$ (resp. $\overline{\mu}_{\mathrm{OFF}}$). A l'instant $t$, nous notons $Y(t) = 1$ si une occurrence a lieu (e.g. un photon a été émis) et 0 sinon. Par conséquent, si la densité d'émission $\overline{X} \equiv \mu$ est constante pendant le laps de temps $I = [t, \dots, t+\tau[$, alors la probabilité de détecter $k$ occurrences dans $I$ est donnée par la loi de Poisson

$$g(k, \mu, \tau) = \frac{(\mu\tau)^k \mathrm{e}^{-\mu\tau}}{k!}. \tag{7.1}$$

Motivées par l'application réelle, les durées respectives des états ON et OFF sont sup-

posées être distribuées selon des lois de puissance comme suit :

$$\tau_{\mathrm{ON}} \sim \mathrm{Pareto}_{(\theta, m_{\mathrm{ON}}-1)}, \tag{7.2}$$

$$\tau_{\mathrm{OFF}} \sim \mathrm{Pareto}_{(\theta, m_{\mathrm{OFF}}-1)}, \tag{7.3}$$

où $\theta$ désigne le temps de coupure, i.e., la taille minimale des temps ON et OFF, et $m_{\mathrm{ON}}$ et $m_{\mathrm{OFF}}$ sont les exposants des lois de puissance.

**Données agrégées.** Dans la pratique, les occurrences sont comptées dans $N$ intervalles de temps successifs de durée $\Delta$, i.e.,

$$(\forall n \in \{1, \ldots, N\}) \quad y_{\Delta,n} = \int_{n\Delta}^{(n+1)\Delta} Y(t)\mathrm{d}t. \tag{7.4}$$

Dès lors, la densité d'émission agrégée

$$(\forall n \in \{1, \ldots, N\}) \quad \overline{x}_{\Delta,n} = \frac{1}{\Delta} \int_{n\Delta}^{(n+1)\Delta} \overline{X}(t)\mathrm{d}t \tag{7.5}$$

possède plus d'états que les deux seuls états ON et OFF. Si durant le $n$-ème intervalle de temps,

$$\overline{x}_{\Delta,n} = \alpha\overline{\mu}_{\mathrm{ON}} + (1-\alpha)\overline{\mu}_{\mathrm{OFF}} \tag{7.6}$$

avec $\alpha \in [0, 1]$, alors

$$y_{\Delta,n} \sim \mathcal{P}((\alpha\overline{\mu}_{\mathrm{ON}} + (1-\alpha)\overline{\mu}_{\mathrm{OFF}})\Delta). \tag{7.7}$$

**Problème.** Idéalement, nous souhaiterions retrouver l'évolution temporelle de $\overline{X}$, mais la seule quantité dont nous disposons est le nombre d'occurrences $\boldsymbol{y}_\Delta$ dans des intervalles de temps sucessifs de taille $\Delta$. Par conséquent, dans les cas pratiques où $\Delta > \theta$ il est tout à fait inenvisageable de pouvoir estimer $\overline{X}$. Toutefois, sous certaines conditions discutées dans la section suivante, nous pouvons prétendre à l'estimation de la distribution de $\tau_{\mathrm{ON}}$ et $\tau_{\mathrm{OFF}}$ pour des durées supérieures à $\Delta$.

### 7.1.2 Impact de l'aggrégation

Dans cette section, nous évaluons l'impact de l'aggrégation, i.e., du rapport $\theta/\Delta$, sur la proportion moyenne d'états où $\overline{\boldsymbol{x}}_\Delta = \overline{\mu}_{\mathrm{ON}}$ ou $\overline{\mu}_{\mathrm{OFF}}$. Nous choisissons le cas particulier où $\overline{\mu}_{\mathrm{ON}}/\overline{\mu}_{\mathrm{OFF}} = 13$. Pour faciliter la discussion, nous supposerons que $m_{\mathrm{ON}} = m_{\mathrm{OFF}} = m$.

A partir de (7.2) et (7.3), si $m > 2$ alors la durée moyenne des temps ON et OFF vaut $(m-1)\theta/m$. Autrement, la distribution des temps ON et OFF est à queue lourde et aucun temps moyen n'existe. Cette observation, qui peut paraître anodine, est importante pour comprendre l'allure de $I_\Delta$. Nous articulerons donc la discussion autour de $m = 2.2$ et $m = 1.8$ en guise d'exemples représentatifs de $m > 2$ et $m < 2$.

Pour $m = 2.2$, la distribution de $\overline{\boldsymbol{x}}_\Delta$ est représentée en bleu pour $\theta/\Delta = 10$ à 0.01 par les Figures 7.1(a)-(d). Plus le rapport $\theta/\Delta$ diminue, plus un grand nombre d'états ni ON ni OFF apparaissent. Si l'on omet les cas où $\overline{\boldsymbol{x}}_\Delta = \overline{\mu}_{\mathrm{ON}}$ ou $\overline{\mu}_{\mathrm{OFF}}$, alors la distribution de $\overline{\boldsymbol{x}}_\Delta$ a une forme de plus en plus « triangulaire » lorsque $\theta/\Delta$ diminue : il est plus probable d'avoir autant d'états ON et OFF dans chaque intervalle $\Delta$ que d'avoir un état prédominant.

Pour $m = 1.8$ (orange), l'allure de la distribution de $\overline{\boldsymbol{x}}_\Delta$ est totalement différente : quel que soit le rapport $\theta/\Delta$, il existe une majorité d'états exclusivement ON ou OFF dans chaque intervalle $\Delta$.

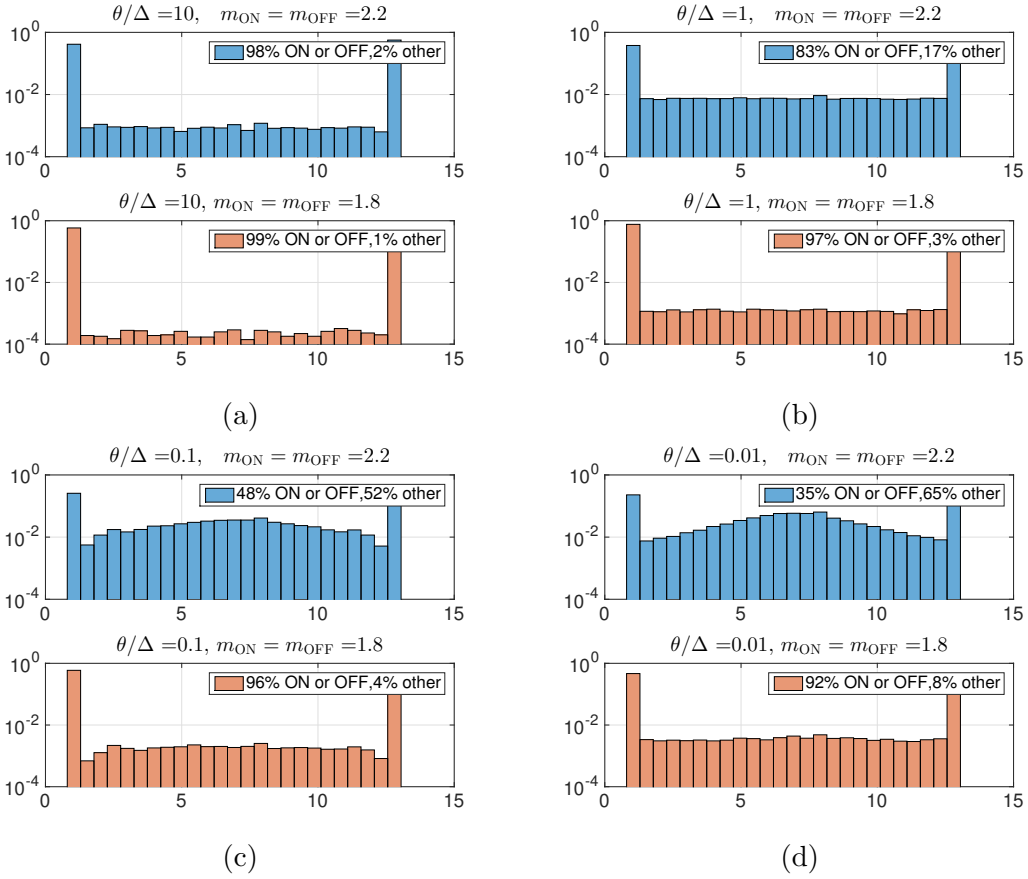



FIGURE 7.1 – **Proportion de temps ON et OFF.** Distribution à queue lourde (orange) ou non (bleu). Impact du rapport $\theta/\Delta$=10, 1, 0.1 et 0.01 de (a) à (d).

Par conséquent, si $m_{\text{ON}} \leq 2$ et $m_{\text{OFF}} \leq 2$, alors nous pouvons prétendre à l'estimation des distributions des temps ON et OFF à partir de $\boldsymbol{y}_\Delta$ lorsque $\theta/\Delta \leq 1$. Dans la suite de ce chapitre, nous nous plaçerons dans ce cas.

## 7.2 Statistique Poissonienne vs. Gaussienne

Les approches TV que nous avons présentées (cf., e.g., Problème 2.2) nous ont permis d'inclure l'*a priori* Gaussien des données, i.e., $\boldsymbol{y} \sim \mathcal{N}(\overline{\boldsymbol{x}}, \sigma^2)$, à partir du choix du terme d'attache aux données $\|\boldsymbol{y} - \boldsymbol{x}\|_2^2$.

Ici, afin de formuler le problème d'estimation des discontinuités de $\overline{\boldsymbol{x}}_\Delta$ par des approches TV, nous proposons deux stratégies. La première consiste à modifier le terme d'attache aux données pour prendre en compte le nouvel *a priori* $\boldsymbol{y}_\Delta \sim \mathcal{P}(\overline{\boldsymbol{x}}_\Delta)$. La seconde consiste à « gaussianniser » $\boldsymbol{y}_\Delta$ ce qui nous permettra d'utiliser les approches précédentes basées sur un terme d'attache aux données $\ell_2$.

### 7.2.1 Divergence de Kullback-Leibler

Le terme d'attache aux données vise à modéliser la vraisemblance de l'observation $\boldsymbol{y}_\Delta$ étant donné une solution candidate $\boldsymbol{x}_\Delta$. Ici, pour tout $n \in \{1, \ldots, N\}$, nous souhaiterions trouver l'intensité $x_{\Delta,n}$ qui maximise le rapport entre la probabilité de détecter $y_{\Delta,n}$ occurrences issues d'une distribution de Poisson d'intensité $x_{\Delta,n}$ et la probabilité de détecter $y_{\Delta,n}$

occurrences issues d'une distribution de Poisson d'intensité $y_{\Delta,n}$, i.e.,

$$\widehat{\boldsymbol{x}}_\Delta = \arg\max_{\boldsymbol{x}_\Delta \in \mathbb{R}^N} \prod_{n=1}^{N} \frac{x_{\Delta,n}^{y_{\Delta,n}} \mathrm{e}^{-x_{\Delta,n}}}{y_{\Delta,n}^{y_{\Delta,n}} \mathrm{e}^{-y_{\Delta,n}}}. \tag{7.8}$$

De manière équivalente, nous pouvons minimiser l'opposé du logarithme de la quantité précédénte, i.e.,

$$\widehat{\boldsymbol{x}}_\Delta = \arg\min_{\boldsymbol{x}_\Delta \in \mathbb{R}^N} \mathrm{DKL}(\boldsymbol{x}_\Delta, \boldsymbol{y}_\Delta) \tag{7.9}$$

où DKL est la divergence de Kullback-Leibler dont nous rappellons la définition :

**Définition 7.1** (Divergence de Kullback-Leibler [124])**.** *On appelle divergence de Kullback-Leibler dirigée de $\boldsymbol{x} \in \mathbb{R}^N$ vers $\boldsymbol{y} \in \mathbb{R}^N$, la quantité définie par*

$$\mathrm{DKL}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{n=1}^{N} \phi_n(x_n) \tag{7.10}$$

*où pour tout $n \in \{1, \dots, N\}$*

$$\phi_n(x_n) = \begin{cases} y_n \log\left(\frac{y_n}{x_n}\right) + x_n - y_n & \text{si } y_n > 0 \text{ et } x_n > 0, \\ x_n & \text{si } y_n = 0 \text{ et } x_n \geq 0, \\ +\infty & \text{autrement.} \end{cases} \tag{7.11}$$

Afin de prendre en compte l'hypothèse que $\widehat{\boldsymbol{x}}_\Delta$ est constant par morceaux, nous rajouterons dans (7.9) un terme de variation totale.

### 7.2.2 Transformée d'Anscombe

Soit $\boldsymbol{y}_\Delta \sim \mathcal{P}(\overline{\boldsymbol{x}}_\Delta)$ une variable aléatoire distribuée selon une loi de Poisson de taux $\overline{\boldsymbol{x}}_\Delta$. Nous souhaiterions trouver une transformation $f$ telle que $\tilde{\boldsymbol{y}}_\Delta = f(\boldsymbol{y}_\Delta) \sim \mathcal{N}(\tilde{\boldsymbol{x}}, \sigma^2 \mathbb{1}_N)$ de valeur moyennes $\tilde{\boldsymbol{x}}_\Delta$.

**Proposition 7.1** (Transformée exacte)**.** *Soit $y \sim \mathcal{P}(\overline{x})$ une variable aléatoire distribuée selon une loi de Poisson de paramètre $\overline{x}$. Alors la quantité $\tilde{y} = f_{(\tilde{x},\sigma^2,\overline{x})}(y)$ où*

$$f_{(\tilde{x},\sigma^2,\overline{x})}(y) = \sigma\sqrt{2}\mathrm{erf}^{-1}\left(2\frac{\Gamma(\lfloor y+1 \rfloor, \overline{x})}{\lfloor y \rfloor !} - 1)\right) + \tilde{x}, \tag{7.12}$$

*est tirée selon une distribution Gaussienne de moyenne $\tilde{x}$ et de variance $\sigma^2$.*

*Démonstration.* Soient $F_y : k \to \mathbb{P}[y \leq k]$ et $F_{\tilde{y}} : \tilde{k} \to \mathbb{P}[\tilde{y} \leq \tilde{k}]$ les fonctions de distribution cumulative respectivement associées à $y$ et $\tilde{y}$. Alors, pour tout $k \in \mathbb{N}$,

$$F_y(k) = \mathbb{P}[y \leq k] = \mathbb{P}[f^{-1}_{(\tilde{x},\sigma^2,\overline{x})}(\tilde{y}) \leq k] \tag{7.13}$$

$$= \mathbb{P}[\tilde{y} \leq f_{(\tilde{x},\sigma^2,\overline{x})}(k)] = F_{\tilde{y}}(f_{(\tilde{x},\sigma^2,\overline{x})}(k)), \tag{7.14}$$

et par conséquent

$$f_{(\tilde{x},\sigma^2,\overline{x})} = F^{-1}_{\tilde{y}} \circ F_y. \tag{7.15}$$

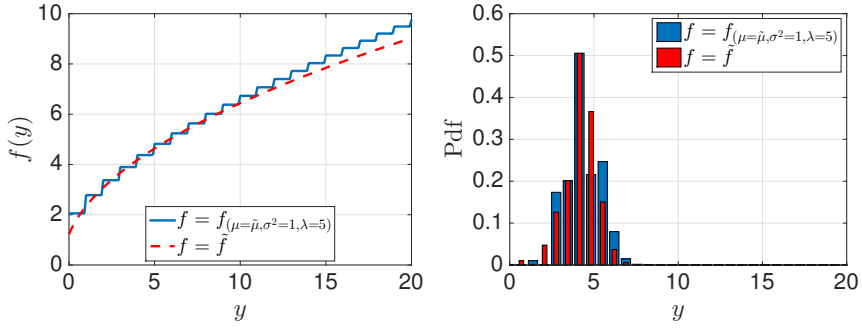



FIGURE 7.2 – **Transformée d'Anscombe**. Gauche : comparaison entre la transformation exacte (bleu) et la transformée d'Anscombe (rouge). Droite : comparaison de la qualité de la distribution gaussienne après chacune des deux transformations.

Si nous rappellons

$$F_y(k) = P_{(\overline{x})}(k) = \frac{\Gamma(\lfloor k+1 \rfloor, \overline{x})}{\lfloor k \rfloor !}, \tag{7.16}$$

$$F_{\tilde{y}}(\tilde{k}) = \phi_{(\tilde{x},\sigma^2)}(\tilde{k}) = \frac{1}{2}\left(1 + \mathrm{erf}\left(\frac{\tilde{k}-\tilde{x}}{\sigma\sqrt{2}}\right)\right), \tag{7.17}$$

où $\Gamma$ et erf sont respectivement la fonction gamma incomplète inférieure et la fonction d'erreur de Gauss, i.e.,

$$(\forall u \in ]0, +\infty[) \quad \Gamma(u, \overline{x}) = \int_0^{\overline{x}} t^{u-1} \mathrm{e}^{-t} \mathrm{d}t, \tag{7.18}$$

$$(\forall u \in \mathbb{R}) \quad \mathrm{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u \mathrm{e}^{-t^2} \mathrm{d}t, \tag{7.19}$$

alors la transformation s'écrit pour tout $y \in \mathbb{R}_+$ :

$$f_{(\tilde{x},\sigma^2,\overline{x})}(y) = \phi^{-1}_{(\tilde{x},\sigma^2)} \circ P_{(\overline{x})} = \sigma\sqrt{2}\mathrm{erf}^{-1}\left(2\frac{\Gamma(\lfloor y+1 \rfloor, \overline{x})}{\lfloor y \rfloor !} - 1)\right) + \tilde{x}. \tag{7.20}$$

□

Un exemple de la fonction $f_{(\tilde{x},\sigma^2,\overline{x})}$ est illustré par la Figure 7.2 (haut gauche, bleu). C'est une fonction naturellement constante par morceaux qui ne change de valeur que toutes les valeurs entières de $y$.

Cependant, son utilisation dans la pratique requière la connaissance de $\overline{\boldsymbol{x}}$ qui n'est pas connu *a priori*. Il est généralement proposé d'utiliser à la place des techniques de stabilisation de la variance telles que la transformée de Bartlett [21], la transformée de Wilson-Hilferty [215] ou la transformée de Haar-Fisz [94]. Ici nous utilisons la transformée d'Anscombe, qui est un cas particulier de la transformée de Bartlett. Nous montrons ici que cetete transformée peut s'interpreter comme une approximation lisse et strictement monotone de la fonction $f_{(\tilde{x},\sigma^2,\overline{x})}$.

**Proposition 7.2** (Transformée d'Anscombe [11])**.** *La transformée d'Anscombe*

$$\tilde{f} : y \to 2\sqrt{y + 3/8} \tag{7.21}$$

*est une bonne approximation de* $f_{(\tilde{x},\sigma^2,\overline{x})}$ *pour*

$$\begin{cases} \tilde{x} & = 2\sqrt{\overline{x} + 3/8} - 1/(4\sqrt{\overline{x}}), \\ \sigma^2 & = 1. \end{cases} \tag{7.22}$$

*Afin d'inverser la transformation* (7.21)*, nous utiliserons la forme explicite et non biaisée proposée dans [138, 137].*

Une illustration de la transformée d'Anscombe est rapportée Figure 7.2 (haut gauche, rouge). Nous remarquons qu'elle constitue effectivement une bonne approximation de la transformée exacte pour de faibles valeurs de $\boldsymbol{y}_\Delta$. En effet, (7.21) peut être obtenu par un développement limité de $f_{(\tilde{x},\sigma^2,\overline{x})}$ en 0 pour les paramètres énoncés dans (7.22). La qualité des distributions gaussiennes après transformation sont comparées Figure 7.2 (haut droite) et apparaissent toute deux satisfaisantes.

## 7.3 Méthodologies

Dans cette section, nous présentons plusieurs procédures de détection de discontinuités dont une méthode de seuillage et différentes approches variationnelles reposant sur l'utilisation de la variation totale.

### 7.3.1 Méthode du seuil

Les méthodes de seuillage sont très largement utilisées dans la pratique pour discriminer les états OFF des états ON du fait de leur simple mise en oeuvre. Le seuil est généralement choisi qualitativement à partir de l'allure de $\boldsymbol{y}_\Delta$ [56]. Ici nous proposons une méthode légèrement améliorée permettant la selection automatique du seuil.

**Estimée par seuillage $\widehat{\widehat{\boldsymbol{x}}}_\Delta^{(\mathrm{Thr.})}$ :**

1. Un histogramme normalisé de $\boldsymbol{y}_\Delta$, pour 100 valeurs équitablement espacées entre $\min(\boldsymbol{y}_\Delta)$ et $\max(\boldsymbol{y}_\Delta)$, est calculé.
2. L'histogramme est ensuite ajusté par un mélange de deux distributions de Poisson, i.e.,

$$\alpha\mathcal{P}_{(\mu_1)} + (1-\alpha)\mathcal{P}_{(\mu 2)} \tag{7.23}$$

   où $\alpha \in [0,1]$ et $\mu_1 \leq \mu_2$. Les valeurs optimales des paramètres, notées $(\widehat{\alpha}, \widehat{\mu}_1, \widehat{\mu}_2)$, sont obtenues par la méthode de Nelder-Mead [126].
3. Le seuil $\zeta$ est choisi de manière à minimiser le recouvrement entre les distributions $\mathcal{P}_{(\widehat{\mu}_1)}$ et $\mathcal{P}_{(\widehat{\mu}_2)}$, i.e.,

$$\widehat{\zeta} = \arg \min_{\widehat{\mu}_1 \leq \zeta \leq \widehat{\mu}_2} F_{(\widehat{\mu}_2)}(\zeta) - F_{(\widehat{\mu}_1)}(\zeta) \tag{7.24}$$

   où $F_{(\mu)}$ désigne la fonction cumulative de $\mathcal{P}_{(\mu)}$.
4. Finalement, nous définissons l'estimée seuillée suivante

$$(\forall n \in \{1,\ldots,N\}) \quad \widehat{\widehat{x}}_{\Delta,n}^{(\mathrm{Thr.})} = \begin{cases} \widehat{\mu}_1 & \text{, si } y_{\Delta,n} \leq \widehat{\zeta}, \\ \widehat{\mu}_2 & \text{, sinon.} \end{cases} \tag{7.25}$$

Un exemple d'histogramme de $\boldsymbol{y}_\Delta$ calculé à l'étape 1 est illustré par la Figure 7.3. Le meilleur ajustement issu de l'étape 2 est affiché en traits pleins noirs alors que le seuil sélectionné à l'étape 3 est indiqué en traits pointillés.

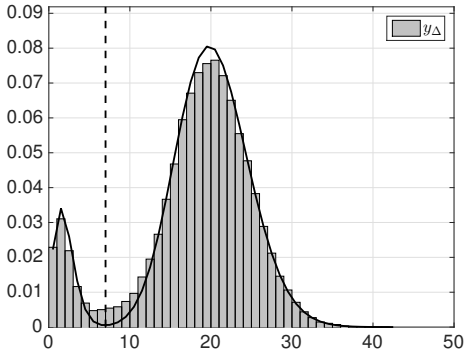



FIGURE 7.3 – **Histogramme de $\boldsymbol{y}_\Delta$.** Meilleur ajustement (traits pleins noirs) et seuil (traits pointillés noirs)

### 7.3.2 Méthodes d'estimation par approche TV

#### 7.3.2.1 Estimées par approche TV

Nous proposons d'estimer les discontinuités de $\overline{\boldsymbol{x}}_\Delta$ en résolvant un problème variationnel comprenant un terme d'attache aux données et un terme de variation totale afin d'imposer le comportement constant par morceaux de la solution. Nous considérons seulement le cas où la variation totale est pénalisée en norme $\ell_1$. Dans la suite, nous envisageons les différentes stratégies listées ci-dessous :

**$\ell_2$-TV.** La première approche, la plus naïve, consiste à résoudre le Problème 2.2 appliqué à $\boldsymbol{y}_\Delta$, i.e.,

$$\widehat{\boldsymbol{x}}_{\Delta,\lambda}^{(\ell_2-\mathrm{TV})} = \arg\min_{\boldsymbol{x}_\Delta \in \mathbb{R}^N} \|\boldsymbol{y}_\Delta - \boldsymbol{x}_\Delta\|^2 + \lambda \mathrm{TV}(\boldsymbol{x}_\Delta). \tag{7.26}$$

où $\mathrm{TV}(\boldsymbol{x}_\Delta) = \|L\boldsymbol{x}_\Delta\|_1$ avec l'opérateur de premières différences $L$ défini en (2.3). Nous indicerons la solution par $(\ell_2 - \mathrm{TV})$ pour rappeler le terme d'attache aux données $(\ell_2)$ et le terme de régularisation (TV). Cette approche possède un coût de calcul très faible grâce à l'algorithme présenté dans [52], mais le terme d'attache aux données n'est pas cohérent avec l'*a priori* sur $\boldsymbol{y}_\Delta$.

**DKL-TV.** Une seconde solution consiste à choisir la divergence de Kullback-Leibler comme terme d'attache aux données. Comme discuté à la Section 7.2.1, c'est le terme d'attache aux données adapté pour les processus de Poisson. Nous proposons alors de résoudre le problème de minimisation convexe suivant

$$\widehat{\boldsymbol{x}}_{\Delta,\lambda}^{(\mathrm{DKL}-\mathrm{TV})} = \arg\min_{\boldsymbol{x}_\Delta \in \mathbb{R}^N} \mathrm{DKL}(\boldsymbol{x}_\Delta, \boldsymbol{y}_\Delta) + \lambda \mathrm{TV}(\boldsymbol{x}_\Delta). \tag{7.27}$$

Contrairement à (7.26), le terme d'attache aux données dans (7.27), i.e., la divergence de Kullback-Leibler, n'est pas pas Lipschitz-différentiable, au même titre que la variation totale. Par conséquent, nous ne pouvons pas utiliser une variante de l'Algorithme 1 forward-backward qui nécessite qu'au moins une des deux fonctions à minimiser soit différentiable. Toutefois, (7.27) peut être résolu efficacement grâce à un algorithme primal-dual [149] reposant sur l'utilisation de l'opérateur proximal suivant :

**Exemple 7.1** (Opérateur proximal de la divergence de Kullback-Leibler)**.** *Soient $\gamma \in ]0, +\infty[$, un vecteur $\boldsymbol{y} \in \mathbb{R}^N$ et $\mathrm{DKL}(\cdot, \boldsymbol{y})$ la divergence de Kullback-Leibler. Alors, l'opérateur proximal*

*de la fonction* $\mathbb{R}^N \to \mathbb{R} : \boldsymbol{u} \mapsto \gamma \mathrm{DKL}(\boldsymbol{u}, \boldsymbol{y})$ *s'écrit*

$$\mathrm{prox}_{\gamma \mathrm{DKL}(\cdot, \boldsymbol{y})}(\boldsymbol{u}) = \frac{1}{2}\left(\boldsymbol{u} - \gamma + \sqrt{(\boldsymbol{u}-\gamma)^2 + 4\gamma \boldsymbol{y}}\right). \tag{7.28}$$

Cette approche a l'avantage d'être cohérente avec les hypothèses du modèle mais présente un coût de calcul plus élevé dû à la nature itérative de l'algorithme pour résoudre (7.27).

**Ans.$\ell_2$-TV.** Une troisième solution nécessite un prétraitement à l'aide de la transformée d'Anscombe pour transformer les données $\boldsymbol{y}_\Delta$, dont la distribution est un mélange de lois de Poisson, en $\tilde{\boldsymbol{y}}_\Delta$ dont la distribution est quasiment un mélange de Gaussiennes de variances 1 (voir Section 7.2.2). Dans ce cas, le problème est correctement modélisé si le terme d'attache aux données est la norme $\ell_2$, i.e.,

$$\tilde{\boldsymbol{x}}_{\Delta,\lambda}^{(\mathrm{Ans.}+\ell_2-\mathrm{TV})} = \arg\min_{\boldsymbol{x}_\Delta \in \mathbb{R}^N} \|\tilde{\boldsymbol{y}}_\Delta - \boldsymbol{x}_\Delta\|^2 + \lambda \mathrm{TV}(\boldsymbol{x}_\Delta) \tag{7.29}$$

Finalement, il ne reste plus qu'à appliquer la transformée d'Anscombe inverse à $\tilde{\boldsymbol{x}}_{\Delta,\lambda}^{(\mathrm{Ans.}+\ell_2-\mathrm{TV})}$ pour obtenir $\widehat{\boldsymbol{x}}_{\Delta,\lambda}^{(\mathrm{Ans.}+\ell_2-\mathrm{TV})}$. Cette approche a le double avantage d'être cohérente avec les hypothèses et de présenter un coût de calcul très faible. Cependant, la qualité de la distribution Gaussienne après la transformée d'Anscombe est discutable.

Les différentes solutions $\widehat{\boldsymbol{x}}_{\Delta,\lambda}$ obtenues sont toutes constantes mais biaisées due à l'utilisation de la norme $\ell_1$. Il convient alors de réestimer *a posteriori* les valeurs moyennes de $\widehat{\boldsymbol{x}}_{\Delta,\lambda}$ sur chaque segment à partir de $\boldsymbol{y}_\Delta$.

De plus, dans le présent contexte, nous souhaiterions estimer les deux seuls états ON et OFF. Par conséquent, il est nécessaire de rajouter une étape supplémentaire de seuillage décrite dans la section suivante.

#### 7.3.2.2 Réestimation à 2 niveaux

Pour chacune des solutions mentionnées précédemment, une estimée à 2 niveaux $\widehat{\widehat{\boldsymbol{x}}}_{\Delta,\lambda}$ peut être obtenue à partir de $\widehat{\boldsymbol{x}}_\Delta$ comme suit :

1. Un histogramme normalisé de $\widehat{\boldsymbol{x}}_{\Delta,\lambda}$, avec 100 valeurs équitablement espacées entre $\min(\widehat{\boldsymbol{x}}_{\Delta,\lambda})$ et $\max(\widehat{\boldsymbol{x}}_{\Delta,\lambda})$, est calculé.
2. L'histogramme est ensuite ajusté par un mélange de deux lois normales, i.e.,
$$\alpha \mathcal{N}_{(\mu_1,\sigma_1^2)} + (1-\alpha)\mathcal{N}_{(\mu_2,\sigma_2^2)}, \tag{7.30}$$
où $\alpha \in [0,1]$ et $\mu_1 \leq \mu_2$. Les valeurs optimales de l'ajustement sont notées $(\widehat{\alpha}, \widehat{\mu}_1, \widehat{\sigma}_1^2, \widehat{\mu}_2, \widehat{\sigma}_2^2)$.
3. Un seuil $\widehat{\zeta}$ est choisi afin de minimiser le recouvrement entre $\mathcal{N}_{(\widehat{\mu}_1,\widehat{\sigma}_1^2)}$ et $\mathcal{N}_{(\widehat{\mu}_2,\widehat{\sigma}_2^2)}$. En d'autre termes,
$$\widehat{\zeta} = \arg\min_{\widehat{\mu}_1 \leq \zeta \leq \widehat{\mu}_2} F_{(\widehat{\mu}_2,\widehat{\sigma}_2^2)}(\zeta) - F_{(\widehat{\mu}_1,\widehat{\sigma}_1^2)}(\zeta) \tag{7.31}$$
où $F_{(\mu,\sigma^2)}$ désigne la fonction cumulative de $\mathcal{N}_{(\mu,\sigma^2)}$.
4. Finalement, nous définissons l'estimée à 2 niveaux suivante
$$(\forall n \in \{1,\ldots,N\}) \quad \widehat{\widehat{x}}_{\Delta,\lambda,n} = \begin{cases} \widehat{\mu}_1 & \text{, si } \widehat{x}_{\Delta,\lambda,n} \leq \widehat{\zeta}, \\ \widehat{\mu}_2 & \text{, sinon.} \end{cases} \tag{7.32}$$

Pour la même réalisation rapportée dans la Figure 7.3, les histogrammes de $\widehat{\boldsymbol{x}}_{\Delta,\lambda}$ des trois solutions présentées précédemment sont illustrés par la Figure 7.4 pour $\lambda = 0.5$. Dans chacun des cas, le meilleur ajustement obtenu à l'étape 2 est indiqué en traits pleins noirs et le seuil estimé à l'étape 3 est affiché en traits pointillés.

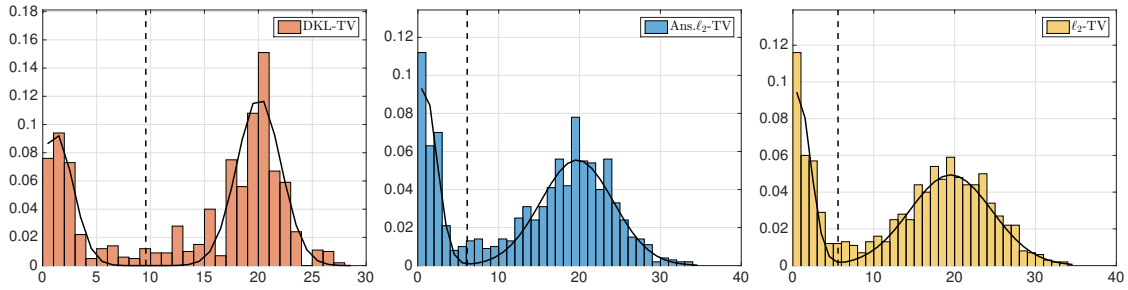



FIGURE 7.4 – **Choix du seuil pour les estimées par approche TV.** Pour chacune des trois méthodes, sont représentés l'histogramme de $\widehat{x}_{\Delta,\lambda}$ pour $\lambda = 0.5$, le meilleur ajustement (traits pleins noir) et le seuil optimal (pointillés noirs).

### 7.3.3 Méthode de segmentation par approche TV

Nous proposons à présent de directement segmenter $\boldsymbol{y}_\Delta$ pour obtenir 2 niveaux, de valeurs moyennes $v_1$ et $v_2$ *a priori* inconnues.

**Estimée à 2 niveaux $\widehat{\widehat{\boldsymbol{x}}}^{(\mathrm{MS})}_{\Delta,\lambda}$ :**

Des estimées initiales de $v_1$ et $v_2$ sont obtenues en ajustant l'histogramme normalisé de $\boldsymbol{y}_\Delta$ par un mélange de distributions de Poisson, i.e., $\alpha\mathcal{P}_{(\widehat{\mu}_1)} + (1-\alpha)\mathcal{P}_{(\widehat{\mu}_2)}$. Cette procédure nous permet d'initialiser $v_1 = \widehat{\mu}_1$ et $v_2 = \widehat{\mu}_2$. L'estimée à 2 niveaux $\widehat{\boldsymbol{x}}^{(\mathrm{MS})}_{\Delta,\lambda}$ est ensuite obtenue après la répétition des étapes suivantes jusqu'à convergence.

1. Résoudre l'homologue 1D du Problème 2.9 afin de segmenter $\boldsymbol{y}_\Delta$ en au plus $Q = 2$ niveaux de valeurs moyennes $v_1$ et $v_2$ :

$$\min_{(\theta_1,\theta_2,\theta_3)\in[0,1]^{3\times N}} \sum_{q=1}^{Q=2} (\theta_q - \theta_{q+1})^\top (\boldsymbol{y}_\Delta - v_q)^2 + \lambda \sum_{q=1}^{Q=2} \mathrm{TV}(\theta_q - \theta_{q+1})$$
$$\text{assujetti à} \quad \begin{cases} \theta_1 = 1, \\ \theta_3 = 0, \\ 1 \geq \theta_2 \geq 0. \end{cases} \tag{7.33}$$

   Les valeurs optimales sont notées $(\widehat{\theta}_1, \widehat{\theta}_2, \widehat{\theta}_3)$ et jouent le rôle de fonctions indicatrices (voir Proposition 2.6) permettant de définir sans ambiguité l'estimée à 2 niveaux $\widehat{\widehat{x}}^{(\mathrm{MS})}_{\Delta,\lambda}$ comme suit :

$$(\forall n \in \{1,\dots,N\}) \quad \widehat{\widehat{x}}^{(\mathrm{MS})}_{\Delta,\lambda,n} = \begin{cases} v_1 \text{ , si} & \widehat{\theta}_{2,n} - \widehat{\theta}_{1,n} = 1, \quad \text{(état OFF)} \\ v_2 \text{ , si} & \widehat{\theta}_{3,n} - \widehat{\theta}_{2,n} = 1. \quad \text{(état ON)} \end{cases} \tag{7.34}$$

2. Réestimer $v_1$ (resp. $v_2$) comme la moyenne de $\boldsymbol{y}_\Delta$ durant les états OFF (resp. ON) de $\widehat{\widehat{\boldsymbol{x}}}^{(\mathrm{MS})}_{\Delta,\lambda}$ .

## 7.4 Comparaison des différentes méthodes

### 7.4.1 Configuration expérimentale

Nous proposons d'évaluer les performances des différentes stratégies sur plusieurs signaux synthétiques reproduisant ceux observés expérimentalement lors de l'enregistrement

**Algorithme 9** Pseudo code pour générer $\overline{\boldsymbol{x}}_\Delta$ et $\boldsymbol{y}_\Delta$.

**Entrée :** Paramètres $\theta$, $m_{\mathrm{ON}}$, $m_{\mathrm{OFF}}$, $\Delta$, $\overline{\mu}_{\mathrm{ON}}$, $\overline{\mu}_{\mathrm{OFF}}$ et $N$.

Fixer la liste $\tau = (\tau_{\mathrm{ON}}(j), \tau_{\mathrm{OFF}}(j))_{1 \leq j \leq \lceil \Delta N/\theta \rceil}$ des durées intercalées des états ON et OFF où $\tau_{\mathrm{ON}} \sim \mathrm{Pareto}_{(\theta, m_{\mathrm{ON}}-1)}$ et $\tau_{\mathrm{OFF}} \sim \mathrm{Pareto}_{(\theta, m_{\mathrm{OFF}}-1)}$.
Fixer $c = \mathrm{cumsum}(\tau)$
Fixer $n = 1$
**Tant que** $n \leq N$ **faire**
    Trouver le premier indice $i$ tel que $c(i) \geq \Delta$.
    Fixer $t = \{c(1), \ldots, c(i-1), \Delta\}$. Fixer $r = c(i) - \Delta$.
    Calculer $\alpha_{\mathrm{ON}}$ (resp. $\alpha_{\mathrm{OFF}}$) la proportion des temps ON (resp. OFF) dans $t$.
    Fixer $\overline{x}_{\Delta,n} = \Delta(\alpha_{\mathrm{ON}}\overline{\mu}_{\mathrm{ON}} + \alpha_{\mathrm{OFF}}\overline{\mu}_{\mathrm{OFF}})$.
    Fixer $c = (r, c(i+1), \ldots c(\mathrm{end}))$.
    Fixer $n = n + 1$.

**Sortie :** $\overline{\boldsymbol{x}}_\Delta$ et $\boldsymbol{y}_\Delta \sim \mathcal{P}(\overline{\boldsymbol{x}}_\Delta)$.

du nombre de photons émis par une boîte quantique (cf., e.g., [125, 108] pour une revue détaillée).

**Fréquence d'émission et taille du signal.** Afin de fixer les ordres de grandeur, nous rappelons les fréquences d'émission de photons typiques des boîtes quantiques industrielles pendant l'état le plus émissif, i.e., l'état ON :

— $f_e = 5$kHz (faible qualité)
— $f_e = 50$kHz (qualité standard)
— $f_e = 200$kHz (meilleure qualité)

Nous choisissons de nous placer dans le cas intermédiaire où $f_e = 50$kHz et considérons que

$$\begin{cases} \overline{\mu}_{\mathrm{ON}} & = f_e \Delta, \\ \overline{\mu}_{\mathrm{OFF}} & = \overline{\mu}_{\mathrm{ON}}/15. \end{cases} \tag{7.35}$$

De plus, si nous considérons que l'enregistrement dure approximatement 10 min et que les photons sont comptés dans des fenêtres de l'ordre la milliseconde, alors nous obtenons des observations de taille $N \sim 10^5$.

**Distribution des temps ON et OFF.** Pour un temps de coupure $\theta = 35\mu$s, nous considérerons que $m_{\mathrm{ON}} = 1.8$ et $m_{\mathrm{OFF}} = 1.95$. Ce sont des valeurs assez grandes utilisées précédemment dans [108]. Pour de telles valeurs, les distributions des temps ON et des temps OFF sont à queue lourde.

**Rapport entre $\theta$ et $\Delta$.** Selon la qualité des capteurs utilisés pour détecter les photons, nous pouvons espérer différents rapports $\theta/\Delta$ :

— $\theta/\Delta \sim 1/100$ pour des caméras standard ($\sim 200$ photons par fenêtre).
— $\theta/\Delta \sim 1/10$ ($\sim 20$ photons par fenêtre).
— $\theta/\Delta \sim 1$ pour les expériences standard ($\sim 2$ photons par fenêtre).
— $\theta/\Delta \sim 10$ ($\sim 0.2$ photons par fenêtre).
— $\theta/\Delta \sim 100$ dans la limite des composants électroniques ($\sim 0.02$ photons par fenêtre).

Nous choisissons de nous placer pour des rapports $\theta/\Delta \leq 1$ où les approches variationnelles basées sur la TV sont adaptées. En particulier, nous étudierons les cas $\theta/\Delta = 0.1$ et 1.

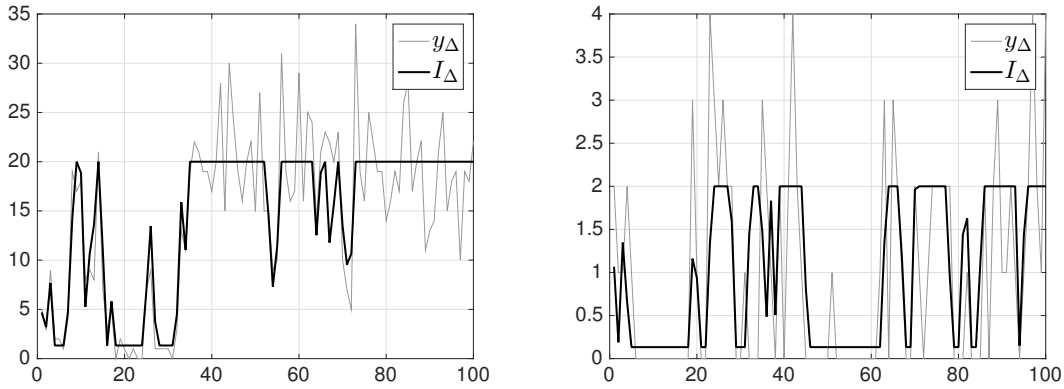



FIGURE 7.5 – **Exemples de réalisation.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$.

**Données synthétiques.** Les données sont générées à l'aide de la procédure décrite par l'Algorithme 9. Deux exemples sont illustrés par la Figure 7.5.

### 7.4.2 Quantification des performances

**Erreur quadratique moyenne relative.** Pour chacune des méthodes, nous comparons l'erreur quadratique moyenne relative, i.e.,

$$\mathrm{rMSE}(\widehat{\widehat{\boldsymbol{x}}}_\Delta) = \|\widehat{\widehat{\boldsymbol{x}}}_\Delta - \overline{\boldsymbol{x}}_\Delta\|^2 / \|\overline{\boldsymbol{x}}_\Delta\|^2 \tag{7.36}$$

en fonction du paramètre de régularisation $\lambda$. Les résultats sont moyennés sur 50 réalisations et rapportés par la Figure 7.6. Globalement, les approches TV offrent une rMSE plus faible ou égale à celle de la méthode du seuil. Notons également que la méthode segmentation (MS) permet d'aboutir systématiquement à une rMSE plus faible que les autres stratégies. Cependant, la rMSE fournit essentiellement une mesure de la qualité d'estimation de $\overline{\mu}_{\mathrm{ON}}$ et $\overline{\mu}_{\mathrm{OFF}}$, et ne nous donne pas d'information sur la qualité de localisation des discontinuités.

Idéalement, nous souhaiterions que $\widehat{\widehat{\boldsymbol{x}}}_\Delta$ soit discontinu dès lors que $\overline{\boldsymbol{x}}_\Delta$ ne vaut ni $\overline{\mu}_{\mathrm{ON}}$ ni $\overline{\mu}_{\mathrm{OFF}}$, afin de ne pas fausser l'estimation de la distribution des temps ON et OFF de $\overline{\boldsymbol{x}}$. Nous proposons de quantifier la similarité entre les discontinuités de $\widehat{\widehat{\boldsymbol{x}}}_\Delta$ et $\overline{\boldsymbol{x}}_\Delta$ à partir de l'indice de Jaccard :

**Indice de Jaccard.** Soit $\boldsymbol{r}$ l'indicatrice des discontinuités de $\overline{\boldsymbol{x}}_\Delta$. Pour chaque méthode, nous noterons $\widehat{\boldsymbol{r}}$ l'indicatrice des discontinuités de la solution à 2 niveaux $\widehat{\widehat{\boldsymbol{x}}}$. L'indice de Jaccard (voir Définition 2.7) entre $\mathcal{G} * \boldsymbol{r}$ et $\mathcal{G} * \widehat{\boldsymbol{r}}$ est calculé pour un filtre passe-bas gaussien à support compact $\mathcal{G}$ de taille 3 points et d'écart-type 0.3. Les résultats, rapportés par la Figure 7.7 en fonction du paramètre de régularisation $\lambda$, montrent que toutes les stratégies donnent des indices de Jaccard plus ou moins similaires. Ces résultats sont toutefois difficiles à interpréter du fait de la différence de nature entre $\widehat{\widehat{\boldsymbol{x}}}_\Delta$ et $\overline{\boldsymbol{x}}_\Delta$, ce qui met en avant la limitation de l'indice de Jaccard.

**Temps d'éxécution.** La comparaison des temps d'éxécution est rapportée Figure 7.8. Sans surprise, les stratégies (Ans.$\ell_2$-TV) et ($\ell_2$-TV) sont les plus rapides grâce à l'utilisation de l'algorithme à la volée proposé dans [52], alors que les méthodes (DKL-TV) et (MS) affichent des temps d'éxécution plusieurs ordres de grandeurs plus élevés.

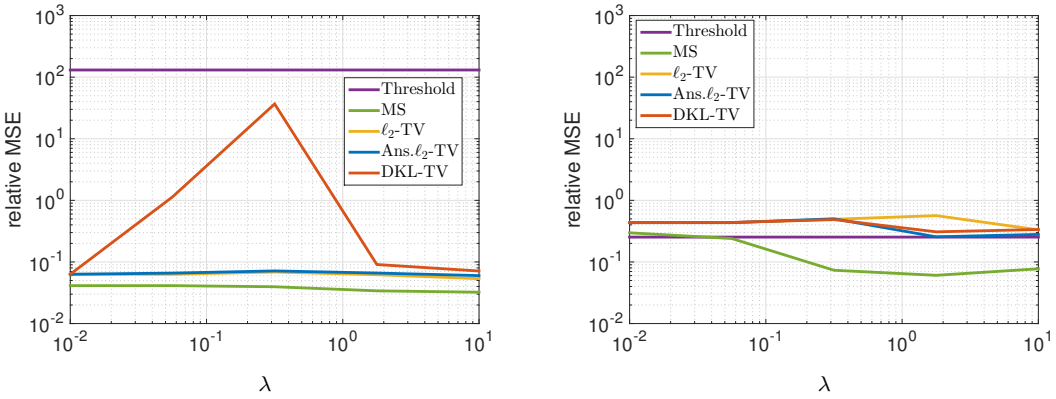



FIGURE 7.6 – **Erreur quadratique moyenne relative.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$

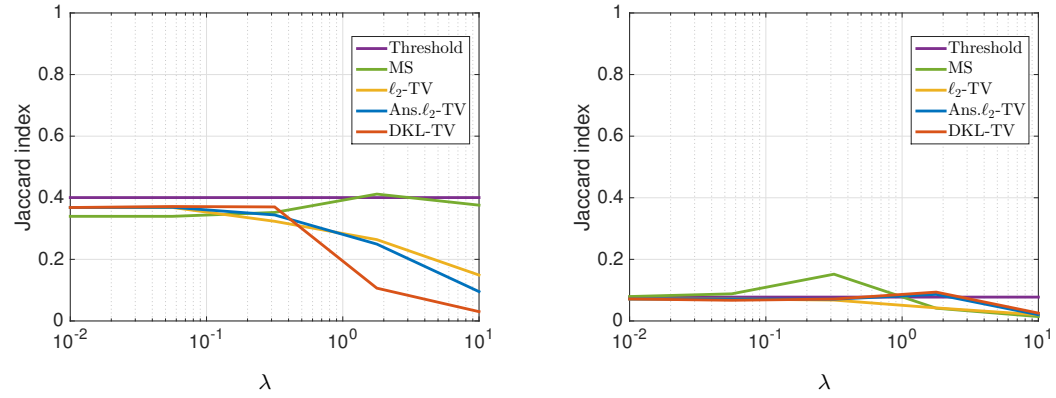



FIGURE 7.7 – **Indice de Jaccard.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$

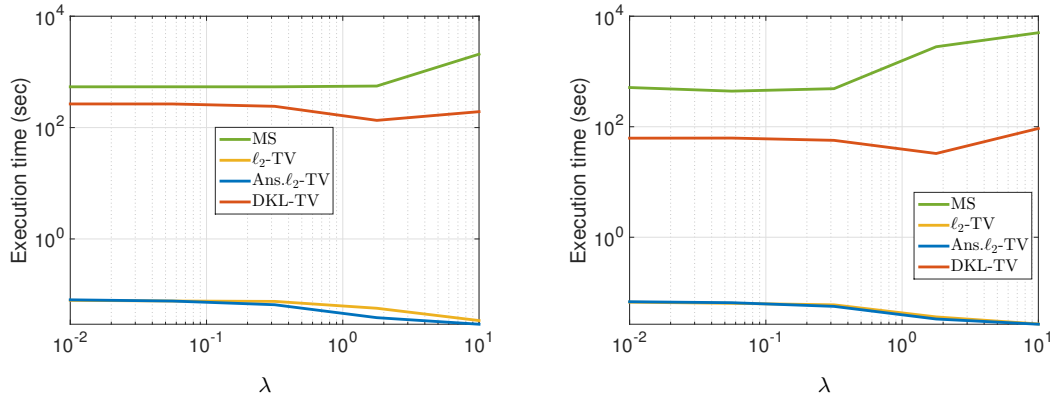



FIGURE 7.8 – **Temps d'éxécution.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$

**Distribution empirique des durées ON et OFF.** Pour chacune des méthodes proposées, les distributions empiriques des temps ON et OFF de la solution $\widehat{\widehat{\boldsymbol{x}}}_{\lambda_{J_{\max}}}$ qui maximise l'indice de Jaccard, sont respectivement rapportées Figure 7.9 et Figure 7.10. Les distributions sont mises à l'échelle pour obtenir la même probabilité que la distribution théorique (noir) dans l'intervalle délimité par les pointillés rouges. Nous rappelons que les densités de probabilité théoriques suivent :

$$\begin{cases} \frac{(m_{\mathrm{ON}}-1)\theta^{m_{\mathrm{ON}}-1}}{\tau_{\mathrm{ON}}^{m_{\mathrm{ON}}}} & \text{pour} \quad \tau_{\mathrm{ON}} \geq \theta, \\ \frac{(m_{\mathrm{OFF}}-1)\theta^{m_{\mathrm{OFF}}-1}}{\tau_{\mathrm{OFF}}^{m_{\mathrm{OFF}}}} & \text{pour} \quad \tau_{\mathrm{OFF}} \geq \theta. \end{cases} \tag{7.37}$$

Pour faciliter la lecture, la densité de probabilité empirique de $\tau_{\mathrm{ON}}$ (resp. $\tau_{\mathrm{ON}}$) est multipliée par le facteur $\tau_{\mathrm{ON}}^{m_{\mathrm{ON}}}$ (resp. $\tau_{\mathrm{OFF}}^{m_{\mathrm{OFF}}}$) et est illustrée par la Figure 7.11 (resp. Figure 7.12).

Nous observons que les diverses méthodes proposées permettent d'estimer plus efficacement la distribution dont l'exposant est le plus faible, i.e., $m_{\mathrm{ON}}$. En effet, plus la distribution des temps ON et OFF de $\overline{\boldsymbol{x}}$ est à queue lourde, plus il est facile d'estimer leur distribution à partir de $\boldsymbol{x}_\Delta$. Il est intéressant de remarquer que, pour la présente configuration avec $\theta/\Delta = 0.1$ (gauche), alors la méthode du seuil donne globalement des résultats similaires aux méthodes d'estimation TV. Ces dernières démontrent un intérêt justifié par exemple lorsque $\theta/\Delta = 1$ (droite). Globalement, les solutions (Ans.$\ell_2$-TV) et (MS) offrent les meilleures performances d'estimation, alors que la solution naive ($\ell_2$-TV) et la solution (DKL-TV) ne parviennent pas à estimer correctement $m_{\mathrm{ON}}$ et $m_{\mathrm{OFF}}$.

## 7.5 Conclusion et perspectives

Dans ce chapitre, nous avons examiné l'apport des approches reposant sur la variation totale pour détecter les discontinuités d'intensité de processus de Poisson agrégés. Les performances ont été évaluées sur des données synthétiques reproduisant les signaux observés en pratique lors d'expériences de scintillement fluorescent de boîtes quantiques. Les résultats préliminaires obtenus sont suffisamment concluants pour appliquer les différentes stratégies sur des données réelles. Cette piste est actuellement envisagée et donnera lieu à une collaboration avec l'équipe *Luminescence* de l'Institut Lumière Matière de l'Université Claude Bernard.

Il serait également intéressant d'examiner la qualité des solutions obtenues lorsque la variation totale est pénalisée en norme $\ell_0$, ce qui nous permettrait d'éviter certains problèmes liés à l'utilisation de la norme $\ell_1$ (biais et effet d'escalier). Pour cela, nous projetons d'utiliser l'algorithme de programmation dynamique [193, 192] pour un terme d'attache aux données $\ell_2$, et l'algorithme récemment proposé dans [75] lorsque le terme d'attache aux données est la divergence de Kullback-Leibler.

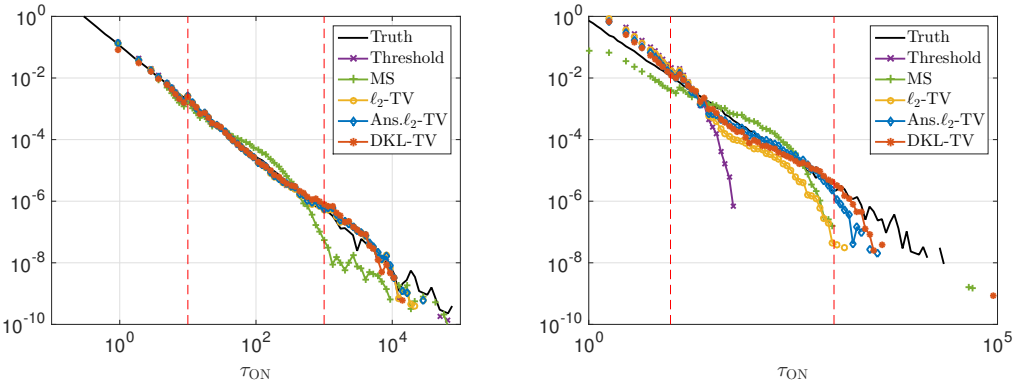



FIGURE 7.9 – **Densité de probabilité empirique de $\tau_{\mathrm{ON}}$.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$.

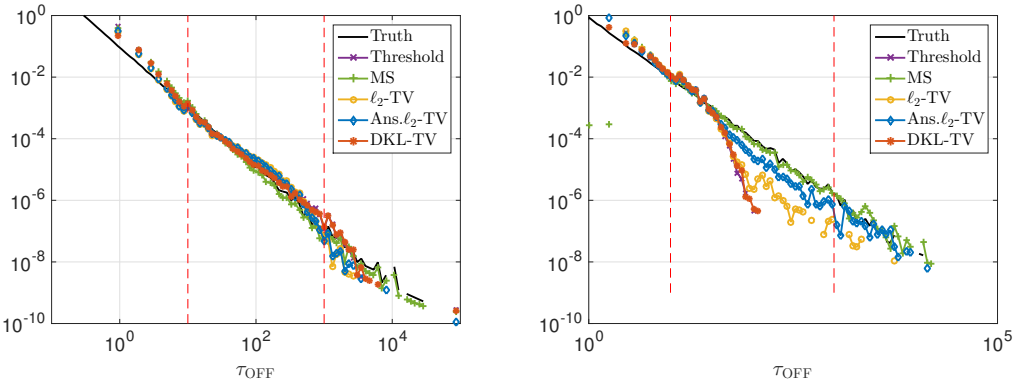



FIGURE 7.10 – **Densité de probabilité empirique de $\tau_{\mathrm{OFF}}$.** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$.

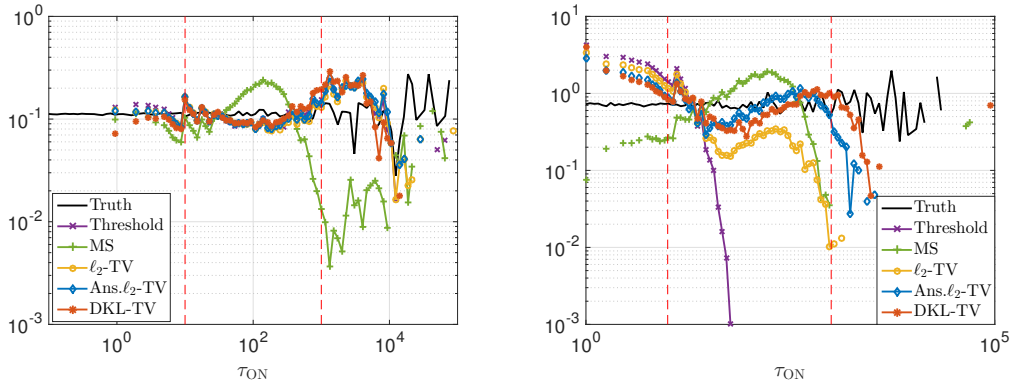



FIGURE 7.11 – **(Densité de probabilité empirique de $\tau_{\mathrm{ON}}$) $\times \tau_{\mathrm{ON}}^{m_{\mathrm{ON}}}$ .** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$.

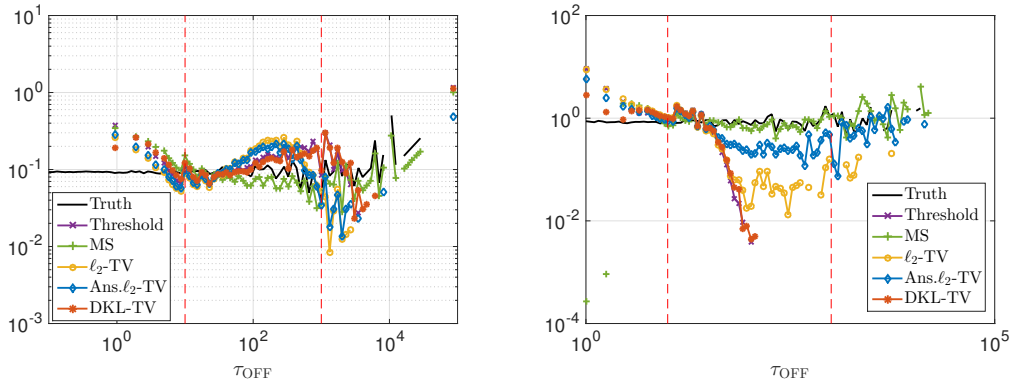



FIGURE 7.12 – **(Densité de probabilité empirique de $\tau_{\mathrm{OFF}}$) $\times \tau_{\mathrm{OFF}}^{m_{\mathrm{OFF}}}$ .** Gauche : $\theta/\Delta = 0.1$. Droite : $\theta/\Delta = 1$.

# Chapitre 8

# Estimation du paramètre de régularisation $\ell_0$-TV

**Sommaire**


**8.1 Paramétrisation du problème . . . . . 150**
8.1.1 Paramétrisation des discontinuités $\boldsymbol{r}$ . . . . . 150
8.1.2 Paramétrisation des valeurs des segments $\boldsymbol{\mu}$ . . . . . 150
8.1.3 Reformulation du problème . . . . . 151
**8.2 Dérivation Bayésienne de la pénalisation $\phi$ . . . . . 151**
8.2.1 Modèle hiérarchique Bayésien . . . . . 151
8.2.2 Critère MAP joint . . . . . 152
**8.3 Solution algorithmique . . . . . 154**
**8.4 Sélection automatique de $\chi$ : Illustration et performances . . . . 155**
8.4.1 Configuration expérimentale et choix des hyperparamètres . . . . . 155
8.4.2 Illustration du principe de sélection automatique de $\chi$. . . . . 155
8.4.3 Quantification des performances . . . . . 157
8.4.4 Comparaison contre de l'inférence bayésienne . . . . . 158
8.4.5 Robustesse de l'estimateur proposé . . . . . 159
**8.5 Conclusion et perspectives . . . . . 160**


**Travaux préliminaires : estimation du paramètre de régularisation.** Les approches TV utilisées tout au long du manuscrit se sont montrées efficaces en raison de leur formalisation reposant sur la minimisation d'une fonctionnelle déterministe et par le développement d'algorithmes de résolution efficaces. Cependant, les performances d'estimation de la solution dépendent fortement du choix du paramètre de régularisation. La question de son estimation reçoit encore de nombreux efforts méthodologiques et théoriques. Il peut par exemple être appris et validé sur une base de donnée de référence par validation croisée ou choisi selon divers critères heuristiques (voir [207] pour une revue détaillée). En outre, lorsque la variance du bruit est connue *a priori*, l'estimateur non biaisé du risque de Stein (SURE) [191, 59], permet de sélectionner $\lambda$ dans le Problème 2.2 afin de minimiser l'erreur quadratique entre $\overline{\boldsymbol{x}}$ et $\widehat{\boldsymbol{x}}_\lambda$. Une interprétation géométrique du $\lambda$ optimal a également été envisagée dans [196] où la méthode proposée vise à utiliser les solutions $(\widehat{\boldsymbol{x}}_\lambda)_{\lambda\geq 0}$ du Problème 2.2 afin de sonder les échelles [1] caractéristiques des régions dans l'observation $\boldsymbol{y}$. Puis le paramètre de régularisation

1. La notion d'échelle est ici à comprendre au sens du rapport entre l'aire et le périmètre.

préservant les régions dont l'échelle dépasse un seuil fixé par l'utilisateur est sélectionné. Toutefois, dans le cas général, les questions de l'interprétation physique et de l'estimation non supervisée du paramètre de régularisation demeurent ouvertes.

**Modèles hiérarchiques Bayésiens.** Les modèles Bayésiens hiérarchiques se prêtent couramment aux problèmes de détection de discontinuités et de débruitage constant par morceaux du fait de leur capacité à englober la question du réglage des hyperparamètres [130, 69]. Cependant, les algorithmes de Monte Carlo par chaînes de Markov (MCMC), généralement utilisés pour fournir des estimées, peuvent s'avérer très coûteux en temps de calcul.

Récemment, différents travaux se sont intéressés au développement d'approches bayésiennes pour estimer le paramètre $\lambda$ du Problème 2.2, reformulé comme un problème d'inférence statistique, dont l'estimateur du maximum *a posteriori* est obtenu en résolvant

$$\max_{\boldsymbol{x}\in\mathbb{R}^N} \left\{ \left(\frac{1}{2\pi\sigma^2}\right)^{N/2} e^{-\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2} \frac{1}{Z(\lambda/\sigma^2)} e^{-\lambda\|L\boldsymbol{x}\|_1} \right\} \tag{8.1}$$

où $Z(\lambda/\sigma^2)$ est une constante de renormalisation. En adoptant un modèle hierarchique bayésien, les hyperparamètres $\lambda$ et $\sigma^2$ peuvent être inclus et estimés conjointement avec $\boldsymbol{x}$. Cependant, la difficulté provient ici de la non-séparabilité de $\exp\left(-\lambda\|L\boldsymbol{x}\|_1\right)$ en $\boldsymbol{x}$, ce qui conduit à $Z(\lambda/\sigma^2)$ ne pouvant pas être exprimé de manière explicite. Afin de pallier cette difficulté, des astuces empiriques ont été proposées dans [15, 34], alors que [165] manie ce problème à l'aide de calculs approchés intensifs.

**Contribution.** Dans ce chapitre nous nous intéressons à l'estimation du paramètre de régularisation $\chi$ présent dans le Problème 2.1. Ainsi, nous proposons de formuler la question du choix de $\chi$ via la résolution d'un problème variationnel impliquant $\boldsymbol{x}$, $\chi$, et $\sigma^2$ comme suit :

**Problème 8.1** (Problème $\ell_0$-TV pénalisé)**.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\chi \geq 0$. On appelle problème $\ell_0$-TV pénalisé, le problème consistant à trouver*

$$(\widehat{x}, \widehat{\chi}, \widehat{\sigma}^2) \in \underset{\boldsymbol{x}\in\mathbb{R}^N, \chi>0, \sigma^2>0}{\operatorname{Argmin}} \frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \frac{\chi}{\sigma^2}\|L\boldsymbol{x}\|_0 + \phi(\chi, \sigma^2). \tag{8.2}$$

Les principaux défis consistent alors à concevoir la fonction $\phi$ afin qu'elle pénalise « correctement » $\chi$ et $\sigma^2$, et à être en mesure de résoudre (8.2) efficacement. Pour y parvenir, une paramétrisation naturelle de $\boldsymbol{x}$ est présentée dans la Section 8.1 afin de reformuler le Problème 8.1. L'expression explicite de $\phi$ est ensuite dérivée dans la Section 8.2 à partir d'un modèle hierarchique bayésien et une solution algorithmique du Problème 8.1 est décrite dans la Section 8.3. La pertinence de l'estimateur $\widehat{\chi}$ proposé et la qualité de la solution $\widehat{\boldsymbol{x}}$ sont illustrées, quantifiées, et comparées dans la Section 8.4 contre une procédure exclusivement bayésienne.

**Références.** Ce chapitre s'appuie sur les contributions [88, 89] et est le résultat d'une collaboration avec N. Dobigeon.

## 8.1 Paramétrisation du problème

En suivant le raisonnement développé dans [129, 130, 69, 67, 66], tout signal constant par morceaux peut être paramétrisé explicitement par un vecteur d'indicatrices de discontinuité et un vecteur des valeurs prises sur chacun des segments.

### 8.1.1 Paramétrisation des discontinuités $\boldsymbol{r}$

Nous introduisons le vecteur d'indicatrices $\boldsymbol{r} \in \mathbb{R}^{N-1}$ des discontinuités de $\boldsymbol{x} \in \mathbb{R}^N$ (voir Définition 2.4). Dans ce chapitre, nous augmentons la dimension de $\boldsymbol{r}$ en adoptant la convention $r_N = 1$ afin d'assurer que le nombre de segments $K$ soit égal au nombre de valeurs non nulles de $\boldsymbol{r}$, i.e.,

$$K = \sum_{n=1}^{N} r_n = \|\boldsymbol{r}\|_0. \tag{8.3}$$

Dans la suite, nous indexerons $K$ par $\boldsymbol{r}$, i.e., $K_{\boldsymbol{r}}$, afin d'expliciter la dépendance entre $K$ et $\boldsymbol{r}$.

Pour tout $k \in \{1, \ldots, K\}$, nous noterons $\mathcal{R}_k \subset \{1, \ldots, N\}$ l'ensemble des indices associés au $k$-ème segment. Dès lors $\mathcal{R}_k \cap \mathcal{R}_{k'} = \{\emptyset\}$ pour $k \neq k'$ et $\cup_{k=1}^{K} \mathcal{R}_k = \{1, \cdots, N\}$.

### 8.1.2 Paramétrisation des valeurs des segments $\boldsymbol{\mu}$

Le vecteur $\boldsymbol{\mu} = (\mu_k)_{1 \leq k \leq K_{\boldsymbol{r}}}$ est utilisé pour indiquer les valeurs prises sur chacun des segments :

$$(\forall k \in \{1, \ldots, K_{\boldsymbol{r}}\})\ (\forall n \in \mathcal{R}_k) \quad x_n = \mu_k. \tag{8.4}$$

### 8.1.3 Reformulation du problème

Au lieu de formuler le problème en fonction de $\boldsymbol{x}$, nous ferons usage du vecteur paramétrique $\boldsymbol{\theta} = \{\boldsymbol{r}, \boldsymbol{\mu}\}$, afin de caractériser entièrement tout signal constant par morceaux. Notons que cette paramétrisation n'est généralement pas employée car elle possède le défaut de formuler le problème en fonction de $\boldsymbol{\theta} \in \mathbb{R}^N \times \mathbb{R}^{K_{\boldsymbol{r}}}$ dont la dimension est *a priori* inconnue puisqu'elle dépend du nombre de discontinuités $K_{\boldsymbol{r}}$. Cette paramétrisation mène au lemme suivant.

**Lemme 8.1.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et $\boldsymbol{\theta} = \{\boldsymbol{r}, \boldsymbol{\mu}\} \in \mathbb{R}^N \times \mathbb{R}^{K_{\boldsymbol{r}}}$. Alors le Problème 8.1 est équivalent à*

$$(\widehat{\theta}, \widehat{\chi}, \widehat{\sigma}^2) = \underset{\substack{\boldsymbol{\theta} \in \mathbb{R}^N \times \mathbb{R}^{K_{\boldsymbol{r}}} \\ \chi > 0, \sigma^2 > 0}}{\arg\min} \frac{1}{2\sigma^2} \sum_{k=1}^{K_{\boldsymbol{r}}} \sum_{i \in \mathcal{R}_k} (y_i - \mu_k)^2 + \frac{\chi}{\sigma^2}(K_{\boldsymbol{r}} - 1) + \phi(\chi, \sigma^2) \tag{8.5}$$

*où $(\mathcal{R}_k)_{1 \leq k \leq K_{\boldsymbol{r}}}$ est défini à partir de $\boldsymbol{r}$ comme indiqué dans la Section 8.1.1.*

*Démonstration.* En effet, le terme d'attache aux données présent dans le Probleme 8.1 peut se réécrire de manière équivalente

$$\|\boldsymbol{y} - \boldsymbol{x}\|^2 = \sum_{k=1}^{K_{\boldsymbol{r}}} \sum_{i \in \mathcal{R}_k} (y_i - x_i)^2 = \sum_{k=1}^{K_{\boldsymbol{r}}} \sum_{i \in \mathcal{R}_k} (y_i - \mu_k)^2. \tag{8.6}$$

De plus, le terme de pénalisation $\ell_0$-TV peut se réécrire

$$\|L\boldsymbol{x}\|_0 = \|\boldsymbol{r}\|_0 - 1 = K_{\boldsymbol{r}} - 1. \tag{8.7}$$

□

## 8.2 Dérivation Bayésienne de la pénalisation $\phi$

Dans cette section, nous rappelons un modèle hierarchique Bayésien couramment utilisé pour l'estimation de signaux constants par morceaux (cf., e.g. [130, 69]) et nous insistons sur la relation formelle qu'il entretient avec le Problème 8.1 afin de choisir la pénalité $\phi$.

### 8.2.1 Modèle hiérarchique Bayésien

L'inférence Bayésienne consiste à dériver la distribution postérieure du vecteur paramétrique $\boldsymbol{\theta}$ à partir de la fonction de vraisemblance associée au modèle d'observation et à partir des distributions antérieures des paramètres inconnus.

Pour ce faire, nous commençons par rappeler que les composantes du bruit sont supposées être des variables gaussiennes i.i.d. de moyenne nulle et de même variance $\sigma^2$ inconnue. Par conséquent, la fonction de vraisemblance jointe des observations $\boldsymbol{y}$ étant donné le modèle constant par morceaux $\boldsymbol{\theta} = \{\boldsymbol{r}, \boldsymbol{\mu}\}$ et la variance du bruit $\sigma^2$ s'écrit :

$$f\left(\boldsymbol{y}|\boldsymbol{r}, \boldsymbol{\mu}, \sigma^2\right) = \prod_{k=1}^{K_{\boldsymbol{r}}} \prod_{i \in \mathcal{R}_k} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big( - \frac{\left(\mu_k - y_i\right)^2}{2\sigma^2} \Big). \tag{8.8}$$

Afin de dériver la distribution postérieure, des distributions antérieures sont assignées aux paramètres $\boldsymbol{r}$ et $\boldsymbol{\mu}$, supposés *a priori* indépendants.

Premièrement, en suivant les choix admis et utilisés dans [130, 172, 69, 67], les composantes $r_n$ de $\boldsymbol{r}$ sont supposées *a priori* indépendantes et identiquement distribuées (i.i.d.) selon une distribution de Bernoulli d'hyperparamètre $p$, quantifiant la probabilité antérieure d'occurence d'une discontinuité, indépendemment de sa position :

$$\begin{aligned} f(\boldsymbol{r}|p) &= \prod_{i=1}^{N-1} p^{r_i}(1-p)^{1-r_i} \\ &= p^{\sum_{i=1}^{N-1} r_i}(1-p)^{(N-1-\sum_{i=1}^{N-1} r_i)} \\ &= \Big(\frac{p}{1-p}\Big)^{(K_{\boldsymbol{r}}-1)}(1-p)^{(N-1)} \end{aligned} \tag{8.9}$$

Deuxièment, un choix naturel pour $f(\boldsymbol{\mu}|\boldsymbol{r})$ d'un point de vue Bayésien est la distribution conjuguée de (8.8) [151, 102]. Dans le cas présent, cela revient à élire des distributions antérieures Gaussiennes indépendantes $\mathcal{N}\left(\mu_0, \sigma_0^2\right)$ pour chacune des valeurs prises sur les segments $\mu_k$ $(k = 1, \ldots, K_{\boldsymbol{r}})$, i.e.,

$$f(\boldsymbol{\mu}|\boldsymbol{r}) = \prod_{k=1}^{K_{\boldsymbol{r}}} \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{\frac{-(\mu_k - \mu_0)^2}{2\sigma_0^2}}. \tag{8.10}$$

De plus, au sein d'un paradigme hiérarchique Bayésien, les paramètres de nuisance tels la variance du bruit $\sigma^2$ et tout autre hyperparamètre impliqué dans les distributions antérieures, peuvent être inclus au sein du modèle et être estimés conjointement avec $\boldsymbol{\theta}$ [130, 69].

Premièrement, afin de modéliser l'absence de connaissance *a priori* sur la variance du bruit $\sigma^2$, la distribution antérieure non informative de Jeffreys est assignée à $\sigma^2$ :

$$f\left(\sigma^2\right) \propto \frac{1}{\sigma^2}. \tag{8.11}$$

Deuxièmement, la distribution beta $\mathcal{B}(\alpha_0, \alpha_1)$ (i.e., la distribution conjuguée de la loi (8.9)) est assignée à l'hyperparamètre $p$ inconnu :

$$f(p) = \frac{\Gamma\left(\alpha_0 + \alpha_1\right)}{\Gamma\left(\alpha_0\right)\Gamma\left(\alpha_1\right)} p^{\alpha_1 - 1}(1-p)^{\alpha_0 - 1}. \tag{8.12}$$

C'est un choix naturel pour modéliser tout paramètre contraint à l'intervalle $[0, 1]$.

Finalement, les seuls hyperparamètres qui subsistent sont $\alpha_0$, $\alpha_1$, $\mu_0$ et $\sigma_0^2$. Leur choix sera établi dans la partie expérimentale.

### 8.2.2 Critère MAP joint

A partir de la fonction de vraisemblance et des fonctions postérieures introduites pour les paramètres et hyperparamètres, la distribution postérieure jointe s'écrit :

$$f(\boldsymbol{\Theta}|\boldsymbol{y}) \propto f\left(\boldsymbol{y}|\boldsymbol{r},\boldsymbol{\mu},\sigma^2\right) f\left(\boldsymbol{\mu}|\boldsymbol{r}\right) f(\boldsymbol{r}|p) f(p) f\left(\sigma^2\right) \quad \text{où} \quad \boldsymbol{\Theta} = \left\{\boldsymbol{r},\boldsymbol{\mu},\sigma^2,p\right\}. \tag{8.13}$$

Toutefois, dériver les estimateurs Bayésiens, tels l'estimateur des moindre carrés (MMSE) et l'estimateur du maximum *a posteriori* (MAP), associés à la distribution postérieure (8.13) n'est pas simple. Ceci est principalement dû à la nature intrinsèquement combinatoriale du problème, engendrée par la dimension variable de l'espace des paramètres $(\boldsymbol{r},\boldsymbol{\mu}) \in \mathbb{R}^N \times \mathbb{R}^{K_{\boldsymbol{r}}}$.

En particulier, l'estimateur MAP est obtenu en maximisant la distribution postérieure jointe (8.13), ce qui peut également être réécrit comme la minimisation de l'opposé du logarithme de $f(\boldsymbol{\Theta}|\boldsymbol{y})$, i.e.,

**Problème 8.2.** *Soient $\boldsymbol{y} \in \mathbb{R}^N$ et des hyperparamètres $(\alpha_0,\alpha_1,\mu_0,\sigma_0^2)$. L'estimateur MAP des paramètres du modèle hierarchique bayésien de débruitage constant par morceaux est obtenu comme la solution du problème suivant*

$$\begin{aligned}\widehat{\boldsymbol{\Theta}} \in \underset{\boldsymbol{\Theta}=\{\boldsymbol{r},\boldsymbol{\mu},\sigma^2,p\}}{\operatorname{Argmin}} \; & \frac{1}{2\sigma^2}\sum_{k=1}^{K_{\boldsymbol{r}}}\sum_{i\in\mathcal{R}_k}(\mu_k - y_i)^2 + (K_{\boldsymbol{r}}-1)\left(\log\left(\frac{1-p}{p}\right) + \frac{1}{2}\log(2\pi\sigma_0^2)\right) \\ & + \frac{N}{2}\log(2\pi\sigma^2) - (N-1)\log(1-p) + \log\sigma^2 \\ & - (\alpha_1-1)\log p - (\alpha_0-1)\log(1-p) \\ & + \frac{1}{2\sigma_0^2}\sum_{k=1}^{K_{\boldsymbol{r}}}(\mu_k-\mu_0)^2 + \frac{1}{2}\log(2\pi\sigma_0^2).\end{aligned} \tag{8.14}$$

Malgré les différences de paramétrisation entre le Problème 8.1 et le Problème 8.2, nous démontrons ici que les deux sont équivalents pour un choix spécifique de $\chi$ et $\phi$.

**Proposition 8.1.** *Les Problème 8.1 et Problème 8.2 sont équivalents pour le choix de paramétrisation*

$$\chi = \sigma^2\left(\log\left(\frac{1-p}{p}\right) + \frac{1}{2}\log(2\pi\sigma_0^2)\right), \tag{8.15}$$

*et*

$$\begin{aligned}\phi(\chi,\sigma^2) = & \frac{N}{2}\log(2\pi\sigma^2) + \log(\sigma^2) - \frac{\chi}{\sigma^2}(N+\alpha_0-2) + \frac{N+\alpha_0-1}{2}\log(2\pi\sigma_0^2) \\ & + (N+\alpha_0+\alpha_1-3)\log\left(1+\exp\left(\frac{\chi}{\sigma^2} - \frac{1}{2}\log(2\pi\sigma_0^2)\right)\right),\end{aligned} \tag{8.16}$$

*sous condition que*

$$\sum_{k=1}^{K_{\boldsymbol{r}}}(\mu_k-\mu_0)^2 \ll \sigma_0^2\log(2\pi\sigma_0^2). \tag{8.17}$$

*Démonstration.* Le principe de la démonstration consiste à être en mesure d'identifier trois termes dans (8.14) : le terme d'attache aux données (8.6), un terme proportionnel à la pénalisation $\ell_0$-TV (8.7) et un troisième terme indépendant du vecteur d'indicatrices des discontinuités $\boldsymbol{r}$. Nous remarquons que c'est pratiquement déjà le cas à condition de pouvoir

négliger le terme $\frac{1}{2\sigma_0^2}\sum_{k=1}^{K_{\boldsymbol{r}}}(\mu_k-\mu_0)^2$ qui dépend explicitement de $\boldsymbol{r}$ via $K_{\boldsymbol{r}}$. Ainsi, si nous choisissons $\sigma_0^2$ suffisamment grand pour que

$$\frac{1}{2\sigma_0^2}\sum_{k=1}^{K_{\boldsymbol{r}}}(\mu_k-\mu_0)^2 \ll \frac{1}{2}\log(2\pi\sigma_0^2), \tag{8.18}$$

alors nous pouvons reconnaître

$$\begin{cases} \chi & = \sigma^2(\log\left(\frac{1-p}{p}\right)+\frac{1}{2}\log(2\pi\sigma_0^2)), \\ \phi(p,\sigma^2) & = \frac{N}{2}\log(2\pi\sigma^2)-(N-1)\log(1-p)+\log\sigma^2 \\ & \quad -(\alpha_1-1)\log p-(\alpha_0-1)\log(1-p)+\frac{1}{2}\log(2\pi\sigma_0^2). \end{cases} \tag{8.19}$$

Il s'en suit que $p$ peut être re-paramétrisé en fonction de $\{\chi,\sigma^2\}$, i.e.,

$$p=\left(1+\exp\left(\frac{\chi}{\sigma^2}-\frac{1}{2}\log(2\pi\sigma_0^2)\right)\right)^{-1}, \tag{8.20}$$

ce qui permet de reformuler $\phi$ en fonction de $\chi$ et $\sigma^2$ (voir (8.16)). □

**Remarque 8.1.** *Une condition suffisante pour satisfaire* (8.17) *est :*

$$\sigma_0^2 \gg N(\max \boldsymbol{y}-\min \boldsymbol{y})^2+1/2\pi \tag{8.21}$$

Plusieurs comportements intéressants peuvent être remarqués à partir de la paramétrisation (8.15) du paramètre de régularisation $\chi$ en fonction de la variance du bruit $\sigma^2$, la probabilité $p$ de présence de discontinuités et la dynamique *a priori* des valeurs prises sur les segments (à travers le terme $\sigma_0^2$).

En particulier, comme espéré, $\chi$ est une fonction croissante de $\sigma^2$, ce qui souligne deux points. Premièrement, il est nécessaire d'augmenter $\chi$ pour régulariser davantage les signaux de plus grande dynamique mais de même SNR. Deuxièmement, à dynamique fixé, il convient de régulariser davantage les observations les plus bruitées.

A l'inverse, $\chi$ est une fonction décroissante de $p$ ce qui, comme nous l'illustrerons dans la suite, indique que la valeur optimale de $\chi$ décroit avec $p$ et doit être choisi de manière plus fine lorsque $p$ augmente.

## 8.3 Solution algorithmique

Nous avons précédemment dérivé, à partir de la Proposition 8.1, une fonction $\phi$ permettant de pénaliser le choix du paramètre de régularisation $\chi$ présent dans le Problème 8.1. Dans cette section, nous mettons au point une stratégie algorithmique permettant d'estimer efficacement la solution $\{\widehat{\boldsymbol{x}}_{\widehat{\chi}},\widehat{\chi},\widehat{\sigma}^2\}$ du Problème 8.1.

Etant donné que le Problème 8.1 est non convexe, alterner la minimisation sur chacune des variables $\boldsymbol{x}$, $\chi$ et $\sigma^2$ n'est pas une stratégie efficace. Afin de réduire une partie de ce problème, nous proposons d'estimer $\chi$ sur une grille discrète $\mathcal{X}$ prédéfinie. Ainsi, une solution candidate du Problème 8.1 peut être obtenue en résolvant

$$(\forall \chi\in\mathcal{X}) \quad (\widehat{\boldsymbol{x}}_\chi,\widehat{\sigma}^2_\chi)\in \underset{\boldsymbol{x}\in\mathbb{R}^N,\sigma^2>0}{\operatorname{Argmin}} \underbrace{\frac{1}{2\sigma^2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2+\frac{\chi}{\sigma^2}\|L\boldsymbol{x}\|_0+\phi(\chi,\sigma^2)}_{J(\boldsymbol{x},\chi,\sigma^2)}, \tag{8.22}$$

**Algorithme 10** Résolution du Problème 8.1 par une approche variationnelle Bayésienne

**Entrée :** Observation $\boldsymbol{y} \in \mathbb{R}^N$.
Ensemble prédéfini de paramètres de régularisation $\mathcal{X}$.
Hyperparamètres $\boldsymbol{\Phi} = \{\alpha_0, \alpha_1, \sigma_0^2\}$.

**Iterations :**

1: **Pour** $\chi \in \mathcal{X}$ **faire**
2: Calculer $\widehat{\boldsymbol{x}}_\chi = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \chi\|L\boldsymbol{x}\|_0$ .
3: Calculer $\widehat{\sigma}_\chi^2 = \|\boldsymbol{y}-\widehat{\boldsymbol{x}}_\chi\|_2^2/(N-1)$.

**Sortie :** Solution $\{\widehat{\boldsymbol{x}}_{\widehat{\chi}}, \widehat{\chi}, \widehat{\sigma}_{\widehat{\chi}}^2\}$ où $\widehat{\chi} = \arg\min_{\chi\in\mathcal{X}} J(\widehat{\boldsymbol{x}}_\chi, \chi, \widehat{\sigma}_\chi^2)$

que nous proposons d'approcher par

$$(\forall \chi \in \mathcal{X}) \quad \begin{cases} \widehat{\boldsymbol{x}}_\chi & = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y}-\boldsymbol{x}\|_2^2 + \chi\|L\boldsymbol{x}\|_0, \\ \widehat{\sigma}_\chi^2 & = \frac{\|\boldsymbol{y}-\widehat{\boldsymbol{x}}_\chi\|^2}{N-1}, \end{cases} \tag{8.23}$$

puis en sélectionnant le triplet $\{\widehat{\boldsymbol{x}}_{\widehat{\chi}}, \widehat{\chi}, \widehat{\sigma}_{\widehat{\chi}}^2\}$ tel que

$$\widehat{\chi} = \arg\min_{\chi\in\mathcal{X}} J(\widehat{x}_\chi, \chi, \widehat{\sigma}_\chi^2). \tag{8.24}$$

Notons que l'approximation faite en (8.23) consiste à utiliser les solutions du Problème 2.1 pour sonder l'espace $\{\boldsymbol{x}, \sigma^2\} \in \mathbb{R}^N \times \mathbb{R}_+$. Par conséquent, les étapes de l'algorithme, rapportées dans l'Algorithme 10, sont très succintes et la complexité de l'algorithme dépend seulement de la capacité à résoudre efficacement le Problème 2.1 pour toutes les valeurs $\chi \in \mathcal{X}$. Pour ce faire, nous utilisons l'algorithme de programmation dynamique *Pottslab* développé dans [193].

## 8.4 Sélection automatique de $\chi$ : Illustration et performances

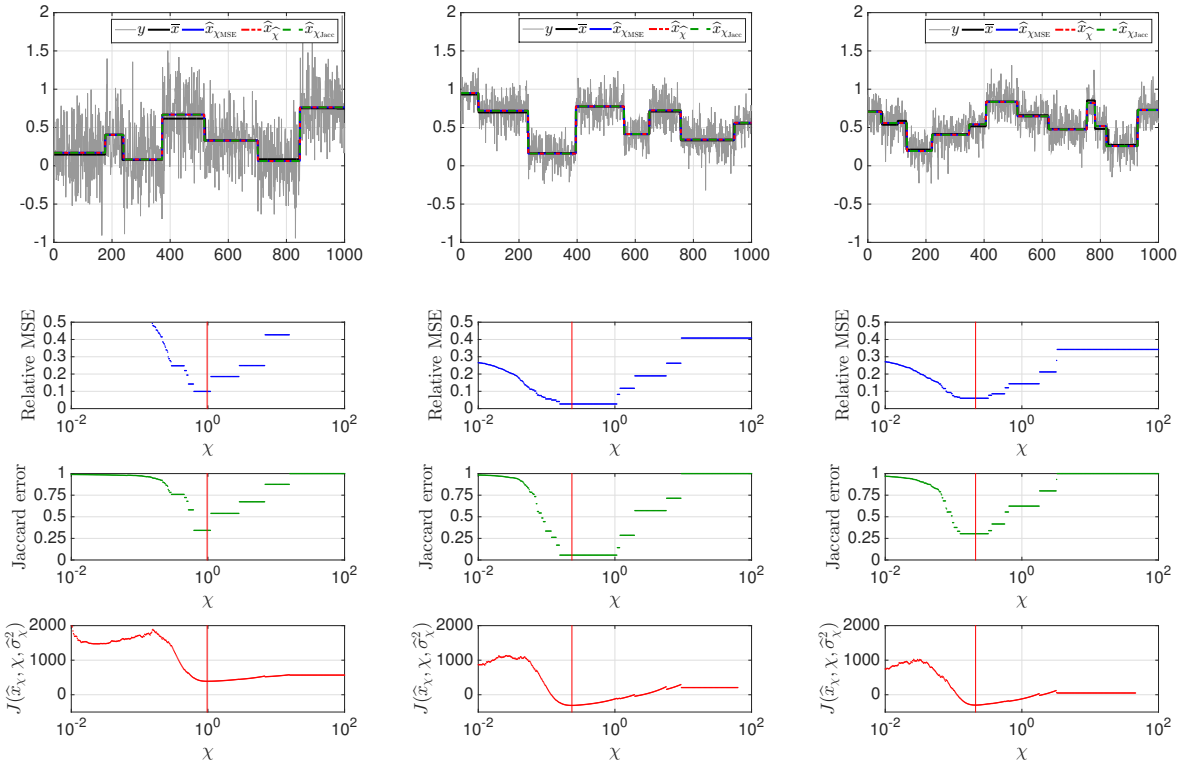
### 8.4.1 Configuration expérimentale et choix des hyperparamètres

**Données synthétiques.** Les observations $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{b} \in \mathbb{R}^N$ sont synthétisées en deux étapes. Premièrement, les discontinuités sont tirées uniformément sur l'ensemble $\{1, \ldots, N\}$ formant ainsi environ $pN$ segments, puis les valeurs prises sur chaque segment sont tirées uniformément entre une valeur minimale $\overline{x}_{\min}$ et une valeur maximale $\overline{x}_{\max}$ également données au préalable. Deuxièmement, on génère $\boldsymbol{b} \sim \mathcal{N}(0, \sigma^2 \mathbb{1}_N)$ et on forme la quantité $\boldsymbol{y} = \overline{\boldsymbol{x}} + \boldsymbol{b}$. Dans la suite, nous noterons $\overline{\boldsymbol{r}}$ (resp. $\overline{\boldsymbol{\mu}}$) le vecteur d'indicatrices des discontinuités (resp. les valeurs sur chaque segment) de $\overline{\boldsymbol{x}}$.

**Choix des hyperparamètres.** D'un côté, on fixe les hyperparamètres $\alpha_0 = \alpha_1 = 1$ afin que la loi bêta, assignée à la probabilité de discontinuités $p$ dans (8.12), se réduise à une loi uniforme (donc non informative) sur l'ensemble $[0, 1]$.

D'un autre côté, la distribution assignée en (8.10) aux valeurs sur chaque segment $\boldsymbol{\mu}$ dépend du choix de $\sigma_0$. D'après la Proposition 8.1, nous devons choisir $\sigma_0^2$ suffisamment large. Dans la suite, nous choisissons $2\pi\sigma_0^2 = 10^4$. L'impact du choix de $\sigma_0^2$ sur la solution sera discutée à la Section 8.4.5.

**Discrétisation de $\mathcal{X}$.** Dans nos expériences, $\mathcal{X}$ est composé de 500 valeurs équitablement espacées (en échelle logarithmique base 10) entre $10^{-5}$ et $10^5$.




(a) SNR = 1, $p = 0.01$. (b) SNR = 2, $p = 0.01$. (c) SNR = 2, $p = 0.015$.

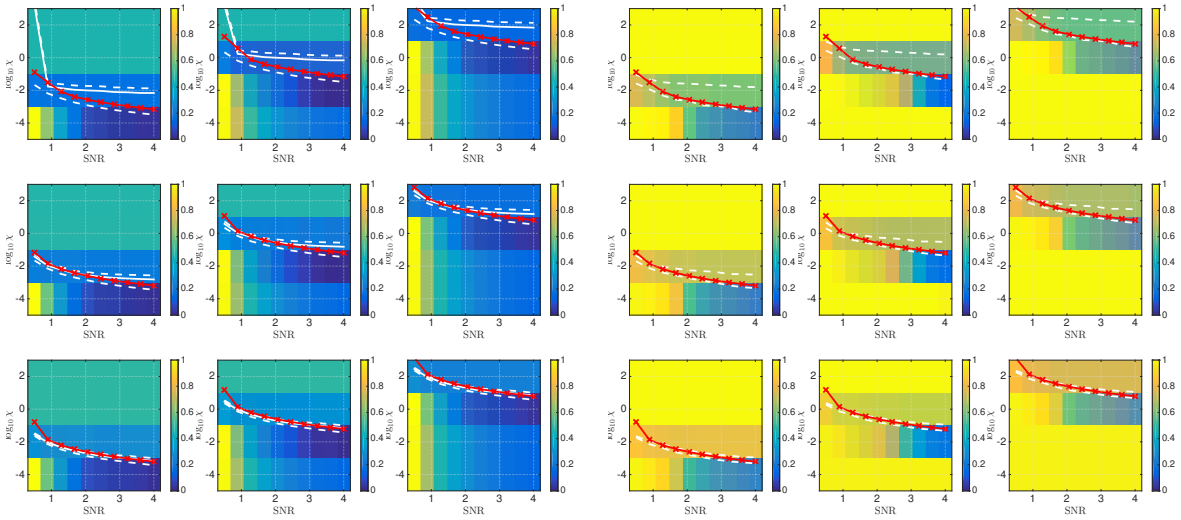
FIGURE 8.1 – **Illustration de la sélection automatique de $\chi$.** Pour chacune des configurations présentées, le critère de sélection proposé est affiché sur la 4e ligne où $\widehat{\chi}$ est indiqué par une ligne verticale rouge. On remarque que la solution $\widehat{\boldsymbol{x}}_\chi$ pour $\chi = \widehat{\chi}$, representée sur la 1ère ligne en rouge, bénéficie systématiquement d'une MSE relative (2e ligne) et d'une erreur de Jaccard (3e ligne) plus faibles que pour tout autre valeur de $\chi$.

**Quantification des performances.** Les performances sont quantifiées en termes d'erreur quadratique moyenne relative et d'erreur de Jaccard (voir Définition 2.7). Le premier évalue les performances d'estimation de $\overline{\boldsymbol{\mu}}$ alors que le second se concentre sur $\overline{\boldsymbol{r}}$. Pour tout estimateur $\widehat{\boldsymbol{r}}$, l'erreur de Jaccard sera calculée entre des versions lissées de $\overline{\boldsymbol{r}}$ et $\widehat{\boldsymbol{r}}$, i.e., $\overline{\boldsymbol{r}} * \mathcal{G}$ et $\widehat{\boldsymbol{r}}_{\widehat{\chi}} * \mathcal{G}$ où $\mathcal{G}$ est un filtre passe-bas gaussien à support compact de taille 5 points et d'écart type 0.5 (voir Section 2.3.3).

### 8.4.2 Illustration du principe de sélection automatique de $\chi$.

La Figure 8.1 illustre le principe de la méthode proposée sur plusieurs configurations de signaux pour lesquels $\overline{x}_{\max} - \overline{x}_{\min} = 1$. La première et la deuxième colonne mettent en évidence l'impact du SNR $= (\overline{x}_{\max} - \overline{x}_{\min})/3\sigma$ : SNR = 1 (1ere colonne) et SNR = 2 (2e colonne). La première et la troisième colonne montrent quant à elles l'impact de $p$ : $p = 0.01$ (1ere colonne) et $p = 0.15$ (3e colonne) pour un SNR fixé.

Pour chacun des exemples, la première ligne affiche le signal original $\overline{\boldsymbol{x}}$ (noir), l'observation dégradée $\boldsymbol{y}$ (gris), la solution obtenue par la méthode proposée $\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ (rouge), et les deux solutions optimales au sens de la MSEr (bleu) et de l'erreur de Jaccard (vert). La deuxième (resp. troisième ligne) présente l'erreur quadratique moyenne relative (resp. erreur de Jaccard) en fonction de $\chi$. Ce sont des fonctions de $\chi$ constantes par morceaux, ce qui indique que la



(a) Erreur quadratique moyenne relative (b) Erreur de Jaccard

FIGURE 8.2 – **Performances d'estimation ($\chi$ vs SNR).** Le paramètre estimé $\widehat{\chi}$ est affiché en rouge en fonction du SNR et comparé à la gamme oracle, délimitée en traits pointillés blancs, pour laquelle la solution correspondante minimise la MSE ou l'erreur de Jaccard. De haut en bas : $p = 0.005$, 0.010 et 0.015. De gauche à droite : $\overline{x}_{\max} - \overline{x}_{\min} = 0.1$, 1 et 10.

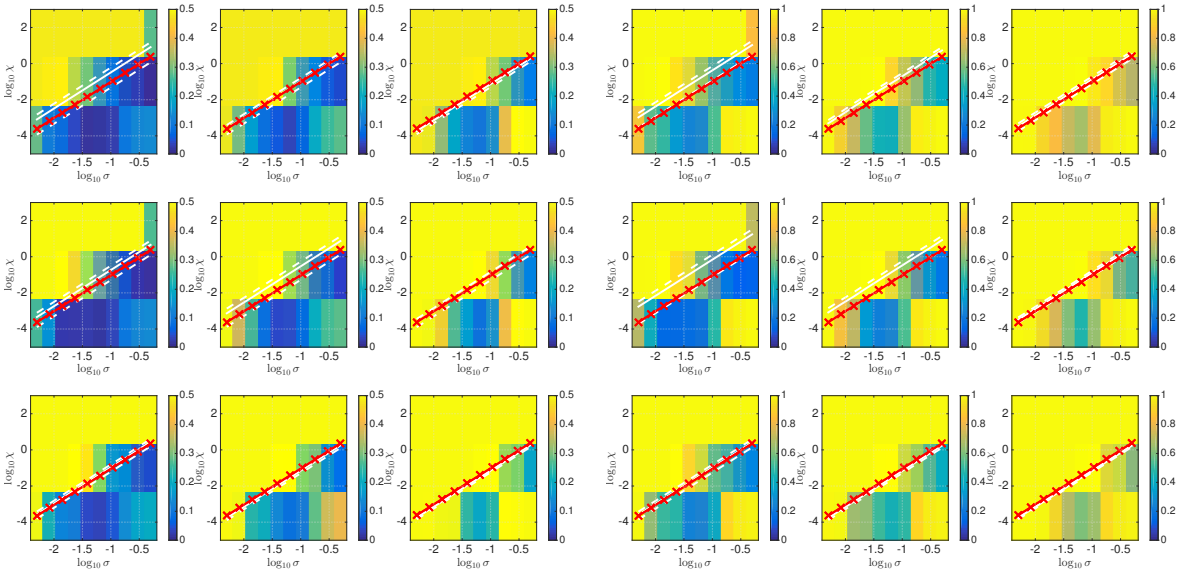
solution $\widehat{\boldsymbol{x}}_\chi$ du Problème 2.1 est identique pour tout un ensemble de valeurs $\chi$. En particulier, nous noterons $\mathcal{X}_{\text{MSE}}$ (resp. $\mathcal{X}_{\text{Jacc}}$) l'ensemble des valeurs de $\chi$ pour lesquelles $\widehat{\boldsymbol{x}}_\chi$ minimise la MSEr (resp. erreur de Jaccard). La dernière ligne affiche le critère proposé $J(\widehat{x}_\chi, \chi, \widehat{\sigma}^2_\chi)$ en fonction de $\chi$. Il est intéressant de remarquer que $J(\widehat{x}_\chi, \chi, \widehat{\sigma}^2_\chi)$ n'est pas constant par morceaux et que $\widehat{\chi}$ minimisant $J(\widehat{x}_\chi, \chi, \widehat{\sigma}^2_\chi)$ est unique. Sur chacune des trois lignes, la valeur de $\widehat{\chi}$ est reportée par une ligne verticale rouge. Idéalement, nous souhaiterions que $\widehat{\chi} \in \mathcal{X}_{\text{MSE}} \cup \mathcal{X}_{\text{Jacc}}$.

Dans chacun des trois cas, nous observons clairement que l'estimée $\widehat{\chi} \in \mathcal{X}_{\text{MSE}} \cup \mathcal{X}_{\text{Jacc}}$. Une examination plus poussée montre que lorsque le SNR diminue (voir Figure 8.1 (a) et (b)), la section d'un $\chi$ optimal est plus difficile, comme indiqué par la diminution des intervalles $\mathcal{X}_{\text{MSE}}$ et $\mathcal{X}_{\text{Jacc}}$. Nous observons également le même comportement lorsque $p$ augmente (voir Figure 8.1 (b) et (c)). Toutefois la méthode proposée parvient à choisir automatiquement un $\widehat{\chi} \in \mathcal{X}_{\text{MSE}} \cup \mathcal{X}_{\text{Jacc}}$. De plus, la solution $\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ apparait visuellement comme un excellent estimateur de $\overline{\boldsymbol{x}}$.

### 8.4.3 Quantification des performances

Dans cette section, nous proposons d'évaluer plus en détails les performances d'estimation de $\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ et d'examiner le comportement de $\widehat{\chi}$ en fonction $\sigma^2$, $\overline{x}_{\max} - \overline{x}_{\min}$, et $p$. Pour cela, nous considérons deux expériences. La première évalue les performances d'estimation en fonction du SNR pour différentes valeurs de $p$ et $\overline{x}_{\max} - \overline{x}_{\min}$ (voir Figure 8.2) alors que la seconde (voir Figure 8.3) se concentre sur le comportement en fonction de $\sigma$, pour différentes valeurs de $p$ et SNR.

Pour chacune des configurations examinées, nous rapportons la MSEr (resp. erreur de Jaccard) entre $\overline{\boldsymbol{x}}$ et $\widehat{\boldsymbol{x}}_\chi$ (resp. $\overline{\boldsymbol{r}}$ et $\widehat{\boldsymbol{r}}_\chi$) en fonction de $\chi$. Les intervalles $\mathcal{X}_{\text{MSE}}$ et $\mathcal{X}_{\text{Jacc}}$ sont délimités en traits pointillés blancs alors que leur valeur moyenne est dépeinte en traits pleins. Nous superposons également l'estimée $\widehat{\chi}$ en rouge. Les résultats sont moyennés sur 50 réalisations.



(a) Erreur quadratique moyenne relative (b) Erreur de Jaccard

FIGURE 8.3 – **Performances d'estimation ($\chi$ vs SNR).** Le paramètre estimé $\widehat{\chi}$ est affiché en rouge en fonction du SNR et comparé à la gamme oracle, délimitée en traits pointillés blancs, pour laquelle la solution correspondante minimise la MSE ou l'erreur de Jaccard. De haut en bas : $p = 0.005$, 0.010 et 0.015. De gauche à droite : SNR = 4, 2 et 1.

La Figure 8.2 illustre le choix optimal de $\chi$ en fonction de $p$, du SNR et de la dynamique $\overline{x}_{\max} - \overline{x}_{\min}$. Premièrement, nous observons que la région optimale (en blanc) se resserre lorsque le SNR et $p$ diminuent, ce qui montre que $\widehat{\chi}$ doit être choisi plus finement. Toutefois, nous remarquons que l'estimée $\widehat{\chi}$ reproduit le bon comportement en fonction de $p$ et $\sigma^2$ et se situe systématiquement entre les traits pointillés blancs. Cela indique que la solution bénéficie des meilleures performances d'estimation en termes de MSEr et d'erreur de Jaccard. De plus, $\widehat{\chi}$ varie correctement avec la dynamique, comme indiqué par la translation des courbes blanches et rouges (voir Figure 8.2 de gauche à droite), ce qui suggère que la méthode proposée fonctionne quelque soit la dynamique des observations.

Considérons à présent une dynamique $\overline{x}_{\max} - \overline{x}_{\min} = 1$ fixée. Dans ce cas, nous remarquons sur la Figure 8.3 que $\widehat{\chi}$ varie linéairement avec $\sigma^2$ et que $\widehat{\chi}$ diminue lorsque $p$ augmente (voir la translation de la courbe rouge en parcourant la Figure 8.3 de haut en bas). Ces deux observations sont cohérentes avec l'analyse dimensionnelle du Problème 2.1 qui énonce que $\chi \sim \sigma^2/2p$.

Dans chacun des cas, nous remarquons que l'estimateur proposé $\widehat{\chi}$ reproduit le même comportement que $\mathcal{X}_{\text{MSE}}$ et $\mathcal{X}_{\text{MAP}}$.

### 8.4.4 Comparaison contre de l'inférence bayésienne

A présent que la méthode est validée, nous proposons de comparer ses performances contre les estimateurs MAP et MMSE Bayésiens classiques associés au modèle hiérarchique dérivé à la Section 8.2.1 et dont la procédure MCMC est rappelée dans l'Annexe D. Le nombre d'itérations de Monte Carlo est fixé à $T_{\text{MC}} = 10^3$ et les hyperparamètres associés à la distribution antérieure $f(\boldsymbol{\mu}|\boldsymbol{r})$ décrite en (8.10) sont fixés à $\mu_0 = \widehat{\mathbb{E}}(\boldsymbol{y})$ et $\sigma_0^2 = 10\widehat{\text{Var}}(\boldsymbol{y})$,

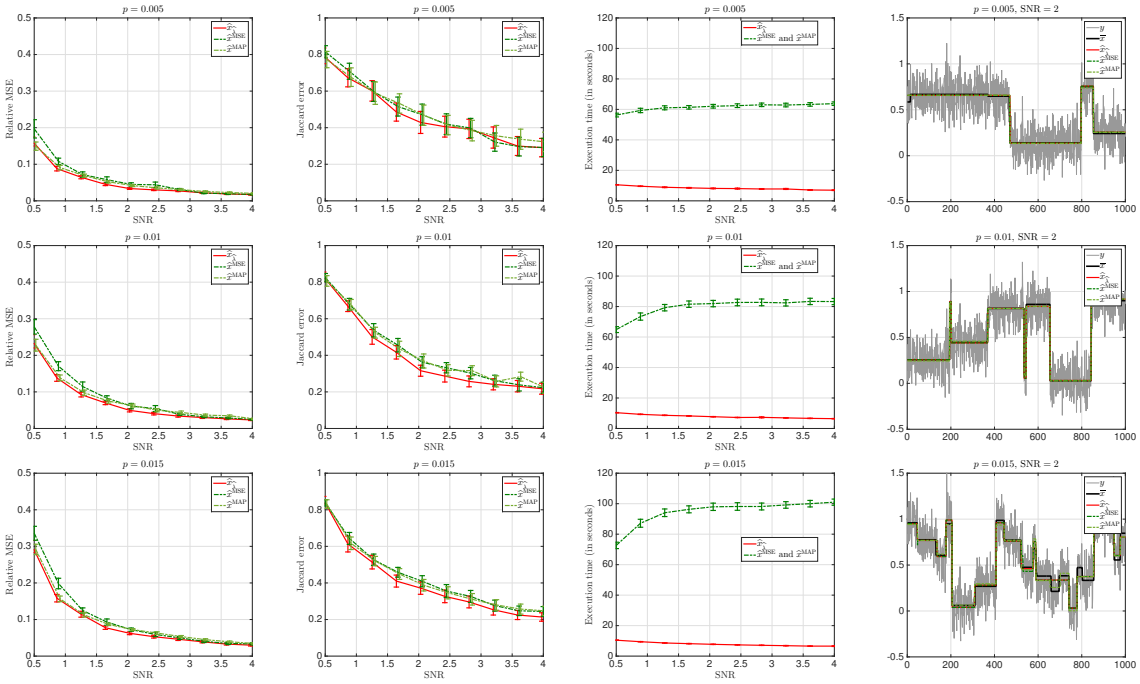



FIGURE 8.4 – **Comparaison contre de l'inférence bayésienne.** De haut en bas : $p = 0.005$, 0.010 et 0.015. De gauche à droite : Erreur quadratique moyenne relative, erreur de Jaccard, temps d'éxécution et exemples de solutions.

où $\widehat{\mathbb{E}}$ et $\widehat{\mathrm{Var}}(\cdot)$ désignent respectivement la moyenne et la variance empirique.

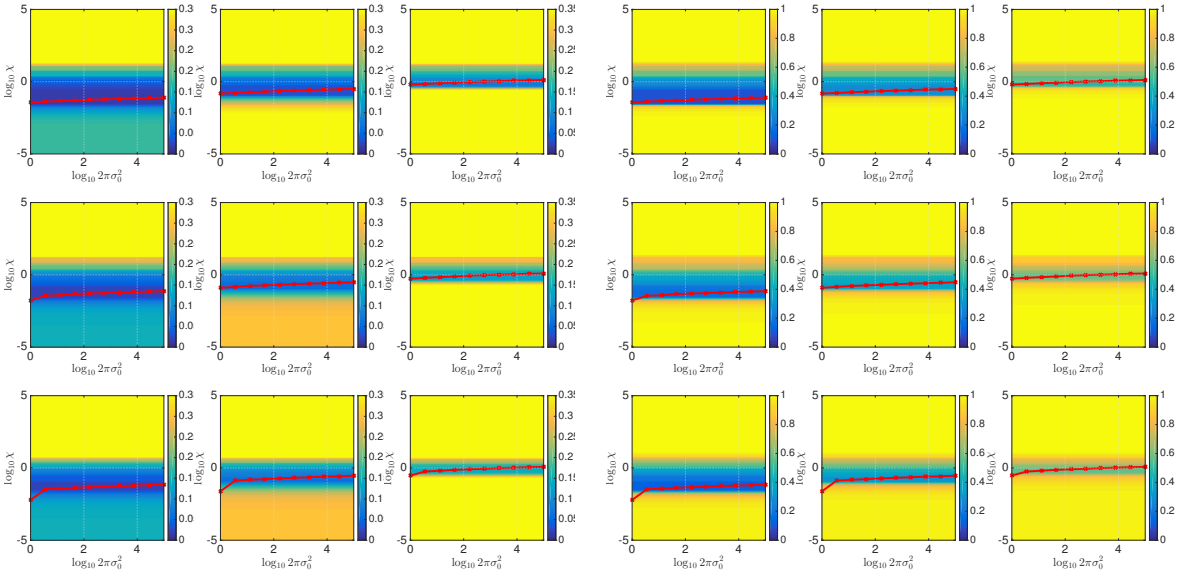
Les performances d'estimations en fonction du SNR sont rapportées par la Figure 8.4, où $\overline{x}_{\max} - \overline{x}_{\min} = 1$ et où chaque ligne correspond à une probabilité $p \in \{0.005, 0.01, 0.15\}$ différente. Les résultats obtenus avec l'estimateur $\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ proposé sont indiqués en rouge alors que les résultats associés aux estimateurs bayésiens $\widehat{\boldsymbol{x}}_{\mathrm{MAP}}$ et $\widehat{\boldsymbol{x}}_{\mathrm{MMSE}}$ sont rapportés en vert clair et vert foncé.

Globalement, les performances d'estimation de $\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ sont systématiquement comparables à celles de $\widehat{\boldsymbol{x}}_{\mathrm{MAP}}$ et $\widehat{\boldsymbol{x}}_{\mathrm{MMSE}}$ en termes de MSEr (1ère colonne) et d'erreur de Jaccard (2ème colonne). La troisième colonne illustre quant à elle l'apport en temps d'éxécution obtenu avec la méthode proposée par rapport aux approches MCMC. Il est intéressant de remarquer que l'approche proposée est déterministe et que son temps d'éxécution dépend principalement de $N$ et $|\mathcal{X}|$, alors que les valeurs des paramètres $p$ et $\sigma^2$ influent fortement sur le temps de calcul des approches MCMC. Des résultats similaires sont obtenus pour des valeurs de $N$ plus grandes. A titre d'exemple, pour $N = 10^4$, les approches MCMC nécessitent plus d'une heure de calcul alors que la méthode proposée est éxécutée en quelques minutes.

### 8.4.5 Robustesse de l'estimateur proposé

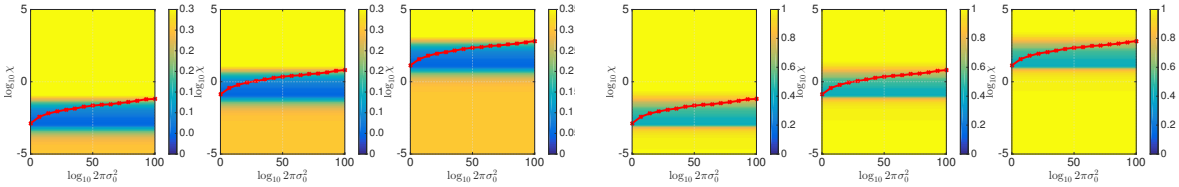
Jusqu'à présent nous avons choisi de fixer arbitrairement $2\pi\sigma_0^2 = 10^4$. Dans cette section, nous relachons ce choix et nous nous intéressons à l'impact de l'hyperparamètre $\sigma_0^2$ sur la solution estimée en fonction de $p$, $(\overline{x}_{\max} - \overline{x}_{\min})$, $\sigma^2$ et $\mathrm{SNR} = (\overline{x}_{\max} - \overline{x}_{\min})/3\sigma$.

**Impact de $p$ et SNR** La Figure 8.5 dépeint les résultats en fonction de $\mathrm{SNR} \in \{1, 2, 4\}$ et $p \in \{0.005, 0.01, 0.015\}$ pour $(\overline{x}_{\max} - \overline{x}_{\min}) = 1$. Dans chaque configuration, nous observons que malgré les variations de $\widehat{\chi}$ en fonction $\sigma_0^2$, les performances d'estimation de l'estimée



(a) Erreur quadratique moyenne relative (b) Erreur de Jaccard

FIGURE 8.5 – **Impact de $\sigma_0^2$ (2).** De haut en bas : $p = 0.005$, 0.010 et 0.015. De gauche à droite : SNR = 4, 2 et 1.



(a) Erreur quadratique moyenne relative (b) Erreur de Jaccard

FIGURE 8.6 – **Impact de $\sigma_0^2$ (1).** De gauche à droite : $\overline{x}_{\max} - \overline{x}_{\min} = 0.1$, 1 et 10.

$\widehat{\boldsymbol{x}}_{\widehat{\chi}}$ sont sensiblement identiques dans la gamme $\log_{10} 2\pi\sigma_0^2 = 1$ à 5 et offrent les meilleurs résultats en termes de MSEr et d'erreur de Jaccard.

**Impact de $\overline{x}_{\max} - \overline{x}_{\min}$.** La Figure 8.6 illustre comment les performances varient en fonction de $\overline{x}_{\max} - \overline{x}_{\min} \in \{0.1, 1, 10\}$ pour $p = 0.01$ et SNR = 1 fixes. Nous considérons également la gamme $\log_{10} 2\pi\sigma_0^2$ étendue de 1 à 100. Nous observons que les performances sont identiques à la translation de $(\overline{x}_{\max} - \overline{x}_{\min})^2$ près, ce qui indique que $\sigma_0^2$ doit être choisi indépendemment de la dynamique du signal. C'est un résultat prévisible dû à la dépendance de $\phi$ en $\sigma_0^2$ alors que, dans les approches Bayésiennes, $\sigma_0^2$ est généralement choisi proportionnellement à $\widehat{\mathrm{Var}}(\boldsymbol{y})$. De plus, nous remarquons que pour $\sigma_0^2 \to +\infty$, les performances d'estimation se dégradent considérablement. En effet, pour un tel cas limite, le terme en $\log(2\pi\sigma_0^2)$ domine tous les autres termes dans le critère $J$, ce que nous souhaitons éviter.

A partir de toutes ces simulations, nous observons que choisir $\log 2\pi\sigma_0^2 = 4$ apparait être un bon compromis entre la remarque théorique visant à sélectionner $\sigma_0^2$ suffisamment grand et le bon comportement numérique.

## 8.5 Conclusion et perspectives

Ce chapitre s'est intéressé au développement d'une méthode de sélection automatique et non supervisée du paramètre de régularisation présent dans le Problème 2.1. L'originalité de ce travail réside dans la mise en évidence de l'équivalence entre un problème d'estimation constant par morceaux bayésien et une version pénalisée du Problème 2.1. Cela nous a ainsi permis de mettre au point une approche variationnelle bayésienne profitant à la fois de la capacité des stratégies hiérarchiques bayésiennes à englober les hyperparamètres dans le modèle, et du faible coût de calcul inhérent aux approches variationnelles. Les performances de la méthode proposée se comparent favorablement contre les estimateurs bayésiens de l'état de l'art à la fois en terme de qualité d'estimation et en terme de temps de calcul.

Un avantage majeur de la méthode proposée est de facilement pouvoir s'étendre à d'autres modèles de dégradation de données univariées. Notamment, le cas du bruit de Poisson, i.e., $\boldsymbol{y} \sim \mathcal{P}(\overline{\boldsymbol{x}})$, pourrait également être abordé de façon similaire à partir de l'algorithme de programmation dynamique développé dans [75] et du modèle hierarchique Bayésien dérivé dans [68]. L'extension au cadre bidimensionnel est également envisagée.

# Conclusion

# Conclusion

## Bilan

Lorsque ces travaux de thèse ont été initiés, les processus invariants d'échelle étaient principalement supposés homogènes et univariés. Dans ce contexte, le formalisme multifractal basé sur les coefficients d'ondelettes dominants fournit un cadre pratique pour l'estimation de la régularité ponctuelle et du spectre multifractal. L'objectif de cette thèse a consisté à répondre aux questions posées par l'analyse des processus invariants d'échelle multivariés et non homogènes, par la formalisation de problèmes variationnels.

**Non homogène.** La première contribution a été d'envisager l'utilisation des approches TV pour détecter les régions au sein desquelles les propriétés d'invariance d'échelle des processus multifractals par morceaux peuvent être considérées homogènes. Le principe de la méthode proposée réside sur deux éléments clefs. Premièrement, nous avons montré que le formalisme multifractal doit être conduit localement en estimant les quantités liées au spectre multifractal, à savoir $(c_1, c_2)$ ou $(C_{1,j}, C_{2,j})_j$, dans une fenêtre glissante suffisamment grande. Deuxièmement, nous avons promu l'application de méthodes d'estimation/segmentation vectorielle conjointe pour détecter lesdites régions.

La deuxième contribution de la thèse s'est portée sur le développement d'un algorithme à la volée pour la détection de discontinuités dans des données vectorielles. Le principe de la méthode repose sur le découplage des conditions de Karush-Kuhn-Tucker à l'aide de variables non négatives afin d'être en mesure de mettre successivement à jour l'encadrement des variables primales et duales. Une solution algorithmique a été proposée où les variables non négatives sont supposées constantes par morceaux et estimées parmi un ensemble prédéfini dont la taille permet de régler le compromis entre la qualité de la solution et le temps de calcul. Nous avons également suggéré différentes façons d'accélérer la méthode proposée en profitant de sa structure parallélisable et de l'ordre de grandeur relatif des observations.

**Multivarié.** Dans un troisième temps, nous nous sommes intéressés à l'étude des processus autosimilaires multivariés en concevant une stratégie de régression non linéaire afin d'estimer non seulement la matrice de Hurst mais également la matrice de covariance des OfBm. L'estimation des paramètres des OfBm a été formulée par la minimisation d'une fonctionnelle non convexe quantifiant l'erreur quadratique entre la fonction de partition mesurée empiriquement et un modèle paramétrique. La principale difficulté a donc consisté à trouver le minimum global parmi un ensemble non convexe délimitant l'espace des solutions où l'OfBm est correctement défini. Pour y parvenir, le Chapitre 5 s'est concentré sur le cas bivarié en proposant une approximation de l'espace de recherche par une union d'ensembles convexes et en mettant au point un algorithme de séparation-évaluation pour trouver le minimum global. Un avantage majeur de la méthode paramétrique proposée est qu'elle bénéficie de bonnes performances d'estimation même pour des données de taille relativement faible ce qui peut

se montrer utile pour traiter des données réelles, qui sont généralement limitées.

La méthodologie a par ailleurs été envisagée et validée au Chapitre 6 pour l'analyse de Biv-OfGn. Contrairement aux Biv-OfBm, les Biv-OfGn offrent un cadre plus souple pour la modélisation jointe de l'invariance d'échelle de processus bivariés et peuvent s'avérer plus appropriés pour modéliser des données réelles. Leur application a été mise en oeuvre sur des données issues du trafic internet, à savoir les séries temporelles du nombre de bytes et paquets reçus. A cet effet, nous avons proposé l'introduction d'un paramètre d'intégration fractionnaire dans la définition des coefficients d'ondelette comme une astuce pratique et efficace pour assurer que les exposants de Hurst mesurés n'excèdent pas les valeurs autorisées par le modèle. Cela nous a ainsi permis de réaliser une procédure d'identification d'anomalies dans le trafic basée sur la différence d'exposants de Hurst.

**Contributions plus spécifiques autour de la variation totale.** Dans la perspective d'étudier le phénomène de fluorescence intermittente, une attention plus spécifque sur le bruit de Poisson a été portée au Chapitre 7 afin de déterminer les alternances entre deux intensités d'émission à partir du compte d'occurrences dans des intervalles de temps réguliers. Nous avons considéré l'utilisation des approches d'estimation TV comme un moyen pratique pour sonder les différents niveaux d'intensité avant d'appliquer un seuil choisi automatiquement et permettant d'identifier deux états d'émission. Les résultats obtenus démontrent l'intérêt d'utiliser la transformée d'Anscombe et bénéficient de temps de calcul suffisamment faibles pour être appliqués sur des données réelles.

La principale limitation des approches TV utilisées réside dans le fait que les performances dépendent fortement du choix d'un hyperparamètre contrôlant la quantité de régularisation à appliquer. La dernière contribution de cette thèse a donc été d'inclure son estimation dans le problème variationnel $\ell_0$-TV en rajoutant un terme de pénalisation. Le choix de ce dernier et le sens donné à l'hyperpararamètre optimal ont été dérivés au Chapitre 8 en établissant une correspondance avec l'estimateur MAP d'un modèle hierarchique Bayésien. Cette double représentation a également permis de concevoir une solution algorithmique efficace profitant à la fois de garanties de convergence et de la sélection automatique du paramètre de régularisation. A présent, nous disposons d'une solution rapide et entièrement non supervisée qui permet d'atteindre les meilleures performances d'estimation des discontinuités et des valeurs moyennes.

## Quelques perspectives

**Analyse multifractale inhomogène.** Puisque les quantités $(c_1, c_2)$ voire $(C_{1,j}, C_{2,j})_j$ sont estimées dans une fenêtre glissante, les estimées sont exponentiellement corrélées. Or les méthodes TV employées supposent que le bruit est blanc gaussien. Dès lors, il serait utile d'appliquer un algorithme de blanchiment de signal pour s'y rapporter.

De plus, nous avons montré que les approches vectorielles conjointes bénéficient de meilleures performances d'estimation que plusieurs approches scalaires indépendantes lorsque le SNR est du même ordre de grandeur sur chacune des composantes. Or, le bruit d'estimation de $(C_{1,j})_j$ et $(C_{2,j})_j$ varie en fonction de l'échelle d'analyse $2^j$. De fait, il serait intéressant d'examiner les performances d'estimation lorsque les termes d'attache aux données sont divisés par la variance d'estimation.

**Segmentation vectorielle.** Nous avons vu que les approches de segmentation vectorielle basées sur la TV dépendent fortement du choix des niveaux $(v_{q,1}, \ldots, v_{q,M})$. Pour la segmentation d'images RGB, les $Q$ classes sont usuellement choisies pour représenter les principales

couleurs susceptibles d'être présentes dans l'image. Cependant, pour la segmentation des propriétés multifractales, il est difficile de prévoir au préalable les différentes classes de valeurs de $(c_1, c_2)$ voire $(C_{1,j}, C_{2,j})_j$ présentes dans la texture à analyser. A cet effet, il serait utile d'envisager une méthode d'estimation préliminaire des $Q$ classes. Alternativement, il serait intéressant de pouvoir proposer $P$ niveaux $\boldsymbol{w_m} = (w_{p,m})_{1 \leq p \leq P}$ indépendemment pour chacune des composantes $m \in \{1, \ldots, M\}$, et d'inclure l'estimation des $Q$ classes à partir des valeurs $\boldsymbol{w}$ dans le problème variationnel.

**Algorithmes à la volée.** Le développement de l'algorithme à la volée soulève deux principales perspectives. Premièrement, il serait utile de pouvoir étendre l'algorithme à la volée au débruitage linéaire par morceaux où $L$ est remplacé par l'opérateur de secondes différences. Cependant, cette perspective implique également quelques difficultés dans sa mise en oeuvre notamment dûes au fait que les conditions de Karush-Kuhn-Tucker ne peuvent pas être découplées.

Deuxièmement, nous avons vu qu'il existe un délai entre l'instant où l'algorithme détecte qu'une discontinuité aurait dû être introduite et la position de la dite discontinuité. Dans l'optique de futurs travaux, nous envisageons l'utilisation de la version en ligne de l'algorithme à la volée présenté à la Section 4.4.4 comme un bon moyen de prédire la présence de discontinuités.

**Identification d'OfBm.** Etant donné que le nombre de paramètres caractérisant les OfBm requiert l'estimation de $M + M^2 + M(M-1)/2$ paramètres pour $M$ composantes, l'algorithme de séparation-évaluation proposé peut s'avérer prohibitif en temps de calcul pour $M > 2$. Dès lors, il serait souhaitable de développer une méthode non paramétrique permettant d'estimer les exposants de Hurst $(H_1, \ldots, H_M)$. Autrement, il serait avantageux de pouvoir estimer efficacement la matrice de mélange $W$.

**Fluorescence intermittente.** Nous avons vu que la solution (Ans.$\ell_2$-TV) permettait d'obtenir des résultats encourageants pour la détection des états d'émission ON et OFF. Une perspective consiste à mettre à profit le caractère à la volée de l'algorithme [52] pour détecter en temps réel les états OFF et, en appliquant un champ électrique à la boîte quantique, permettre le retour dans l'état ON. Cette stratégie ouvrirait ainsi la voie à la conception de nanostructures fortement émissives à intensité constante.

**Sélection du paramètre de régularisation.** Comme nous l'avons noté, l'extension de la méthode proposée à d'autres termes d'attache aux données et au cadre bidimensionnel pourrait s'avérer très utile. De même, il serait intéressant de considérer le cas où la variance du bruit et/ou la probabilité de présence de discontinuitiés varient dans le temps. Dès lors, le problème consisterait non plus à estimer un scalaire mais un vecteur modélisant le fait que le paramètre de régularisation optimal varie également dans le temps.

# Annexe

# Annexe A

# Définitions utiles en optimisation convexe

Les résultats d'optimisation convexe énoncés dans ce manuscrit concernent une classe particulière de fonctions convexes $\Gamma_0(\mathbb{R}^N)$ dont la définition requiert les notions suivantes.

**Définition A.1.** *(Epigraphe d'une fonction). Soit $f : \mathbb{R}^N \to [-\infty, +\infty]$. L'épigraphe de $f$ est l'ensemble noté* epi $f$ *défini par :*

$$\operatorname{epi} f = \{(x, \alpha) \in \mathbb{R}^N \times \mathbb{R} \mid f(x) \leq \alpha\}. \tag{A.1}$$

**Définition A.2.** *(Fonction semi-continue inférieurement). Soit $f : \mathbb{R}^N \to [-\infty, +\infty]$ et soit $u \in \mathbb{R}^N$. La fonction $f$ est dite semi-continue inférieurement (s.c.i.) si* epi $f$ *est fermé.*

**Définition A.3.** *(Domaine d'une fonction et fonction propre). Soit une fonction $f : \mathbb{R}^N \to [-\infty, +\infty]$. Le domaine de $f$ est l'ensemble noté* dom $f$ *défini par :*

$$\operatorname{dom} f = \Big\{u \in \mathbb{R}^N \mid f(u) < +\infty\Big\}. \tag{A.2}$$

*On dit que $f$ est une fonction propre si* dom $f \neq \emptyset$ *(i.e., si $f$ n'est pas identiquement égale à l'infini).*

**Définition A.4.** *(Classe $\Gamma_0(\mathbb{R}^N)$). Soit $\Gamma_0(\mathbb{R}^N)$ la classe des fonctions $f : \mathbb{R}^N \to ]-\infty, +\infty]$ convexes, semi-continues inférieurement et propres.*

**Définition A.5.** *(Enveloppe conique). Soit $\mathcal{C}$ un ensemble convexe. L'enveloppe conique de $\mathcal{C}$ est l'ensemble noté* cone $\mathcal{C}$ *défini par :*

$$\operatorname{cone} \mathcal{C} = \bigcup_{\lambda > 0} \Big\{\lambda x \mid x \in \mathcal{C}\Big\}. \tag{A.3}$$

**Définition A.6.** *(Intérieur relatif fort). Soit $\mathcal{C}$ un ensemble convexe. L'intérieur relatif fort de $\mathcal{C}$ est l'ensemble noté* sri $\mathcal{C}$ *défini par :*

$$\operatorname{sri} \mathcal{C} = \Big\{x \in \mathcal{C} \mid \operatorname{cone}(\mathcal{C} - x) = \overline{\operatorname{Vect}}(\mathcal{C} - x)\Big\} \tag{A.4}$$

*où* $\operatorname{Vect}(\mathcal{C} - x)$ *est l'ensemble des combinaisons linéaires des vecteurs de $(\mathcal{C} - x)$ et où* $\overline{\operatorname{Vect}}(\mathcal{C} - x)$ *désigne son adhérence.*

# Annexe B

# Algorithme de segmentation vectorielle couplée

## B.1 Enoncé du problème

Nous considérons l'extension du Problème 2.10 au cadre vectoriel où cette fois-ci la segmentation n'est pas imposée conjointe à toutes les composantes, i.e., l'hypothèse $\boldsymbol{\theta_q} \equiv \boldsymbol{\theta_{q,m}}$ $(\forall m \in \{1,\dots,M\})$ est relâchée et nous considérons que $\boldsymbol{\theta} \in \mathbb{R}^{(Q+1)\times M\times|\Omega|}$. Par conséquent, l'ensemble $(\mathrm{D}\boldsymbol{\theta_q})_{\underline{\ell}} \in [-1,1]^{M\times 2}$ n'est plus réduit à un singleton et les coefficients $(\mathrm{D}\boldsymbol{\theta_q})_{\underline{\ell}}$ définissent alors une éllipsoïde centrée en 0 de dimension $0 \leq r \leq 2$ dans l'espace $[-1,1]^2$ où $r$ est le nombre de valeurs singulières. Les valeurs singulières $\zeta_{q,\underline{\ell}} = (\zeta_{q,\underline{\ell},1}, \zeta_{q,\underline{\ell},2})$ correspondent aux longueurs respectives des axes principaux de cet ellipsoïde.

Afin de segmenter la quantité $\mathbf{y} = (\boldsymbol{y_1},\dots,\boldsymbol{y_M})^\top$ tout en désirant que les masques $(\Omega_{1,m},\dots,\Omega_{Q,m})$ soient plus ou moins similaires sur chaque composante $m \in \{1,\dots,M\}$, nous proposons d'introduire des corrélations entre les composantes en modifiant le Problème 3.2 comme suit :

**Problème B.1** (Problème $\ell_0$-TV 2D vectoriel à $Q$ niveaux avec couplage entre composantes)**.** *Soit $\mathbf{y} = (\boldsymbol{y_1},\dots,\boldsymbol{y_M})^\top \in \mathbb{R}^{M\times|\Omega|}$ et un ensemble de valeurs de niveaux $(v_{q,m})_{1\leq q\leq Q, 1\leq m\leq M}$ avec comme convention $v_{q,m} \leq v_{q+1,m}$. Alors le problème consiste à résoudre*

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}^\top\in[0,1]^{(Q+1)\times M\times|\Omega|}}{\arg\min} \sum_{q=1}^{Q}\sum_{m=1}^{M} (\boldsymbol{\theta_{q,m}} - \boldsymbol{\theta_{q+1,m}})^\top (\boldsymbol{y_m} - v_{q,m})^2 + \lambda \sum_{q=1}^{Q} STV(\boldsymbol{\theta_q})$$

$$\textit{assujetti à} \qquad (\forall m \in \{1,\dots,M\}) \quad \begin{cases} \boldsymbol{\theta_{1,m}} \equiv 1, \\ \boldsymbol{\theta_{Q+1,m}} \equiv 0, \\ 1 \geq \boldsymbol{\theta_{2,m}} \geq \ldots \geq \boldsymbol{\theta_{Q,m}} \geq 0, \end{cases} \tag{B.1}$$

*où, pour tout $q \in \{1,\dots,Q\}$, le tenseur de structure des pénalisations vaut*

$$STV(\boldsymbol{\theta_q}) = \sum_{\underline{\ell}=1}^{N} \|\zeta_{q,\underline{\ell}}\|_p \quad \textit{avec} \quad p \geq 1$$

*et où $\zeta_{q,\underline{\ell}} = (\zeta_{q,\underline{\ell},1}, \zeta_{q,\underline{\ell},2}) \in \mathbb{R}^2$ est défini à partir de la décomposition en valeurs singulières de $(D\boldsymbol{\theta_q})_{\underline{\ell}} \in \mathbb{R}^{M\times 2}$ :*

$$(D\boldsymbol{\theta_q})_{\underline{\ell}} = U_{q,\underline{\ell}} Z_{q,\underline{\ell}} (V_{q,\underline{\ell}})^\top \tag{B.2}$$

*où*

$$\begin{cases} (U_{q,\underline{\ell}})^\top U_{q,\underline{\ell}} = \mathbb{1}_M, \\ V_{q,\underline{\ell}}(V_{q,\underline{\ell}})^\top = \mathbb{1}_2, \\ Z_{q,\underline{\ell}} = \begin{pmatrix} \zeta_{q,\underline{\ell},1} & 0 & \dots & \dots & 0 \\ 0 & \zeta_{q,\underline{\ell},2} & 0 & \dots & 0 \end{pmatrix}^\top. \end{cases} \tag{B.3}$$

Le Problème B.1 peut être interprêté comme une version discrète de la relaxation proposée dans [195].

## B.2 Solution algorithmique

Afin de proposer un algorithme efficace pour résoudre le Problème B.1, nous reformulons d'abord le critère (B.1) comme suit

$$\widehat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}^\top \in [0,1]^{(Q+1)\times M\times|\Omega|}}{\arg\min} \sum_{q=2}^{Q}\sum_{m=1}^{M} \theta_{q,m}^\top \Big( (\boldsymbol{y_m} - v_{q,m})^2 - (\boldsymbol{y_m} - v_{q-1,m})^2 \Big) + \lambda \sum_{q=2}^{Q} \mathrm{STV}(\boldsymbol{\theta_q}) + \iota_{E_0}(\boldsymbol{\theta}) + \iota_{E_1}(\boldsymbol{\theta}) + \iota_{E_2}(\boldsymbol{\theta}) \tag{B.4}$$

où, pour tout $k \in \{0,1,2\}$, $\iota_{E_k}$ désigne la fonction indicatrice sur l'ensemble convexe fermé et non vide $E_k \subset \mathbb{R}^{(Q-1)\times M\times\Omega}$, c'est-à-dire $\iota_{E_k}(\boldsymbol{\theta}) = 0$ si $\boldsymbol{\theta} \in E_k$ et $+\infty$ sinon. L'ensemble $E_0$ correspond à la contrainte sur la dynamique qui impose que $\boldsymbol{\theta}$ appartienne à $[0,1]^{(Q-1)\times M\times|\Omega|}$, i.e.,

$$E_0 = \{\boldsymbol{\theta} \in [0,1]^{(Q-1)\times M\times|\Omega|}\}$$

et où

$$E_1 = \Big\{ \boldsymbol{\theta} \in \mathbb{R}^{(Q-1)\times M\times|\Omega|} \,|\, \boldsymbol{\theta_{2q}} - \boldsymbol{\theta_{2q+1}} \geq 0, (\forall q \in \{1,\dots,\lfloor (Q-1)/2 \rfloor\} \Big\} \tag{B.5}$$

et

$$E_2 = \Big\{ \boldsymbol{\theta} \in \mathbb{R}^{(Q-1)\times M\times|\Omega|} \,|\, \boldsymbol{\theta_{2q+1}} - \boldsymbol{\theta_{2q+2}} \geq 0, (\forall k \in \{1,\dots,\lfloor (Q-2)/2 \rfloor\} \Big\}. \tag{B.6}$$

Le critère (B.4) est la somme de 5 fonctions convexes possiblement non lisses, semi-continues inférieurement et propres, et implique un tenseur de structure des pénalisations comprenant un opérateur linéaire $D$. Par conséquent, il peut être minimisé à l'aide d'un algorithme proximal dont les itérations ont été introduites dans [200, 53] et son résumées dans l'Algorithme 11. Sous certaines hypothèses techniques assurant l'existence de la solution (voir [200, 53]), les itérées $\left(\boldsymbol{\theta}^{[k]}\right)_{k\in\mathbb{N}}$ convergent vers le minimiseur de (B.4).

Les opérateurs proximaux présents dans l'Algorithme 11 ont une expression explicite. En effet, la forme explicite de $\mathrm{prox}_{\|\cdot\|_p}$ pour $p = 2$ est donnée dans [166], alors que le cas $p = 1$ correspond à l'opérateur de seuillage doux (voir Exemple 2.2). Il convient de noter que lorsque $p = 2$, l'étape de décomposition en valeurs singulières peut être évitée [41]. D'un autre côté, nous avons noté

$$\psi_{q,m}(\boldsymbol{\theta}) = \boldsymbol{\theta}^\top \Big( (\boldsymbol{y_m} - v_{q,m})^2 - (\boldsymbol{y_m} - v_{q-1,m})^2 \Big) \tag{B.7}$$

dont l'opérateur proximal s'écrit

$$\mathrm{prox}_{\sigma^{-1}\psi_{q,m}}\, \boldsymbol{\theta} = \boldsymbol{\theta} - \sigma^{-1}\Big( (\boldsymbol{y_m} - v_{q,m})^2 - (\boldsymbol{y_m} - v_{q-1,m})^2 \Big). \tag{B.8}$$

**Algorithme 11** Résolution du Problème B.1 basée sur l'algorithme primal-dual

**Entrée :** Paramètres $\tau > 0$ et $\sigma \in ]0, \tau^{-1}\big(\|D^\top D\| + 3\big)^{-1}[$.

Initialiser $\boldsymbol{\theta}^{[0]} = (\boldsymbol{\theta_{q,m}}^{[0]})_{2\leq q\leq Q, 1\leq m\leq M} \in \mathbb{R}^{(Q-1)\times M\times|\Omega|}$.

Initialiser $\widetilde{\boldsymbol{y}}^{[0]} \in \mathbb{R}^{(Q-1)\times M\times 2|\Omega|}$.

Initialiser $\widetilde{\widetilde{\boldsymbol{y}}}^{[0]}, \bar{\boldsymbol{y}}^{[0]}, \bar{\bar{\boldsymbol{y}}}^{[0]} \in \mathbb{R}^{(Q-1)\times M\times|\Omega|}$.

**Pour** $k = 0, 1, \ldots$ **faire**

  **Etapes primales : mise à jour de** $\boldsymbol{\theta}^{[k+1]}$

  **Pour** $q \in \{2, \ldots, Q\}$ **faire**

    **Pour** $m \in \{1, \ldots, M\}$ **faire**

      $z_{q,m}^{[k]} = \theta_{q,m}^{[k]} - \tau\big(D^\top \widetilde{y}_{q,m}^{[k]} - \widetilde{\widetilde{y}}_{q,m}^{[k]} - \bar{y}_{q,m}^{[k]} - \bar{\bar{y}}_{q,m}^{[k]}\big)$

  $\theta^{[k+1]} = P_{E_0} z^{[k]}$

  $\widetilde{\theta}^{[k+1]} = 2\theta^{[k+1]} - \theta^{[k]}$

  **Etapes duales : mise à jour de** $\widetilde{y}^{[k+1]}$, $\widetilde{\widetilde{y}}^{[k+1]}$, $\bar{y}^{[k+1]}$, $\bar{\bar{y}}^{[k+1]}$

  **Pour** $q \in \{2, \ldots, Q\}$ **faire**

    **Pour** $m \in \{1, \ldots, M\}$ **faire**

      $\widetilde{u}_{q,m}^{[k+1]} = \widetilde{y}_{q,m}^{[k]} + \sigma D\widetilde{\theta}_{q,m}^{[k+1]}$

  $\widetilde{\widetilde{u}}^{[k+1]} = \widetilde{\widetilde{y}}^{[k]} + \sigma\widetilde{\theta}^{[k+1]}$

  $\bar{u}^{[k+1]} = \bar{y}^{[k]} + \sigma\widetilde{\theta}^{[k+1]}$

  $\bar{\bar{u}}^{[k+1]} = \bar{\bar{y}}^{[k]} + \sigma\widetilde{\theta}^{[k+1]}$

  **Pour** $q \in \{2, \ldots, Q\}$ **faire**

    **Pour** $\underline{\ell} \in \{1, \ldots, N\}$ **faire**

      Calculer $\zeta_{q,\underline{\ell},1}^{[k+1]}$ et $\zeta_{q,\underline{\ell},2}^{[k+1]}$ à partir de $\widetilde{u}_{q,\cdot,\underline{\ell}}^{[k+1]}$ (cf. (B.2))

      $\eta_{q,\underline{\ell},\cdot}^{[k+1]} = \zeta_{q,\underline{\ell},\cdot}^{[k+1]} - \frac{\sigma}{\lambda}\mathrm{prox}_{\frac{\lambda}{\sigma}\|\cdot\|_p}\big(\frac{\lambda}{\sigma}\zeta_{q,\underline{\ell},\cdot}^{[k+1]}\big)$

      Calculer $\widetilde{y}_{q,\underline{\ell}}^{[k+1]}$ à partir de $\eta_{q,\underline{\ell},\cdot}^{[k+1]}$ (cf. (B.2))

    **Pour** $m \in \{1, \ldots, M\}$ **faire**

      $\widetilde{\widetilde{y}}_{q,m}^{[k+1]} = \widetilde{\widetilde{u}}_{q,m}^{[k+1]} - \sigma\mathrm{prox}_{\sigma^{-1}\psi_{q,m}}\big(\sigma^{-1}\widetilde{\widetilde{u}}_{q,m}^{[k+1]}\big)$

  $\bar{y}^{[k+1]} = \bar{u}^{[k+1]} - \sigma P_{E_1}(\sigma^{-1}\bar{u}^{[k+1]})$

  $\bar{\bar{y}}^{[k+1]} = \bar{\bar{u}}^{[k+1]} - \sigma P_{E_2}(\sigma^{-1}\bar{\bar{u}}^{[k+1]})$

**Sortie :** Solution ...

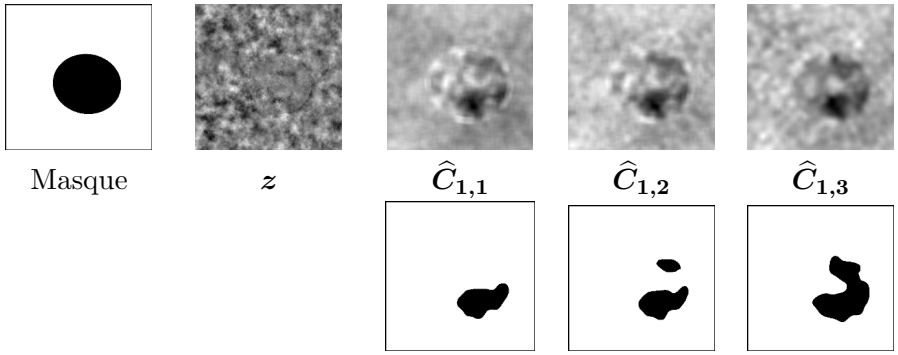



FIGURE B.1 – **Performances de segmentation couplée.** 1ère ligne (de gauche à droite) : masque permettant de générer les données, données originales $\boldsymbol{z}$, estimées des valeurs moyennes de $\widehat{\boldsymbol{C}}_{\mathbf{1},\boldsymbol{j}}$ pour $j = 1$, $j = 2$, et $j = 3$. 2ème ligne (de gauche à droite) : Solution du Problème 3.2 pour $j = 1$, $j = 2$, et $j = 3$. 3ème ligne (de gauche à droite) : Solution du Problème 3.3 pour $j = 1$, $j = 2$, et $j = 3$.

Finalement, les projections sur $E_0, E_1$, et $E_2$ se réduisent à des projections sur des hyperslabs [22, Exemple 28.17]

D'autres solutions primal-dual auraient pu utilisées, comme celle dérivée dans [38, 50].

## B.3 Quantification des performances

Nous proposons de comparer les performances de segmentation de la solution du Problème B.1 contre celles de la solution du Problème 3.2 sur une texture concue à partir de l'inclusion d'un patch 2D d'une réalisation de marche aléatoire multifractale de paramètres $(c_1, c_2)_A = (0.8, -0.005)$ sur un fond de paramètres $(c_1, c_2)_B = (0.5, -0.05)$. Une réalisation de telle texture est représentée sur la Figure B.1. Dans nos simulations, nous utilisons $\mathbf{y} = \widehat{\boldsymbol{C}}_{\mathbf{1}}$, $Q = 2$, $\lambda = 20$, et $p = 2$.

**Performances de segmentation.** La solution proposée, dont le résultat est dépeint sur la Figure. B.1-(ligne du bas), atteint un taux plus faible de coefficients mal classés à chaque échelle, ce qui illustre l'intérêt de l'approche vectorielle. Cependant, la segmentation est différente à chaque échelle. Afin de segmenter la texture originelle $\boldsymbol{z}$, l'utilisateur doit mettre au point une procédure pour combiner les segmentations obtenues à chaque échelle. Des conclusions similaires sont obtenues pour différentes réalisations et configurations.

# Annexe C

# Démonstrations

## C.1 Démonstration du Théorème 5.2

Considérons une séquence $\{\Theta_N\}_{N\in\mathbb{N}} \in \Xi$ (i.e., non nécessairement composée de minima). Nous affirmons que

$$C_N(\Theta_N) \xrightarrow{P} 0 \quad \Rightarrow \quad \Theta_N \xrightarrow{P} \overline{\Theta}. \tag{C.1}$$

Par contradiction, supposons que l'on choisisse une sous-séquence $\{\Theta_{N(r)}\}_{r\in\mathbb{N}}$ tel que, avec probabilité positive,

$$C_{N(r)}(\Theta_{N(r)}) < r^{-1} \quad \text{et} \quad \|\Theta_{N(r)} - \overline{\Theta}\| \geq C_0 > 0. \tag{C.2}$$

Alors, (5.40) et (5.42) impliquent qu'il existe des indices $j^*$, $i_1^*$ et $i_2^*$, une constante $\delta > 0$ et une séquence d'ensembles

$$E_{\delta,N(r)} = \{\omega : |\log_2 |E_{i_1^*,i_2^*}(2^{j^*}, \Theta_{N(r)})| - \log_2 |E_{i_1^*,i_2^*}(2^{j^*}, \overline{\Theta})|| \geq \delta\} \tag{C.3}$$

tels que

$$P(E_{\delta,N(r)}) \geq C_1 > 0 \tag{C.4}$$

pour un certain $C_1 > 0$. Alors, pour $\varepsilon \in (0, \delta)$, le Théorème 5.1 implique que pour un certain $C_2$,

$$\begin{aligned}\frac{1}{r} > \sum_{j=j_1}^{j_2} \sum_{1\leq i_1\leq i_2\leq M} &\{\log_2 |S_{i_1,i_2}(2^j)| - \log_2 |E_{i_1,i_2}(2^j, \overline{\Theta})| \\ &+ \log_2 |E_{i_1,i_2}(2^j, \overline{\Theta})| - \log_2 |E_{i_1,i_2}(2^j, \Theta_{N(r)})|\}^2 \\ &\geq \left\{|\log_2 |E_{i_1^*,i_2^*}(2^{j^*}, \overline{\Theta})| - \log_2 |E_{i_1^*,i_2^*}(2^{j^*}, \Theta_{N(r)})|| - \varepsilon\right\}^2 \\ &\geq C_2 > 0 \end{aligned} \tag{C.5}$$

avec probabilité non nulle (contradiction). Par conséquent, (C.1) est vérifié. A présent, considérons la séquence (5.43), et remarquons que

$$0 \leq C_N(\widehat{\Theta}_N^M) = \inf_{\Theta\in\Xi} C_N(\Theta) \leq C_N(\overline{\Theta}) \xrightarrow{P} 0, \tag{C.6}$$

grâce au Théorème 5.1. Donc, par (C.1), la limite (5.44) est satisfaite. □

## C.2 Démonstration du Théorème 5.3

Réécrivons

$$C_N(\Theta) = \sum_{j=j_1}^{j_2} \sum_{1\le i_1\le i_2\le M} (f_N)_{i_1,i_2,j}(\Theta) \tag{C.7}$$

avec

$$(f_N)_{i_1,i_2,j}(\Theta) = \{\log_2 |S_{i_1,i_2}(2^j)| - \log_2 |E_{i_1,i_2}(2^j,\Theta)|\}^2. \tag{C.8}$$

Il apparait donc clairement que pour $k \in \mathbb{N}^*$,

$$C_N^{(k)}(\Theta) = \sum_{j=j_1}^{j_2} \sum_{1\le i_1\le i_2\le M} (f_N^{(k)})_{i_1,i_2,j}(\Theta). \tag{C.9}$$

Fixons un triplet $(i_1, i_2, j)$. Par (5.39), (5.41) et (5.42), les trois premières dérivées de $(f_N)_{i_1,i_2,j}(\Theta)$ par rapport à $\Theta$ sont bien définies dans $\operatorname{int}\Xi$. Les deux premières dérivées peuvent s'exprimer en fonction de $\Theta$ comme suit

$$\begin{aligned}(f'_N)_{i_1,i_2,j}(\Theta) &= \nabla_\Theta (f_N)_{i_1,i_2,j}(\Theta)^* \\ &= -2\{\log_2 |S_{i_1,i_2}(2^j)| - \log_2 |E_{i_1,i_2}(2^j,\Theta)|\}\Lambda_{i_1,i_2}(2^j,\Theta),\end{aligned} \tag{C.10}$$

et

$$\begin{aligned}(f''_N)_{i_1,i_2,j}(\Theta) &= \nabla_\Theta[\nabla_\Theta (f_N)_{i_1,i_2,j}(\Theta)^*] \\ &= 2\Big\{\Lambda_{i_1,i_2}(2^j,\Theta)\Lambda_{i_1,i_2}(2^j,\Theta)^* \\ &\quad - [\log_2 |S_{i_1,i_2}(2^j)| - \log_2 |E_{i_1,i_2}(2^j,\Theta)|]\nabla_\Theta \Lambda_{i_1,i_2}(2^j,\Theta)\}.\end{aligned} \tag{C.11}$$

De manière similaire, $(f'''_N)_{i_1,i_2,j}(\Theta)$ est composé de sommes et produits de $\log_2 |S(2^j)| - \log_2 |E(2^j,\Theta)|$ et des dérivées de $\Lambda_{i_1,i_2}(2^j,\Theta)$.

Réécrivons $C'_N(\Theta) = \{(C'_N)_l(\Theta)\}_{l=1,\dots,\dim\Xi}$ via un développement de Taylor du second ordre de $(C'_N)_l(\Theta)$ avec reste de Lagrange,

$$\begin{aligned}\mathbb{R} \ni (C'_N)_l(\widehat{\Theta}_N) - (C'_N)_l(\overline{\Theta}) &= \Big\{\nabla_\Theta (C'_N)_l(\overline{\Theta})(\widehat{\Theta}_N - \overline{\Theta}) \\ &\quad + (\widehat{\Theta}_N - \overline{\Theta})^* \frac{\nabla_\Theta[\nabla_\Theta (C'_N)_l((\widetilde{\Theta}_N)_l)^*]}{2!}(\widehat{\Theta}_N - \overline{\Theta})\Big\},\end{aligned} \tag{C.12}$$

où chaque entrée $(\widetilde{\Theta}_N)_l$, $l = 1,\dots,\dim\Xi$, est une valeur du paramètre appartenant au segment entre $\overline{\Theta}$ et $\widehat{\Theta}_N$. Par conséquent,

$$\begin{aligned}\mathbb{R}^{\dim\Xi} \ni C'_N(\widehat{\Theta}_N) - C'_N(\overline{\Theta}) &= \Big\{C''_N(\overline{\Theta})(\widehat{\Theta}_N - \overline{\Theta}) \\ &\quad + \Big((\widehat{\Theta}_N - \overline{\Theta})^* \frac{\nabla_\Theta[\nabla_\Theta (C'_N)_l((\widetilde{\Theta}_N)_l)^*]}{2!}\Big)_{l=1,\dots,\dim\Xi}(\widehat{\Theta}_N - \overline{\Theta})\Big\},\end{aligned} \tag{C.13}$$

où chaque entrée $(\widehat{\Theta}_N - \overline{\Theta})^*\nabla_\Theta[\nabla_\Theta (C'_N)_l((\widetilde{\Theta}_N)_l)^*]$, $l = 1,\dots,\dim\Xi$, est un vecteur colonne. Par (5.39) et pour de grands $N$, $\widehat{\Theta}_N \in \operatorname{int}\Xi$ avec probabilité tendant vers 1. Par conséquent, $C'_N(\widehat{\Theta}_N) = 0$. Résoudre (C.13) pour $\widehat{\Theta}_N - \overline{\Theta}$ mène à

$$\begin{aligned}\sqrt{N}(\widehat{\Theta}_N - \overline{\Theta}) = -\Big\{&C''_N(\overline{\Theta}) + \Big((\widehat{\Theta}_N - \overline{\Theta})^* \\ &\cdot \frac{\nabla_\Theta[\nabla_\Theta C'_N((\widetilde{\Theta}_N)_l)^*]}{2!}\Big)_{l=1,\dots,\dim\Xi}\Big\}^{-1}\sqrt{N}C'_N(\overline{\Theta}).\end{aligned} \tag{C.14}$$

Sous (5.42), de part la consistance de l'estimateur $\widehat{\Theta}_N$ de $\overline{\Theta}$ et l'expression de $(f_N''')_{i_1,i_2,j}(\Theta)$, nous avons pour $l = 1, \ldots, \dim \Xi$,

$$\nabla_\Theta[\nabla_\Theta C_N'((\widetilde{\Theta}_N)_l)^*] \xrightarrow{P} \nabla_\Theta[\nabla_\Theta C_N'(\Theta_0)^*]. \tag{C.15}$$

L'inversibilité de la matrice entre crochets dans (C.14) est assurée avec probabilité tendant vers 1 grâce à la Condition (5.45) et le Théorème 5.1, puisque l'expression (C.11) entraine

$$C_N''(\overline{\Theta}) \xrightarrow{P} 2 \sum_{j=j_1}^{j_2} \sum_{1 \le i_1 \le i_2 \le M} \Lambda_{i_1,i_2}(2^j, \overline{\Theta}) \Lambda_{i_1,i_2}(2^j, \overline{\Theta})^*. \tag{C.16}$$

Soit $\|\cdot\|_{l^1}$ la norme matricielle $\ell_1$ entrée par entrée. A partir des relations (C.9) et (C.10), et du développement de Taylor au premier ordre de $\log_2|\cdot|$ autour de $E_{i_1,i_2}(2^j, \overline{\Theta})$ sous condition que (5.41) soit vérifié, on peut reformuler $\sqrt{N} C_N'(\overline{\Theta})$ comme suit :

$$-2 \sum_{j=j_1}^{j_2} 2^{j/2} \sqrt{K_j} \sum_{1 \le i_1 \le i_2 \le M} \frac{S_{i_1,i_2}(2^j) - E_{i_1,i_2}(2^j, \overline{\Theta})}{(\log 2) E_{i_1,i_2}(2^j, \overline{\Theta})} \Lambda_{i_1,i_2}(2^j, \overline{\Theta})$$

$$+ o\Big( \sum_{j=j_1}^{j_2} \sqrt{K_j} \|S(2^j) - E(2^j, \overline{\Theta})\|_{l^1} \Big) \in \mathbb{R}^{\dim \Xi}. \tag{C.17}$$

Il s'en suit alors que la limite faible (5.46) est une conséquence de (C.14), (C.15), (C.16), (C.17), du Théorème 5.1 et du dispositif de Cramér-Wold. □

## C.3 Discussion à propos de la Remarque 5.5

La condition (5.48) revient à exiger le plein rang de la somme de termes de rang 1. Afin de fixer les idées, considérons une somme de la forme $vv^* + ww^*$, où $v, w \in \mathbb{R}^2 \backslash \{0\}$. Alors, la somme est de rang 1 si et seulement si $v$ et $w$ sont colinéaires. En effet, supposons qu'il existe un vecteur $u \neq 0$ tel que $u^*\{vv^* + ww^*\}u = 0$. Alors, $u^*vv^*u = 0 = u^*ww^*u$, d'où la colinéarité.

Afin de simplifier les notations, supposons qu'il soit possible de réécrire le spectre en ondelettes comme

$$|E_{i_1,i_2}(2^j, \Theta)| = a_{i_1,i_2} 2^{j2h_1} + b_{i_1,i_2} 2^{j(h_1+h_2)} + c_{i_1,i_2} 2^{j2h_2}, \tag{C.18}$$

où $i_1, i_2 = 1, \ldots, M$ (c.f. [4, Lemme 4.2]. Supposons en plus que les seuls à paramètres soient $h_1 < h_2$. Alors, pour $j$ et $(i_1, i_2)$ fixes,

$$\frac{\partial}{\partial h_1} \log_2 |E_{i_1,i_2}(2^j, \Theta)| = \frac{1}{\log 2} \frac{a_{i_1,i_2} \log(2^{2j}) 2^{j2h_1} + b_{i_1,i_2} \log(2^j) 2^{j(h_1+h_2)}}{a_{i_1,i_2} 2^{j2h_1} + b_{i_1,i_2} 2^{j(h_1+h_2)} + c_{i_1,i_2} 2^{j2h_2}}, \tag{C.19}$$

et

$$\frac{\partial}{\partial h_2} \log_2 |E_{i_1,i_2}(2^j, \Theta)| = \frac{1}{\log 2} \frac{b_{i_1,i_2} \log(2^{2j}) 2^{j(h_1+h_2)} + c_{i_1,i_2} \log(2^j) 2^{j2h_2}}{a_{i_1,i_2} 2^{j2h_1} + b_{i_1,i_2} 2^{j(h_1+h_2)} + c_{i_1,i_2} 2^{j2h_2}}. \tag{C.20}$$

Les équations (C.19) et (C.20) suggèrent que pour au moins deux triplets $(i_1, i_2, j)$, les vecteurs

$$\Big( \frac{\partial}{\partial h_1} \log_2 |E_{i_1,i_2}(2^j, \Theta)|, \frac{\partial}{\partial h_2} \log_2 |E_{i_1,i_2}(2^j, \Theta)| \Big)^* \tag{C.21}$$

ne seront pas colinéaires en pratique pour la plupart des paramétrisations.

# Annexe D

# Estimateurs bayésiens

L'estimateur du maximum *a posteriori* (MAP) ou l'estimateur des moindres carrés (MMSE) associés à la distribution posterieure jointe $f(\boldsymbol{\theta}|\boldsymbol{y})$ dans (8.13) peut être approché à l'aide d'une procédure MCMC qui repose essentiellement sur un échantillonneur de Gibbs « partially collasped » [204] similaire à l'algorithme dérivé dans [69]. Le principe consiste à générer itérativement des échantillons (indicés par $\cdot^{[t]}$) selon les distributions postérieures conditionnelles associées à la distribution postérieure jointe (8.13). Les itérations sont rapportées par l'Algorithme 12. L'ensemble des échantillons $\boldsymbol{\vartheta} = \left\{\boldsymbol{r}^{[t]}, \boldsymbol{\mu}^{[t]}, \sigma^{2[t]}, p^{[t]}\right\}_{t=1}^{T_{\mathrm{MC}}}$ qui en résultent sont alors asymptotiquement distribués selon (8.13). Ces échantillons peuvent ensuite être utilisés pour approcher les estimateurs MAP et MMSE des paramètres d'intérêt [147]. Les solutions correspondantes sont notées $\widehat{\boldsymbol{x}}_{\mathrm{MAP}}$ et $\widehat{\boldsymbol{x}}_{\mathrm{MMSE}}$ dans la Section 8.4.4.

**Algorithme 12** Débruitage constant par morceaux Bayésien

**Entrée :** Observations $\boldsymbol{y} \in \mathbb{R}^N$.
Hyperparamètres $\boldsymbol{\Phi} = \left\{\alpha_0, \alpha_1, \mu_0, \sigma_0^2\right\}$.

**Iterations :**

**Pour** $t = 1, \ldots, T_{\mathrm{MC}}$ **faire**
  **Pour** $n = 1, \ldots, N-1$ **faire**
    Générer $r_n^{[t]} \sim f\left(r_n|\boldsymbol{y}, \boldsymbol{r}_{\setminus n}, p, \sigma^2, \mu_0, \sigma_0^2\right)$
  **Pour** $k = 1, \ldots, \sum_{n=1}^{N} r_n^{[t]}$ **faire**
    Générer $\mu_k^{[t]} \sim f\left(\mu_k|\boldsymbol{y}, \boldsymbol{r}, \sigma^2, \mu_0, \sigma_0^2\right)$
  Générer $\sigma^{2[t]} \sim f\left(\sigma^2|\boldsymbol{y}, \boldsymbol{r}, \boldsymbol{\mu}\right)$
  Générer $p^{[t]} \sim f\left(p|\boldsymbol{r}, \alpha_0, \alpha_1\right)$

**Sortie :** $\boldsymbol{\vartheta} = \left\{\boldsymbol{r}^{[t]}, \boldsymbol{\mu}^{[t]}, \sigma^{2[t]}, p^{[t]}\right\}_{t=1}^{T_{\mathrm{MC}}}$, $\widehat{\boldsymbol{x}}_{\mathrm{MAP}}$ et $\widehat{\boldsymbol{x}}_{\mathrm{MMSE}}$.

# Bibliographie


[1] MAWI Traffic Archive.

[2] Abry, P., Baraniuk, R., Flandrin, P., Riedi, R., and Veitch, D. (2002). Multiscale nature of network traffic. *IEEE Signal Process. Mag.*, 19(3) :28–46.

[3] Abry, P., Chainais, P., Coutin, L., and Pipiras, V. (2009). Multifractal random walks as fractional Wiener integrals. *IEEE Trans. Inform. Theory*, 55(8) :3825–3846.

[4] Abry, P. and Didier, G. (2015). Wavelet estimation for operator fractional Brownian motion. To appear in Bernoulli, arXiv preprint arXiv :1501.06094.

[5] Abry, P., Goncalves, P., and Flandrin, P. (1993). Wavelet-based spectral analysis of 1/f processes. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, volume 3, pages 237–240, Minneapolis, MN, USA.

[6] Achard, S., Bassett, D., Meyer-Lindenberg, A., and Bullmore, E. (2008). Fractal connectivity of long-memory networks. *Phys. Rev. E*, 77(3) :036104.

[7] Adler, A., Emiya, V., Jafari, M., Elad, M., Gribonval, R., and Plumbley, M. (2012). Audio inpainting. *IEEE Trans. Audio, Speech Language Process.*, 20(3) :922–932.

[8] Afonso, M. V. and Sanches, J. M. R. (2015). Blind inpainting using $\ell_0$ and total variation regularization. *IEEE Trans. Image Process.*, 24(7) :2239–2253.

[9] Amblard, P.-O. and Coeurjolly, J.-F. (2011). Identification of the multivariate fractional Brownian motion. *IEEE Trans. Signal Process.*, 59(11) :5152–5168.

[10] Amblard, P.-O., Coeurjolly, J.-F., Lavancier, F., and Philippe, A. (2012). Basic properties of the multivariate fractional Brownian motion. *Bulletin de la Société Mathématique de France, Séminaires et Congrès*, 28 :65–87.

[11] Anscombe, F. J. (1948). The transformation of Poisson, binomial and negative-binomial data. *Biometrika*, 35(3/4) :246–254.

[12] Antoine, J.-P., Murenzi, R., Vandergheynst, P., and Ali, S. T. (2008). *Two-dimensional wavelets and their relatives*. Cambridge University Press.

[13] Arneodo, A., Audit, B., Decoster, N., Muzy, J.-F., and Vaillant, C. (2002). Wavelet based multifractal formalism : applications to dna sequences, satellite images of the cloud structure, and stock market data. In *The science of Disasters*, pages 26–102. Springer.

[14] Aujol, J.-F., Ladjal, S., and Masnou, S. (2010). Exemplar-based inpainting from a variational point of view. *SIAM J. Math. Anal.*, 42(3) :1246–1285.

[15] Babacan, S. D., Molina, R., and Katsaggelos, A. K. (2007). Total variation image restoration and parameter estimation using variational posterior distribution approximation. In *Proc. Int. Conf. Image Process.*, volume 1, pages I–97, San Antonio, TX, USA.

[16] Bacry, E., Muzy, J. F., and Arnéodo, A. (1993). Singularity spectrum of fractal signals from wavelet analysis : Exact results. *J. Stat. Phys.*, 70(3) :635–674.

[17] Bae, E. and Tai, X.-C. (2009). *Efficient Global Minimization for the Multiphase Chan-Vese Model of Image Segmentation*, pages 28–41. Springer Berlin Heidelberg, Berlin, Heidelberg.

[18] Barbero, A. and Sra, S. (2011). Fast Newton-type methods for total variation regularization. In Getoor, L. and Scheffer, T., editors, *Int. Conf. Machine Learning*, pages 313–320. Omnipress.

[19] Barbero, Á. and Sra, S. (2014). Modular proximal optimization for multidimensional total-variation regularization. *arXiv preprint arXiv :1411.0589*.

[20] Barford, P., Kline, J., Plonka, D., and Ron, A. (2002). A Signal Analysis of Network Traffic Anomalies. In *Proc. ACM SIGCOMM Conf. Internet Measurement*, pages 71–82, Marseille, France.

[21] Bartlett, M. (1936). The square root transformation in analysis of variance. *J. Royal Stat. Soc. B*, 3(1) :68–78.

[22] Bauschke, H. H. and Combettes, P. L. (2011). *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York.

[23] Beck, A. and Teboulle, M. (2009). Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.*, 18(11) :2419–2434.

[24] Benedek, A. and Panzone, R. (1961). The space $l^p$ , with mixed norm. *Duke Math. J.*, 28(3) :301–324.

[25] Bertsekas, D. (1995). *Nonlinear Programming*. Athena Scientific.

[26] Bleakley, K. and Vert, J. P. (2011). The group fused Lasso for multiple change-point detection. Technical report, Preprint, Hal-00602121.

[27] Borgnat, P., Dewaele, G., Fukuda, K., Abry, P., and Cho, K. (2009). Seven years and one day : Sketching the evolution of Internet traffic. pages 711–719, Rio de Janeiro, Brazil.

[28] Boxma, O. J. and Cohen, J. W. (2000). *Self-Similar Network Traffic and Performance Evaluation*, chapter The Single Server Queue : Heavy Tails and Heavy Traffic. Wiley-Interscience.

[29] Boyd, S., Parikh, N., Chu, E., Peleato, B., and Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. *Found. Trends Mach. Learn.*, 3(1) :1–122.

[30] Boyd, S. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press, New York, NY, USA.

[31] Briceño-Arias, L. M., Combettes, P. L., Pesquet, J.-C., and Pustelnik, N. (2011). Proximal algorithms for multicomponent image recovery problems. *J. Math. Imag. Vis.*, 41(1-2) :3–22.

[32] Candès, E. and Recht, B. (2009). Exact matrix completion via convex optimization. *Found. Comput. Math.*, 9(6) :717–772.

[33] Cannon, J., Krokhmal, P. A., Lenth, R. V., and Murphey, R. (2010). An algorithm for online detection of temporal changes in operator cognitive state using real-time psychophysiological data. *Biomedical Signal Processing and Control*, 5(3) :229–236.

[34] Chaari, L., Pesquet, J.-C., Tourneret, J.-Y., and Ciuciu, P. (2011). Parameter estimation for hybrid wavelet-total variation regularization. In *Proc. IEEE Workshop Stat. Sign. Proc.*, Nice, France.

[35] Chambolle, A. (2004). An algorithm for total variation minimization and applications. *J. Math. Imag. Vis.*, 20(1-2) :89–97.

[36] Chambolle, A., Caselles, V., Cremers, D., Novaga, M., and Pock, T. (2010). An introduction to total variation for image analysis. In *Theoretical Foundations and Numerical Methods for Sparse Recovery*, volume 9, pages 263–340. De Gruyter.

[37] Chambolle, A., Cremers, D., and Pock, T. (2012). A convex approach to minimal partitions. *SIAM J. Imaging Sci.*, 5(4) :1113–1158.

[38] Chambolle, A. and Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imag. Vis.*, 40(1) :120–145.

[39] Chan, T. F. and Vese, L. A. (2001). Active contours without edges. *IEEE Trans. Image Process.*, 10(2) :266–277.

[40] Chaux, C., Combettes, P. L., Pesquet, J.-C., and Wajs, V. R. (2007). A variational formulation for frame-based inverse problems. *Inverse Problems*, 23(4) :1495–1518.

[41] Chierchia, G., Pustelnik, N., Pesquet-Popescu, B., and Pesquet, J.-C. (2014). A nonlocal structure tensor based approach for multicomponent image recovery problems. *IEEE Trans. Image Process.*, 23(12) :5233–5248.

[42] Cho, K., Mitsuya, K., and Kato, A. (2000). Traffic data repository at the WIDE project. In *USENIX Annual Technical Conference : FREENIX Track*, pages 263–270, San Diego, CA, USA.

[43] Ciuciu, P., Abry, P., and He, B. J. (2014). Interplay between functional connectivity and scale-free dynamics in intrinsic fMRI networks. *NeuroImage*, 95 :248–263.

[44] Clausen, J. (1997). Branch and bound algorithms-principles and examples. *Parallel Computing in Optimization*, pages 239–267.

[45] Coeurjolly, J.-F., Amblard, P.-O., and Achard, S. (2013). Wavelet analysis of the multivariate fractional Brownian motion. *ESAIM : Probability and Statistics*, 17 :592–604.

[46] Coifman, R. R. and Donoho, D. L. (1995). *Translation-Invariant De-Noising*, pages 125–150. Springer.

[47] Combettes, P. L. (2004). Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53 :475–504.

[48] Combettes, P. L. and Pesquet, J.-C. (2007). A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery. *IEEE J. Selected Topics Signal Process.*, 1(4) :564–574.

[49] Combettes, P. L. and Pesquet, J.-C. (2010). Proximal splitting methods in signal processing. In Bauschke, H. H., Burachik, R., Combettes, P. L., Elser, V., Luke, D. R., and Wolkowicz, H., editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, pages 185–212. Springer-Verlag, New York.

[50] Combettes, P. L. and Pesquet, J.-C. (2011). Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. *Set-Valued Var. Anal.*, 20(2) :307–330.

[51] Combettes, P. L. and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Model. and Simul.*, 4(4) :1168–1200.

[52] Condat, L. (2013a). A direct algorithm for 1D total variation denoising. *IEEE Signal Process. Lett.*, 20(11) :1054–1057.

[53] Condat, L. (2013b). A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. *J. Optim. Theory Appl.*, 158(2) :460–479.

[54] Condat, L. (2016). Discrete Total Variation : New Definition and Minimization. Research report, GIPSA-lab. Rapport interne de GIPSA-lab.

[55] Condat, L. and Pustelnik, N. (2015). Segmentation d'image par optimisation proximale. In *Proc. GRETSI*, pages pp :x+5, Lyon, France. in French.

[56] Crouch, C. H., Sauter, O., Wu, X., Purcell, R., Querner, C., Drndic, M., and Pelton, M. (2010). Facts and artifacts in the blinking statistics of semiconductor nanocrystals. *Nano Letters*, 10(5) :1692–1698.

[57] Crovella, M. E. and Bestavros, A. (1997). Self-similarity in world wide web traffic : Evidence and possible causes. *IEEE Trans. Networking*, 5(6) :835–846.

[58] Davies, P. and Kovac, A. (2001). Local extremes, runs, strings and multiresolution. *Ann. Stat.*, 29(1) :1–65.

[59] Deledalle, C., Vaiter, S., Peyré, G., Fadili, J. M., and Dossal, C. (2012). Unbiased risk estimation for sparse analysis regularization. In *Proc. Int. Conf. Image Process.*, Orlando, FL, USA.

[60] Deledalle, C.-A., Papadakis, N., and Salmon, J. (2015). On debiasing restoration algorithms : Applications to total-variation and nonlocal-means. In Aujol, J.-F., Nikolova, M., and Papadakis, N., editors, *Scale Space and Variational Methods in Computer Vision*, pages 129–141. Springer International Publishing, Lège-Cap Ferret, France.

[61] Deshayes, J. and Picard, D. (1986). *Detection of Abrupt Changes in Signals and Dynamical Systems*, chapter Off-line statistical analysis of change-point models using non parametric and likelihood methods, pages 103–168. Springer Berlin Heidelberg, Berlin, Heidelberg.

[62] Dewaele, G., Fukuda, K., Borgnat, P., Abry, P., and Cho, K. (2007). Extracting hidden anomalies using sketch and non gaussian multiresolution statistical detection procedure. In *Proc. ACM SIGCOMM Workshop on Large Scale Attack Defense*, pages 145–152.

[63] Didier, G., Helgason, H., and Abry, P. (2015). Demixing multivariate-operator self-similar processes. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, Brisbane, Australia.

[64] Didier, G. and Pipiras, V. (2011). Integral representations and properties of operator fractional Brownian motions. *Bernoulli*, 17(1) :1–33.

[65] Didier, G. and Pipiras, V. (2012). Exponents, symmetry groups and classification of operator fractional Brownian motions. *J. Theor. Probab.*, 25(2) :353–395.

[66] Dobigeon, N. and Tourneret, J.-Y. (2007). Joint segmentation of wind speed and direction using a hierarchical model. *Comput. Stat. Data Anal.*, 51(12) :5603–5621.

[67] Dobigeon, N., Tourneret, J.-Y., and Davy, M. (2007a). Joint segmentation of piecewise constant autoregressive processes by using a hierarchical model and a Bayesian sampling approach. *IEEE Trans. Signal Process.*, 55(4) :1251–1263.

[68] Dobigeon, N., Tourneret, J.-Y., and Scargle, J. D. (2005). Change-point detection in astronomical data by using a hierarchical model and a baysesian sampling approach. In *Proc. IEEE Workshop Stat. Sign. Proc.*, pages 369–374, Bordeaux, France.

[69] Dobigeon, N., Tourneret, J.-Y., and Scargle, J. D. (2007b). Joint segmentation of multivariate astronomical time series : Bayesian sampling with a hierarchical model. *IEEE Trans. Signal Process.*, 55(2) :414–423.

[70] Donoho, D. L. (1995). De-noising by soft-thresholding. *IEEE Trans. Inform. Theory*, 41(3) :613–627.

[71] Doret, M., Helgason, H., Abry, P., Goncalves, P., Gharib, C., and Gaucherand, P. (2011a). Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor. *American journal of perinatology*, 28(4) :259–266.

[72] Doret, M., Helgason, H., Abry, P., Goncalves, P., Gharib, C., and Gaucherand, P. (2011b). Multifractal analysis of fetal heart rate variability in fetuses with and without severe acidosis during labor. *Am. J. Perinatol.*, 28(04) :259–266.

[73] Doret, M., Spilka, J., Chudáček, V., Gonçalves, P., and Abry, P. (2015). Fractal analysis and hurst parameter for intrapartum fetal heart rate variability analysis : a versatile alternative to frequency bands and LF/HF ratio. *PLoS ONE*, 10(8) :e0136661.

[74] Douglas, J. and Rachford, H. H. (1956). On the numerical solution of the heat conduction problem in two and three space variables. *Trans. Amer. Math. Soc.*, 82(2) :421–439.

[75] Duan, J., Soussen, C., Brie, D., Idier, J., Wang, Y. P., and Wan, M. (2015). An optimal method to segment piecewise Poisson distributed signals with application to sequencing data. In *IEEE Conf. Eng. Med. Biol. Soc.*, pages 6465–6468.

[76] Dümbgen, L. and Kovac, A. (2009). Extensions of smoothing via taut strings. *Electron. J. Statist.*, 3 :41–75.

[77] Durand, S. and Froment, J. (2003). Reconstruction of wavelet coefficients using total variation minimization. *SIAM J. Sci. Comput.*, 24(5) :1754–1767.

[78] Eckstein, J. and Bertekas, D. P. (1992). On the Douglas-Rachford splitting methods and the proximal point algorithm for maximal monotone operators. *Math. Program.*, 55(3) :293–318.

[79] El Zehiry, N., Xu, S., Sahoo, P., and Elmaghraby, A. (2007). Graph cut optimization for the Mumford-Shah model. In *Proc. Int. Conf. Visualization, Imaging and Image Process.*, VIIP '07, pages 182–187, Anaheim, CA, USA. ACTA Press.

[80] Falconer, K. (1990). *Fractal geometry : mathematical foundations and applications.* J. Wiley & sons, Chichester, New York, Weinheim.

[81] Fenchel, W. (1953). *Convex cones, sets and functions.* Princeton University, Philadelphia, PA.

[82] Flandrin, P. (1992). Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Trans. Inform. Theory*, 38(2) :910–917.

[83] Fontugne, R., Abry, P., Fukuda, K., Borgnat, P., Mazel, J., Wendt, H., and Veitch, D. (2015). Random projection and multiscale wavelet leader based anomaly detection and address identification in internet traffic. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 5530–5534, Brisbane, Australia.

[84] Fontugne, R., Borgnat, P., Abry, P., and Fukuda, K. (2010). MAWILab : Combining diverse anomaly detectors for automated anomaly labeling and performance benchmarking. *Proc. ACM Co-NEXT.*

[85] Frecon, J., Didier, G.and Pustelnik, N., and Abry, P. (2016a). Non-linear wavelet regression and branch & bound optimization for the full identification of bivariate operator fractional brownian motion. *IEEE Trans. Signal Process.* To appear.

[86] Frecon, J., Fontugne, R., Didier, G., Pustelnik, N., Fukuda, K., and Abry, P. (2016b). Non-linear regression for bivariate self-similarity identification - application to anomaly detection in Internet traffic based on a joint scaling analysis of packet and byte counts. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, pages 4184–4188, Shanghai, China.

[87] Frecon, J., Pustelnik, N., Abry, P., and Condat, L. (2016c). On-the-fly approximation of multivariate total variation minimization. *IEEE Trans. Signal Process.*, 64(9) :2355–2364.

[88] Frecon, J., Pustelnik, N., Dobigeon, N., Wendt, H., and Abry, P. (2014). Hybrid Bayesian variational scheme to handle parameter selection in total variation signal denoising. In *Proc. Eur. Sig. Proc. Conference*, pages 1716–1720, Lisbon, Portugal.

[89] Frecon, J., Pustelnik, N., Dobigeon, N., Wendt, H., and Abry, P. (2016d). Bayesian driven selection of the regularization parameter of the $\ell_0$ in piecewise constant signal denoising problems. *IEEE Trans. Signal Process.* To be submitted.

[90] Frecon, J., Pustelnik, N., Wendt, H., and Abry, P. (2015a). Multivariate optimization for multifractal-based texture segmentation. In *Proc. Int. Conf. Image Process.*, pages 4957–4961, Quebec City, Canada.

[91] Frecon, J., Pustelnik, N., Wendt, H., and Abry, P. (2015b). Variation totale multivariée pour la détection de changement du spectre multifractal. In *Proc. GRETSI*, Lyon, France.

[92] Frecon, J., Pustelnik, N., Wendt, H., Condat, L., and Abry, P. (2016e). Multifractal-based texture segmentation using variational procedure. In *IEEE IVMSP Workshop : Perception and Visual Signal Analysis*, Bordeaux, France.

[93] Friedrich, F., Kempe, A., Liebscher, V., and Winkler, G. (2008). Complexity penalized m-estimation : fast computation. *J. Comput. Graph. Statist.*, 17(1) :201–224.

[94] Fryzlewicz, P. and Nason, G. P. (2004). A Haar-Fisz algorithm for Poisson intensity estimation. *Journal of computational and graphical statistics*, 13(3) :621–638.

[95] Gagne, Y. (1987). *Etude expérimentale de l'intermittence et des singularités dans le plan complexe en turbulence développée*. PhD thesis, Grenoble 1.

[96] Girshick, M. A. and Rubin, H. (1952). A Bayes approach to a quality control model. *Ann. Stat.*, 23(1) :114–125.

[97] Godinez, M., Jimenez, A., Ortiz, R., and Pena, M. (2003). On-line fetal heart rate monitor by phonocardiography. *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3141–3144.

[98] Gramfort, A., Kowalski, M., and Hämäläinen, M. (2012). Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Physics in Medicine and Biology*, 57(7) :1937–1961.

[99] Gupta, H., Ribeiro, V. J., and Mahanti, A. (2010). A longitudinal study of small-time scaling behavior of internet traffic. In Crovella, M., Feeney, L. M., Rubenstein, D., and Raghavan, S. V., editors, *NETWORKING 2010*, Lecture Notes in Computer Science, pages 83–95. Springer.

[100] Hamon, R., Borgnat, P., Flandrin, P., and Robardet, C. (2014). Discovering the structure of complex networks by minimizing cyclic bandwidth sum. *arXiv preprint arXiv :1410.6108*.

[101] Hamza, A. B., Luque-Escamilla, P. L., Martínez-Aroza, J., and Román-Roldán, R. (1999). Removing noise and preserving details with relaxed median filters. *J. Math. Imag. Vis.*, 11(2) :161–177.

[102] Hannart, A. and Naveau, P. (2009). Bayesian multiple change points and segmentation : Application to homogenization of climatic series. *Water Ressource Research*, 45 :1–14.

[103] Hansen, P. C. (2000). The l-curve and its use in the numerical treatment of inverse problems. In *in Computational Inverse Problems in Electrocardiology, ed. P. Johnston, Advances in Computational Bioengineering*, pages 119–142. WIT Press.

[104] Helgason, H., Pipiras, V., and Abry, P. (2011a). Fast and exact synthesis of stationary multivariate Gaussian time series using circulant embedding. *Signal Process.*, 91(5) :1123 – 1133.

[105] Helgason, H., Pipiras, V., and Abry, P. (2011b). Synthesis of multivariate stationary series with prescribed marginal distributions and covariance using circulant matrix embedding. *Signal Process.*, 91(8) :1741 – 1758.

[106] Hinkley, D. and Hinkley, E. (1970). Inference about the change-point in a sequence of binomial variables. *Biometrika*, 57(3) :477–488.

[107] Hoefling, H. (2010). A path algorithm for the fused Lasso signal approximator. *J. Comput. Graph. Statist.*, 19(4) :984–1006.

[108] Houel, J., Doan, Q. T., Cajgfinger, T., Ledoux, G., Amans, D., Aubret, A., Dominjon, A., Ferriol, S., Barbier, R., Nasilowski, M., et al. (2015). Autocorrelation analysis for the unbiased determination of power-law exponents in single-quantum-dot blinking. *ACS nano*, 9(1) :886–893.

[109] Ichida, K. and Fujii, Y. (1979). An interval arithmetic method for global optimization. *Computing*, 23(1) :85–97.

[110] Jaccard, P. (1901). Distribution de la flore alpine dans le bassin des Dranses et dans quelques régions voisines. *Bulletin de la Société Vaudoise des Sciences Naturelles*, (37) :241–272.

[111] Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., and Tsai, T. T. (2005). An algorithm for optimal partitioning of data on an interval. *IEEE Signal Process. Lett.*, 12(2) :105–108.

[112] Jaffard, S. (1991). Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publicacions Matemàtiques*, 35(1) :155–168.

[113] Jaffard, S. (1999). The multifractal nature of lévy processes. *Probab. Theor. Relat. Field*, 114(2) :207–227.

[114] Jaffard, S. (2004). Wavelet techniques in multifractal analysis. In Lapidus, M. and van Frankenhuijsen, M., editors, *Fractal Geometry and Applications : A Jubilee of Benoît Mandelbrot, M. Lapidus and M. van Frankenhuijsen Eds., Proceedings of Symposia in Pure Mathematics*, volume 72, pages 91–152. AMS.

[115] Jaffard, S., Abry, P., Roux, S. G., Vedel, B., and Wendt, H. (2007). The contribution of wavelets in multifractal analysis. In Ta-Tsien, L., Damlamian, A., and Jaffard, S., editors, *The Zuhai Conference on Wavelets and Applications*, Zuhai, China. Higher education press et world scientific publishing.

[116] Jain, A. K. (1989). *Fundamentals of Digital Image Processing*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[117] Jaulin, L., Kieffer, M., and Didrit, O. (2001). *Applied interval analysis : with examples in parameter and state estimation, robust control and robotics*. Springer, London.

[118] Jennane, R., Ohley, W. J., Majumdar, S., and Lemineur, G. (2001). Fractal analysis of bone x-ray tomographic microscopy projections. *IEEE Trans. Med. Imag.*, 20(5) :443–449.

[119] Johnson, C., Messier, P., Sethares, W., Klein, A., Brown, C., Do, A., Klausmeyer, P., Abry, P., Jaffard, S., Wendt, H., Roux, S., Pustelnik, N., Van Noord, N., Van Der Maaten, L., Postma, E., Coddington, J., Daffner, L., Murata, H., Wilhelm, H., Wood, S., and Messier, M. (2014). Pursuing automated classification of historic photographic papers from raking light images. *Journal of the American Institute for Conservation*, 53(3) :159–170.

[120] Kearfott, R. B. (1992). An interval branch and bound algorithm for bound constrained optimization problems. *J. Global. Optim.*, 2(3) :259–280.

[121] Kolmogorov, V., Pock, T., and Rolinek, M. (2015). Total variation on a tree. *ArXiv e-prints*.

[122] Kowalski, M. (2009). Sparse regression using mixed norms. *Appl. Comp. Harm. Analysis*, 27(3) :303–324.

[123] Krishnamurthy, B., Sen, S., Zhang, Y., and Chen, Y. (2003). Sketch-based change detection : Methods, evaluation, and applications. In *Proc. ACM SIGCOMM Conf. Internet Measurement*, pages 234–247, Miami, FL, USA.

[124] Kullback, S. and Leibler, R. A. (1951). On information and sufficiency. *Ann. Math. Stat.*, 22(1) :79–86.

[125] Kulzer, F. and Orrit, M. (2004). Single-molecule optics. *Annu. Rev. Phys. Chem.*, 55 :585–611.

[126] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder–Mead simplex method in low dimensions. *SIAM J. Opt.*, 9(1) :112–147.

[127] Lakhina, A., Crovella, M., and Diot, C. (2005). Mining Anomalies Using Traffic Feature Distributions. *ACM SIGCOMM '05*, pages 217–228.

[128] Lashermes, B., Jaffard, S., and Abry, P. (2005). Wavelet leader based multifractal analysis. In *Proc. Int. Conf. Acoust., Speech Signal Process.*, volume 4, pages iv–161. IEEE.

[129] Lavielle, M. (1998). Optimal segmentation of random processes. *IEEE Trans. Signal Process.*, 46(5) :1365–1373.

[130] Lavielle, M. and Lebarbier, E. (2001). An application of MCMC methods for the multiple change-points problem. *Signal Process.*, 81(1) :39–53.

[131] Lavielle, M. and Moulines, E. (1999). Least-squares estimation of an unknown number of shifts in a time series. *Journal of Time Series Analysis*, 21(1) :33–59.

[132] Leland, W. E., Taqqu, S., Willinger, W., and Wilson, D. V. (1994). On the self-similar nature of ethernet traffic. *IEEE Trans. Networking*, 2(1) :1–15.

[133] Lellmann, J., Kappes, J., Yuan, J., Becker, F., and Schnörr, C. (2009). Convex multiclass image labeling by simplex-constrained total variation. In *Scale Space and Variational Methods in Computer Vision*, pages 150–162. Springer.

[134] Lopes, R. and Betrouni, N. (2009). Fractal and multifractal analysis : a review. *Medical Image Analysis*, 13(4) :634–649.

[135] Lorden, G. (1971). Procedures for reacting to a change in distribution. *Ann. Stat.*, 42(6) :1897–1908.

[136] Louchet, C. and Moisan, L. (2011). Total variation as a local filter. *SIAM J. Imaging Sci.*, 4(2) :651–694.

[137] Makitalo, M. and Foi, A. (2011a). A closed-form approximation of the exact unbiased inverse of the Anscombe variance-stabilizing transformation. *IEEE Trans. Image Process.*, 20(9) :2697–2698.

[138] Makitalo, M. and Foi, A. (2011b). Optimal inversion of the Anscombe transformation in low-count Poisson image denoising. *IEEE Trans. Image Process.*, 20(1) :99–109.

[139] Mallat, S. (2008). *A Wavelet Tour of Signal Processing, Third Edition : The Sparse Way*. Academic Press, 3rd edition.

[140] Mallat, S. and Hwang, W. (1992). Singularity detection and processing with wavelets. *IEEE Trans. Inform. Theory*, 38(2) :617–643.

[141] Mammen, E. and Van de Geer, S. (1997). Locally adaptive regression splines. *Ann. Stat.*, 25(1) :1–434.

[142] Mandelbrot, B. (1975). *Les objets fractals : forme, hasard et dimension*. Nouvelle bibliothèque scientifique. Flammarion, Paris.

[143] Mandelbrot, B. (1983). *The fractal geometry of nature*. W. H. Freeman and Co. cop., New York.

[144] Mandelbrot, B. (1999a). Intermittent turbulence in self-similar cascades : divergence of high moments and dimension of the carrier. In *Multifractals and 1/f noise*, pages 317–357. Springer.

[145] Mandelbrot, B. (1999b). *Survey of multifractality in finance*, volume 1238 of *Cowles Foundation discussion paper*. The Foundation, New Haven, CT, USA.

[146] Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10(4) :422–437.

[147] Marin, J. and Robert, C. (2007). *Bayesian Core : A Practical Approach to Computational Bayesian Statistics*. Springer, New York, NY, USA.

[148] Mazel, J., Fontugne, R., and Fukuda, K. (2014). Taxonomy of anomalies in backbone network traffic. In *Proceedings of the fifth International Workshop on TRaffic Analysis and Characterization (TRAC '14)*, pages 30–36, Nicosia, Cyprus.

[149] Melot, C., Boursier, Y., Aujol, J.-F., and Anthoine, S. (2012). Some proximal methods for Poisson intensity CBCT and PET. *Inverse Probl. Imag.*, 6(4) :p. 565–598.

[150] Micchelli, C. A., Shen, L., and Xu, Y. (2011). Proximity algorithms for image models : denoising. *Inverse Problems*, 27(4) :045009.

[151] Mohammad-Djafari, A. and Féron, O. (2006). Bayesian approach to change points detection in time series. *Int. J. Imag. Syst. Tech.*, 16(5) :215–221.

[152] Moore, R., Hansen, E., and Leclerc, A. (1992). Recent advances in global optimization. chapter Rigorous Methods for Global Optimization, pages 321–342. Princeton University Press, Princeton, NJ, USA.

[153] Moore, R. E. (1966). *Interval analysis*. Prentice-Hall Inc., Englewood Cliffs, N.J.

[154] Moore, R. E., Kearfott, R. B., and Cloud, M. J. (2009). *Introduction to Interval Analysis*. Society for Industrial and Applied Mathematics.

[155] Moreau, J. J. (1963). Fonctionnelles sous-différentiables. *C. R. Acad. Sci., Paris*, 257 :4117–4119.

[156] Moreau, J. J. (1965). Proximité et dualité dans un espace hilbertien. *Bull. Soc. Math. France*, 93 :273–299.

[157] Nedic, A. and Bertsekas, D. (2001). Incremental subgradient methods for nondifferentiable optimization. *SIAM J. Opt.*, 12(1) :109–138.

[158] Nelson, J., Nafornta, C., and Isar, A. (2016). Semi-local scaling exponent estimation with box-penalty constraints and total-variation regularization. *IEEE Trans. Image Process.*, 25(6) :3167–3181.

[159] Nesterov, Y. (2009). Primal-dual subgradient methods for convex problems. *Math. Program.*, 120(1) :221–259.

[160] Ninin, J. (2010). *Optimisation Globale basée sur l'Analyse d'Intervalles : Relaxation Affine et Limitation de la Mémoire*. PhD thesis, Institut National Polytechnique de Toulouse, Toulouse.

[161] Nirmal, M., Dabbousi, B. O., Bawendi, M. G., Macklin, J. J., Trautman, J. K., Harris, T. D., and Brus, L. E. (1996). Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature*, 383(6603) :802–804.

[162] Nychis, G., Sekar, V., Andersen, D. G., Kim, H., and Zhang, H. (2008). An empirical evaluation of entropy-based traffic anomaly detection. *ACM IMC '08*, pages 151–156.

[163] Page, E. (1954). Continuous inspection schemes. *Biometrika*, 41(1-2) :100–115.

[164] Paxson, V. and Floyd, S. (1995). Wide area traffic : The failure of Poisson modeling. *IEEE Trans. Networking*, 4(3) :209–223.

[165] Pereyra, M., Dobigeon, N., Batatia, H., and Tourneret, J.-Y. (2013). Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm. *IEEE Trans. Image Process.*, 22(6) :2385–2397.

[166] Peyré, G. and Fadili, J. (2011). Group sparsity with overlapping partition functions. In *Proc. Eur. Sig. Proc. Conference*, pages x+5, Barcelona, Spain.

[167] Picard, F., Robin, S., Lavielle, M., Vaisse, C., and Daudin, J.-J. (2005). A statistical approach for array CGH data analysis. *BMC Bioinformatics*, 6 :27.

[168] Pock, T., Schoenemann, T., Graber, G., Bischof, H., and Cremers, D. (2008). A convex formulation of continuous multi-label problems. In *Proc. European Conference on Computer Vision*, pages 792–805. Springer Berlin Heidelberg.

[169] Poljak, B. T. (1987). *Introduction to optimization*. Optimization Software.

[170] Pollak, M. (1985). Optimal detection of a change in distribution. *Ann. Stat.*, 13(1) :206–227.

[171] Pollak, M. (1987). Average run lengths of an optimal method of detecting a change in distribution. *Ann. Stat.*, 15(2) :749–779.

[172] Punskaya, E., Andrieu, C., Doucet, A., and Fitzgerald, W. (2002). Bayesian curve fitting using MCMC with applications to signal segmentation. *IEEE Trans. Signal Process.*, 50(3) :747–758.

[173] Pustelnik, N. (2010). *Proximal methods for the resolution of inverse problems : application to positron emission tomography*. Theses, Université Paris-Est.

[174] Pustelnik, N., Chaux, C., and Pesquet, J.-C. (2011). Parallel proximal algorithm for image restoration using hybrid regularization. *IEEE Trans. Image Process.*, 20(9) :2450–2462.

[175] Pustelnik, N., Chaux, C., Pesquet, J.-C., and Comtat, C. (2010). Parallel algorithm and hybrid regularization for dynamic PET reconstruction. In *IEEE Med. Imag. Conf.*, pages 2423–2427, Knoxville, Tennessee. 5p.

[176] Pustelnik, N., Wendt, H., Abry, P., and Dobigeon, N. (2016). Local regularity, wavelet leaders and total variation based procedures for texture segmentation. Technical report, arXiv :1504.05776.

[177] Quan, L., Heidemann, J., and Pradkin, Y. (2013). Trinocular : Understanding Internet reliability through adaptive probing. *Proc. ACM SIGCOMM*, pages 255–266.

[178] Ratschek, H. and Rokne, J. (1988). *New Computer Methods for Global Optimization*. Halsted Press, New York, NY, USA.

[179] Robert, R. and Vargas, V. (2010). Gaussian multiplicative chaos revisited. *The Annals of Probability*, 38(2) :605–631.

[180] Rockafellar, R. T. (1967). Duality and stability in extremum problems involving convex functions. *Pacific J. Math.*, 21(1) :167–187.

[181] Roux, S. G., Arneodo, A., and Decoster, N. (2000). A wavelet-based method for multifractal image analysis. III. Applications to high-resolution satellite images of cloud structure. *Eur. Phys. J. B*, 15(4) :765–786.

[182] Rudin, L., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Phys.D*, 60(1-4) :259–268.

[183] Salmenkivi, M., Kere, J., and Mannila, H. (2002). Genome segmentation using piecewise constant intensity models and reversible jump mcmc. *Bioinformatics*, 18(suppl 2) :S211–S218.

[184] Samorodnitsky, G. and Taqqu, M. (1994). *Stable non-Gaussian random processes*. Chapman and Hall, New York.

[185] Seuret, S. and Véhel, J. L. (2002). The local Hölder function of a continuous function. *Appl. Comp. Harm. Analysis*, 13(3) :263–276.

[186] Shiryaev, A. (1961). The problem of the most rapid detection of a disturbance in a stationary process. *Soviet Math. Dokl.*, 2 :795–799.

[187] Shiryaev, A. (1963). On optimum methods in quickest detection problems. *Theor. Probab. Appl.*, 8(1) :22–46.

[188] Shor, N. Z., Kiwiel, K. C., and Ruszcayǹski, A. (1985). *Minimization Methods for Non-differentiable Functions*. Springer-Verlag New York, Inc., New York, NY, USA.

[189] Silveira, F., Diot, C., Taft, N., and Govindan, R. (2010). ASTUTE : Detecting a different class of traffic anomalies. In *Proc. ACM SIGCOMM*, pages 267–278, New Delhi, India. ACM.

[190] Spilka, J., Frecon, J., Leonarduzzi, R., Pustelnik, N., Abry, P., and Doret, M. (2015). Intrapartum fetal heart rate classification from trajectory in sparse SVM feature space. In *IEEE Conf. Eng. Med. Biol. Soc.*, pages 2335–2338, Milan, Italy.

[191] Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Stat.*, 9(6) :1135–1151.

[192] Storath, M. and Weinmann, A. (2014). Fast partitioning of vector-valued images. *SIAM J. Imaging Sci.*, 7(3) :1826–1852.

[193] Storath, M., Weinmann, A., and Demaret, L. (2014). Jump-sparse and sparse recovery using Potts functionals. *IEEE Trans. Signal Process.*, 62(14) :3654–3666.

[194] Storath, M., Weinmann, A., and Unser, M. (2016). Exact algorithms for $L^1$-TV regularization of real-valued or circle-valued signals. *J. Sci. Comput.*, 38(1) :614–630.

[195] Strekalovskiy, E., Chambolle, A., and Cremers, D. (2012). A convex representation for the vectorial Mumford-Shah functional. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1712–1719, Providence, RI, USA.

[196] Strong, D. M., Aujol, J.-F., and Chan, T. F. (2006). Scale recognition, regularization parameter selection, and Meyer's G norm in total variation regularization. *Multiscale Model. and Simul.*, 5(1) :273–303.

[197] Taqqu, M. S., Willinger, W., and Sherman, R. (1997). Proof of a Fundamental Result in Self-Similar Traffic Modelling. *Comp. Commun. Rev.*, 27 :5–23.

[198] Thorup, M. and Zhang, Y. (2004). Tabulation based 4-universal hashing with applications to second moment estimation. In *ACM SIAM Symp. Discrete Algorithms*, pages 615–624.

[199] Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison Wesley.

[200] Vũ, B. C. (2011). A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Adv. Comput. Math.*, 38 :667–681.

[201] Vaiter, S., Golbabaee, M., Fadili, J., and Peyré, G. (2015). Model selection with low complexity priors. *Information and Inference*, page 52.

[202] Vaiter, S., Peyre, G., Dossal, C., and Fadili, J. (2013). Robust sparse analysis regularization. *IEEE Trans. Inform. Theory*, 59(4) :2001–2016.

[203] Van der Vaart, A. (2000). *Asymptotic Statistics*, volume 3. Cambridge University Press.

[204] Van Dyk, D. and Park, T. (2008). Partially collapsed Gibbs samplers : Theory and methods. *J. American Statist. Ass.*, 103(482) :790–796.

[205] Veitch, D. and Abry, P. (1999). A wavelet-based joint estimator of the parameters of long-range dependence. *IEEE Trans. Inform. Theory*, 45(3) :878–897.

[206] Vese, L. A. and Chan, T. F. (2002). A multiphase level set framework for image segmentation using the Mumford and Shah model. *Int. J. Comp. Vis.*, 50(3) :271–293.

[207] Vogel, C. (2002). *Computational Methods for Inverse Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA.

[208] Vogel, C. R. and Oman, M. E. (1996). Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1) :227–238.

[209] Wahlberg, B., Boyd, S., Annergren, M., and Wang, Y. (2012). An ADMM algorithm for a class of total variation regularized estimation problems. In *Preprints of the 16th IFAC Symposium on System Identification*, pages 83–88. QC 20121112.

[210] Wald, A. (1947). *Sequential Analysis*. John Wiley and Sons, 1st edition.

[211] Watkins, L. P. and Yang, H. (2005). Detection of intensity change points in time-resolved single-molecule measurements. *J Phys Chem B*, 109 :617–628.

[212] Wendt, H. (2008). *Contributions of wavelet leaders and bootstrap to multifractal analysis : images, estimation performance, dependence structure and vanishing moments. confidence intervals and hypothesis tests.* PhD thesis, Ecole Normale Supérieure de Lyon.

[213] Wendt, H., Abry, P., and Jaffard, S. (2007). Bootstrap for empirical multifractal analysis. *IEEE Signal Process. Mag.*, 24(4) :38–48.

[214] Willinger, W., Taqqu, M., and Erramilli, A. (1996). *Stochastic networks : Theory and applications*, chapter A Bibliographical Guide to Self-Similar Traffic and Performance Modeling for Modern High-Speed Networks, pages 339–366. Clarendon Press (Oxford University Press).

[215] Wilson, E. B. and Hilferty, M. M. (1931). The distribution of chi-square. *Proc. Nat.. Acad. Sci. Unit. States Am.*, 17(12) :684–688.

[216] Winkler, G., Kempe, A., Liebscher, V., and Wittich, O. (2005). *Innovations in Classification, Data Science, and Information Systems : Proceedings of the 27th Annual Conference of the Gesellschaft für Klassifikation e.V., Brandenburg University of Technology, Cottbus, March 12–14, 2003*, chapter Parsimonious Segmentation of Time Series by Potts Models, pages 295–302. Springer Berlin Heidelberg, Berlin, Heidelberg.

[217] Winkler, G. and Liebscher, V. (2002). Smoothers for discontinuous signals. *Journal of Nonparametric Statistics*, 14(1-2) :203–222.

[218] Y., L., Gabbouj, M., Astola, J., and Neuvo, Y. (1993). Optimal weighted median filters under structural constraints. In *Circuits and Systems, 1993., ISCAS '93, 1993 IEEE International Symposium on*, pages 942–945.

[219] Yao, Y. (1988). Estimating the number of change points via Schwarz criterion. *Stat. and Probab. Lett.*, 6 :181–189.

[220] Young Yang, T. and Kuo, L. (2001). Bayesian binary segmentation procedure for a Poisson process with multiple changepoints. *J. Comput. Graph. Statist.*, 10(4) :772–785.

[221] Zalinescu, C. (2002). *Convex Analysis in General Vector Spaces*. World Scientific.

[222] Zhang, K. and Yang, H. (2005). Photon-by-photon determination of emission bursts from diffusing single chromophores. *J Phys Chem B*, 109(46) :21930–7.